Nancy Cartwright's Philosophy of Science

Edited by
Stephan Hartmann,
Carl Hoefer and Luc Bovens

Routledge Studies in the Philosophy of Science

Nancy Cartwright's Philosophy of Science

Routledge Studies in the Philosophy of Science

1. Cognition, Evolution and Rationality
A Cognitive Science for the Twenty-First Century
Edited by António Zilhão

2. Conceptual Systems
Harold I. Brown

3. Nancy Cartwright's Philosophy of Science
Edited by Stephan Hartmann, Carl Hoefer and Luc Bovens

Nancy Cartwright's Philosophy of Science

Edited by
Stephan Hartmann,
Carl Hoefer and Luc Bovens

Routledge
Taylor & Francis Group
New York London

First published 2008
by Routledge
270 Madison Ave, New York, NY 10016

Simultaneously published in the UK
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

This edition published in the Taylor & Francis e-Library, 2008.

Library of Congress Cataloging in Publication Data

Cartwright, Nancy.
Nancy Cartwright's philosophy of science / edited by Stephan Hartmann, Carl Hoefer, and Luc Bovens.
p. cm. -- (Routledge studies in the philosophy of science ; 3)
Includes bibliographical references and index.
ISBN 978-0-415-38600-5
1. Science--Philosophy. 2. Science--Methodology. I. Hartmann, Stephan, Dr. II. Hoefer, Carl. III. Bovens, Luc. IV. Title. V. Title: Philosophy of science.
Q175.C373 2008
501--dc22 2007049822

ISBN 0-203-89546-0 Master e-book ISBN

ISBN10: 0-415-38600-4 (hbk)
ISBN10: 0-203-89546-0 (ebk)

ISBN13: 978-0-415-38600-5 (hbk)
ISBN13: 978-0-203-89546-7 (ebk)

Dedicated to the memory of
Daniela Bailer-Jones (1969–2006)

Contents

PART I.
Models and Representations

PART II.
Causes and Capacities

PART III.
Antifundamentalism and the Disunity of Science

Acknowledgments

Earlier versions of most of the chapters of this volume were presented at the workshop *Nancy Cartwright's Philosophy of Science* at the University of Konstanz, Germany (December 16–17, 2002). The conference was supported by the Thyssen Foundation, the Center for Junior Research Fellows at the University of Konstanz, and, through a Sofja Kovalevskaja Award, by the Alexander von Humboldt Foundation, the Federal Ministry of Education and Research, and the Program for the Investment in the Future (ZIP) of the German Government. We are grateful for their support. Each paper was refereed by two anonymous referees. We thank them for their constructive comments, and we thank Leah McClimans and Alice Obrecht for preparing the final manuscript.

Stephan Hartmann, Tilburg University
Carl Hoefer, ICREA and Universitat Autònoma de Barcelona
Luc Bovens, London School of Economics and Political Science

1 Introducing Nancy Cartwright's Philosophy of Science

Carl Hoefer

OVERVIEW: CARTWRIGHT'S EMPIRICISM AND THE STANFORD SCHOOL

Nancy Cartwright's philosophy of science is, in her view, a form of empiricism but empiricism in the style of Neurath and Mill, rather than of Hume or Carnap. Her concerns are not with the problems of skepticism, induction, or demarcation; she is concerned with how actual science achieves the successes it does, and what sort of metaphysical and epistemological presuppositions are needed to understand that success.

Cartwright, like many working scientists themselves, takes a rather pragmatic/realist stance toward observations and interventions made by scientists and engineers and particularly toward their connections to causality: Scientists *see impurities causing signal loss in a cable*, and they *stimulate an inverted population, causing it to lase*. Given these starting points, there can be no question of a skeptical attitude toward *causation*, in either singular or generic form. The fundamental role (or better, *roles*) played by causation in scientific practice is undeniable; what Cartwright does, then, is reconfigure empiricism from the ground up based on this insight. In the reconfiguration process, many mainstays of the received view of science take a beating; especially, as we will see, the fundamentality of laws of nature. We will come back to this point, as well as Cartwright's views on causation, below.

Rather than claiming allegiance to some traditional philosophical standpoint, Cartwright likes to think of her work as an example of the practice of the Stanford School of history/philosophy of science. This school is formed by fortuitous spatiotemporal proximity and a family resemblance in philosophy styles; the best way to describe it is by listing its practitioners: Pat Suppes, John Dupré, Ian Hacking, Margaret Morrison, Peter Galison, and of course Nancy Cartwright. One thing that unites Stanford School practitioners is a strong respect for scientific practice—*actual* scientific practice, as displayed in the best examples of scientific discovery and creation. If science has delivered genuine knowledge about our world—as it surely has—then studying its actual practices is the surest guide to an understanding of how that knowledge is gained. Case studies are indispensable for philosophy of science. Though not an end in themselves, they are invaluable for keeping our metaphysical and methodological speculations on track with real science.

Examples of this method are prominent in all three of Cartwright's main books (HTLPL, NCATM, and DW).[1] In the first book, Cartwright looks into the theory and engineering of lasers, finding that the fundamental, laws-driven "received view" of how such applied-physics items should relate to fundamental physics proves false at every turn. In the second book, economics examples are brought to bear on the metaphysical issues of causation and capacities to show that economic practice can only make sense if read as presupposing the existence of stable causal capacities, over and above regularities and probabilities. In the third, the BCS model of superconductivity is examined with the aim of arguing that quantum mechanics is a theory with definite, built-in *limits* in what it can pretend to cover—and hence no refuge for the beleaguered fundamentalist.

Stanford School philosophers are usually "empiricists" in some broad use of the term but, unlike their teachers, do not shy away from metaphysics when it is built into, and hence justified by, successful scientific practice. In Cartwright's work, metaphysics appears in some very specific guises concerning causation and related issues (natural kinds, properties, dispositions, counterfactuals . . .). But more generally, Cartwright has a metaphysical Big Picture that emerges with increasing clarity, becoming explicit in DW. She does not make her chief objective that of defending this big-picture view overtly. Instead, she points out how it emerges from her studies of science as a natural and largely overlooked alternative to the more traditional Humean/empiricist and Rationalist big-pictures. The name of this alternative view is, of course, the "Dappled World".

Cartwright is no theist of course, but it is nevertheless correct to say that for her, God is an Englishman rather than a Frenchman. This means that the world is more than a bit untidy and poorly organized; it has superficial rules rather than deep, necessary principles. Another useful contrast is this: Cartwright is more Aristotelian than Platonist. Universal, eternal forms, if they exist, are certainly no use to us in accounting for what actually happens in the world, whereas things' *natures* and *capacities* certainly are. The two central features of this worldview are the insistence on the reality of causation (and of causal capacities, or powers, etc.) and the insistence that so-called fundamental laws are no genuine, true part of nature. A *consequence* of these tenets is one of the Stanford School's central views, the *disunity of science*.

LYING LAWS AND FUNDAMENTALISTS

In her first book, HTLPL, Cartwright mounts her first sustained attack on two aspects of philosophy of science that she believes are deeply mistaken:

its rejection, based on a tradition beginning with Hume and reinforced by Russell, of causality and causal laws and its claim that finding and applying true laws of nature (typically in physics) is central to the success of science. Ironically Cartwright herself was still under the sway of this second tradition, for in rejecting the truth of the laws of physics, she took herself to be defending a partly antirealist view.[2] Only in DW did Cartwright come to see her opponent not as scientific realism but, rather, merely that law-centred metaphysical picture which she aptly named *fundamentalism*.

HTLPL discusses laws of all sorts: fundamental physical laws, less-fundamental equations, high-level phenomenological laws, and causal laws. Cartwright's arguments go to show that only causal laws, and some high-level phenomenological laws in physics, can be held to be literally true, even in a restricted domain of application; and all true laws are to be understood as merely true *ceteris paribus*—all else being equal, or better: when conditions are right. Why is *truth* such a rare and hedged quality for the laws of physics?

We can distinguish at least two lines of argument for this view. First, Cartwright argues that even fundamental laws such as Newton's law of gravity and Maxwell's equations are false in most real-world situations. She believes that Newton's law tells how an object behaves (falls) when there are no other causes operating on it; but if a body is charged and moving in an E-M field, or subject to air friction, then it is literally false. The movement predicted by the law is not what we see. Most philosophers of science find this argument puzzling; they take Newton's law as a description of a force that exists on a body in virtue of gravity, not as a prediction about motion. Newton's second law, $\sum_i \vec{F}_i / m = \vec{a}$, makes a prediction about motion, and it is always true, as long as we include *all* the forces present in our vector-addition summation. When both Newton's and Maxwell's laws are involved, the sum has two or more terms, but that does not make either of those laws false individually; they truly report *one component* of the forces present and at work.

Cartwright rejects this story; component forces, she argues, are not real, and—being intrinsically unmeasurable and unobservable—are not fit entities for an empiricist to postulate. There is some truth to the traditional story, of course—but only when read as a *causal* story rather than a story about true fundamental laws. Newton's gravity law correctly tells us about one cause of motion. Thanks to the fact that physics is simple (compared, say, to economics), in the second law, we have a simple rule about how multiple causes of motion combine. But even that law is only true *ceteris paribus*: when things affect a body for which we don't have a force law—e.g., when a child picks up the ball and carries it—then the second law is false along with Newton's gravity law and Maxwell's laws. We know perfectly well what is going on causally, but our physics is telling us only lies. The traditionalist believes that physics *does* entail a time-dependent force on the ball for the situation of the child carrying it, but it is simply too hard to calculate. This is an example of fundamentalist faith, which we return to below.

Cartwright's second line of attack on the truth of laws proceeds by looking at detailed cases of the application of physics to real phenomena which we take ourselves to have mastered—e.g. laser physics or quantum damping, which induces broadening of spectral lines. Looking carefully at what physicists actually do, Cartwright finds that their procedures argue against the truth of fundamental equations again. Those equations are used in the process of deriving less-fundamental equations—generally, in building a model—but so are a host of nonfundamental tools: approximations, ad-hoc corrections for known causal disturbances, etc. Moreover, a phenomenon such as quantum damping may have several theoretical models—derivations starting in part from the fundamental level (but only in part) and achieving a correct phenomenological equation. Physicists not only tolerate such a profusion of derivations, they seem to luxuriate in it. (The same is true of derivations of Einstein's Field Equations of gravity, e.g., in Misner, Thorne, and Wheeler's classic text *Gravitation*.) By contrast, Cartwright notes, physicists are not happy if they possess more than one rival *causal* explanation for a phenomenon such as quantum damping. There should be one, and just one, correct causal story. What this shows is that, whatever the official rhetoric may be, physicists are realists about causation and antirealists about theoretical explanations. This, for Cartwright, is the right attitude: The laws of physics do a lot of explanatory work for us, but that does not argue for their truth. Inference to the best explanation makes sense when one is inferring to the *most probable cause* but not when one is inferring to the alleged truth of a fundamental equation.

In DW, Cartwright resumes the attack on fundamental laws, but this time her aim is directed squarely against the imperialism she sees in fundamentalism: The belief that physics will, some day, give us The Truth, the equation or equations that are true everywhere and everywhen, and govern all happenings in the world, is a faith that Cartwright sees as both empirically unsupported and damaging to science. Whereas in HTLPL Cartwright took for granted the pretension of physics' laws to be universal and argued that they could not then be true, in DW she grants the truth of some physical laws and theories—but only in very restricted domains. By very clever engineering, we can sometimes get nature to behave in the regular ways found in physical equations; it takes a lot of causal knowledge, in general, to do this. But if we see the "truth" of fundamental laws displayed in such carefully contrived circumstances (such as a bubble chamber, or the Gravity Probe B), does this mean we should conclude that the laws are actually true *everywhere*, only hidden from our gaze by a veil of complexity? Not at all, urges Cartwright. Instead, we should make only a much more limited inductive move: In well-controlled circumstances of such-and-such kind, law L is true. Again, unlike causal powers, laws have no right to be "exported" beyond where we see them work.

In DW Cartwright goes beyond the view of science that she offered in HTLPL by offering a reconceptualized understanding of laws of nature (causal or otherwise) and a metaphysics (the dappled world) with which to replace the fundamentalist's reductionist world of particles moved by laws. Laws, to the extent that we need them, arise because of, and are true only in, *nomological machines*: setups, usually made by us but sometimes found in nature, that combine a simple/stable structure and sufficient shielding from outside influences so as to give rise to regular behavior. We will return to nomological machines and the dappled world, with its "patchwork" conception of both nature and science, below.

CAUSATION: CAUSAL LAWS, CAPACITIES, AND SINGULAR CAUSINGS

In HTLPL Cartwright has two opponents: the Humean-inspired empiricist who thinks that there are no such thing as *causal* laws, over and above mere regularities of association (especially statistical regularities); and the scientific realist who infers from the successes of science to the truth of fundamental laws in physics. The case against the former opponent is largely made in Essay 1, "Causal Laws and Effective Strategies", probably Cartwright's single most influential paper.

Here Cartwright argues for two main theses: first, that there is no way to reduce facts about causation to facts about probabilistic (statistical) relations; second, that in order to understand the effective strategies we use to achieve desired results, we need to accept that there are genuine *causal laws* in nature. Often, it is the fact that it is a causal law that C brings about E (or raises the level of E or makes E more probable . . .) that grounds our having an effective strategy for E. "If indeed, it *isn't true that* buying a TIAA policy is an effective way to lengthen one's life, but stopping smoking is, the difference between the two depends on the causal laws of our universe, and on nothing weaker" (Cartwright 1983: 22).

Cartwright does not define causal laws overtly, but rather *via* an implicit definition: At least, the true statements "$C \hookrightarrow E$" that pass the test of principle *CC* should be counted as causal laws. Principle *CC* gives what Cartwright believes is the strongest link that can be made between probabilities and generic causal facts. Here is the 1983 version of *CC*:

> $C \hookrightarrow E$ iff $\mathrm{Prob}(E|C \,\&\, K_j) > \mathrm{Prob}(E|K_j)$ for all state descriptions K_j over the set $\{C_i\}$, where $\{C_i\}$satisfies
>
> (i) $C_i \in \{C_i\} \Rightarrow C_i \hookrightarrow +/- E$
> (ii) $C \notin \{C_i\}$
> (iii) $\forall D\ (D \hookrightarrow +/- E \Rightarrow D = C \text{ or } D \in \{C_i\})$
> (iv) $C_i \in \{C_i\} \Rightarrow \neg\,(C \hookrightarrow C_i)$.[3]

In words: Cs cause Es if and only if the probability of E given C is greater than the probability of E simpliciter in each subpopulation in which we hold fixed the occurrence or nonoccurrence of each of the other factors C_i that are causes (or preventers) of E. The main work of the paper shows that this conditioning over all the other causes of E is really necessary; without it, i.e. in the total population or an improperly chosen subpopulation, the statistics may show Prob(E | C & B) > Prob(E | B) (where B is just a background condition), and yet C may not be a cause of E. This is the case with the TIAA insurance example Cartwright starts with: The probability of living past 75 if you are a TIAA member is higher than it is for the population at large; but it is just not true that joining TIAA is a cause of longer life.

CC is the strongest link that can be forged between probabilities and causation; but it is obviously not a candidate for a reductive analysis of causation, since the symbol standing for "causes" appears on both sides of the "iff". *CC* also makes clear the troubles that await any attempt to infer causal relationships from statistical data. On a superficial reading, what *CC* is telling us is that before we can use statistical data to determine whether C's cause E's, we need to have *already* established all the other positive and negative causes of E. In fact *CC* does not entail anything quite this strong. It may be that there are ways of ruling out, with reasonable confidence, misleading statistical correlations so that we can infer that C's cause E's without knowing all of E's potential causers and preventers. Cartwright thinks that randomised controlled trials aim to do exactly this. Nevertheless a serious epistemic problem is being laid bare here, namely the difficulty of inferring causal relationships from statistics: more about this problem below.

This paper sounded, in effect, the death knell for attempts to reductively define causation in terms of probabilities. It was also crucial in bringing philosophers of science back to the table to think about and discuss causality seriously and, in light of real-world examples, as a necessary ingredient of our overall understanding of science and the natural world. But what are these causal laws, then, what is their nature and status? Cartwright's views have evolved in the years since 1983. At first Cartwright advocated a kind of realism about general causal laws, understood as *ceteris paribus* laws: *ceteris paribus* because the stated cause–effect relation does not *always* occur but, rather, *always, if nothing gets in the way or goes wrong*. What accounts for this imperfect, but still very real, causal relationship? By the time of NCATM Cartwright wanted to answer this question.

Cartwright's 1983 discussion of causal laws takes place in the context of that book's project of undermining the traditional logical empiricist, covering-law account of laws and explanation. But by 1989 Cartwright was ready to put forward her own alternative picture of how science functions, and the roles of causation in scientific practice. As we will see, that picture gives preeminent place not to causal laws but to *causal capacities* taken as genuine ingredients of reality.

Using a wide variety of examples from both physics and economics to illustrate her points, in NCATM Cartwright mounts an impressive argument for the reality of *natures* and of *causal capacities* as indispensable ingredients of the worldview presupposed by modern science and its methods. The argument proceeds by stages throughout the book, starting with "laws of association". Such mere regularities, whether probabilistic or not, were the meat and potatoes of an earlier generation of empiricists. In NCATM Cartwright reviews and strengthens the argument of "Causal Laws and Effective Strategies": not only do we need to have prior causal knowledge in order to sort out genuine causal laws from mere correlations, but frequently we even need to know *singular causal* facts (i.e. facts about the presence or absence of a singular-causal relation on a given occasion.) Principle *CC* turns out to be false unless the statistical test population is homogenized in *just the right way*; and that way involves knowing singular causal facts:

> . . . what counts as the right populations in which to test causal laws by probabilities will depend not only on what other causal laws are true, but on what singular causal processes obtain as well. One must know, in each individual where *F* occurs, whether its occurrence was produced by *C*, or whether it came about in some other way. Otherwise the probabilities do not say anything, one way or the other, about the hypothesis in question. (Cartwright 1989: 96)

But once the primacy of singular causation is admitted—and Cartwright argues at length that this cannot be avoided—then what we end up with are not merely true causal laws (the "right" laws of association, the ones on which to base effective strategies) but, rather, causal capacities. Once one has tested the effect of taking birth control pills on thrombosis rates in the *right* populations—those in which the action or nonaction of other causes and preventers of thrombosis are held fixed in just the right ways—what emerges is a conclusion about whether birth control pills carry the capacity to cause thrombosis and what the strength of that capacity is in each such population. But those strengths, or rather the probabilistic causal laws that express them, are inevitably tied to contingent facts about populations, test- or real-world; they are not ontologically basic in any way. What *is* basic is the causal capacity carried by a thing, in virtue of its properties, that leads to singular causing of the effect when circumstances are right.

In the later chapters of NCATM Cartwright develops the notion of a capacity (and the related Aristotelian notion of a *nature*), showing that science both presupposes in all its experimental work, and aims at discovering, capacities and natures. Her chief examples, pursued in remarkable depth, come from economics and physics. But while the reconstruction of scientific methods used in both disciplines is clear and convincing, ironically the *ontological* conclusion in favor of capacities strikes one as stronger in physics—the traditional home of Russell's anticausalism—than in economics. "Keynes

. . . maintained that economic phenomena were probably not atomistic—that is, in the terminology of this book, economic life is not governed by stable capacities. John Stuart Mill believed that it was. . . ." (Cartwright 1989: 170). But who was more right, Keynes or Mill? Once we get over any initial discomfort with capacities to begin with, it is perhaps easier to see as real the capacity of undamped vibration to swamp signals in a sensitive detector than the capacity of money supply increase to cause inflation.

Cartwright's more recent work on causality continues the trend away from causal laws but adds some layers of refinement to her views on causal capacities. There are two main strands to the work: one negative, making war on those who neglect the lessons of HTLPL and NCATM regarding the links between statistical probabilities and causal conclusions; and one positive, elaborating a pluralistic and pragmatic approach to causality.

Despite the strict lessons of *CC* and its successors in NCATM (which only sharpened the difficulty of deriving causes from statistics), there is an undying desire among some philosophers of science to develop an account of the epistemology of causation that would (in principle) be applicable directly to statistics and allow us to draw causal conclusions without presupposing any causal knowledge at the start. The research groups of Pearl, and Spirtes, Glymour, and Scheines,[4] have developed mathematical frameworks, and computer programs that implement them, to do precisely this job. These "causal nets" or "Bayes nets" theorists are aiming at an ambitious and laudable goal: using sophisticated philosophy of science to create methodological tools that can actually be used to good effect by real-world scientists and statisticians. And Cartwright praises these goals; but her own studies impel her to raise several important caution flags.

One set of flags has to do with two necessary presuppositions of the Bayes nets methods: Faithfulness and the Causal Markov Condition (CMC). A set of statistical data giving the correlations between a number of variables is *faithful* if, whenever there is a genuine causal connection between two variables, that is manifested in a probabilistic correlation (positive or negative). Cartwright points out that there is no reason to suppose that even ideal statistics (reflecting the "true" probabilities) should meet this condition; some causal systems may be so structured as to balance a positive causal path between C and E against a negative one—leading to statistical independence of C and E, when in fact C both causes and prevents E. When we recall that we are never presented with the true probabilities in this world but, rather, only "imperfect" finite statistical data, things look even worse for faithfulness. We should expect it to be violated from time to time, just by "chance", in real data samples. Can we risk drawing conclusions about the absence of causality between two variables on the basis of an undetected failure of faithfulness?

Cartwright's attack on the CMC has been even more serious and sustained, leading to sharp exchanges between herself and Woodward and Hausman, who defend the condition. The Causal Markov Condition needs to be understood as a condition that may hold (or fail) in a causal model, where the model consists of a set of variables that may enter into various causal relations and a directed acyclic graph (DAG) encoding their presumed causal relationships. DAGs are the node and arrow diagrams that are becoming familiar to philosophers of science. In addition to a DAG, a causal model specifies the statistical relationships among all the variables modeled. (See Woodward, this volume, for examples.)

CMC says that once we hold fixed (conditionalize on) the parents of a variable *C*—that is, all the direct causes of *C*, the arrows into *C*—then *C* is statistically independent of any other variable in the model except its "descendents", i.e. its effects or the effects of its effects (etc.). As Cartwright notes, CMC tries to encode two familar notions about causality: a prohibition against causation across "temporal gaps" and Reichenbach's common cause principle, which says that "a full set of parents screens off the joint effects of any of these parents from each other" (Cartwright 1999: 107).

CMC is crucial to many of the theorems and search-techniques of the Bayes nets groups. But Cartwright thinks that we don't have any good grounds for assuming that it *always* holds, among the types of real-world variables for which we have data and can formulate causal models. The arguments put forward in defense of CMC presuppose both underlying determinism and a perfect reflection of the "true" probabilities in finite data samples, which we know is unlikely.[5] They also presuppose a fine-grainedness of event types that is unlikely to be achievable in real-world causal modeling, and for which in any event we have little or no evidence.

Despite these and other warning flags, Cartwright offers no blanket condemnation of Bayes nets methods; she merely withholds a full endorsement and urges care in their application. Before applying them we need to think carefully about whether their presuppositions are likely to be true in the area of interest—and preferably, use evidence to reach our conclusions about this. But like any other method used in parts of science, e.g. randomised controlled trials, Bayes nets can be a valuable addition to our toolbox of scientific methods. What her criticism of Bayes nets methods shows is another instance of a lesson she draws quite generally: The results of using these methods will only be right if the kind of model they presuppose is there to be had. In Cartwright's slogan: no model in, no causes out.

Much of Cartwright's recent work on causality is collected in a new book, *Hunting Causes and Using Them* (Cartwright 2007).

THE DAPPLED WORLD: THE METAPHYSICS AND METHODS OF GOOD SCIENCE

The "dappled world" we live in, according to Cartwright, is subject to the rule of law only in a patchy, piecemeal way. "For all we know," she writes in the Introduction to DW, "most of what occurs in nature occurs by hap, subject to no law at all." An astrophysicist might beg to differ, but at least on Earth we must admit that Cartwright's alternative metaphysical picture deserves serious consideration. It is, she claims, the picture best supported by the way science achieves its successes (where it does) and by the limitations and failures of science that we tend too often to ignore or set aside.

For fundamentalist philosophers of science, the whole idea of the patchwork of law-governed and non-law-governed domains can be hard to comprehend. How are we to understand the laws' failing to hold in some places and contexts when they hold so perfectly and precisely in others? Is not a He atom a He atom, whether in a gas spectrometer or a child's balloon? If its stability is explained by the solutions of the Schrödinger equation in isolation, can that stability have some different explanation in nonisolated contexts?

The keys to understanding how Cartwright's dappled world functions lie in the concept of a *nature* (developed in NCATM and invoked also in DW), and in the notion of a *nomological machine*. Things—both structured systems, such as a gas in a laser cavity, and putatively simple things, such as an electron—should be viewed as behaving the way that they do because of their natures, in an Aristotelian sense, or their causal *capacities*. Helium atoms are stable in isolation but also in both spectrometers and balloons, because it is in the nature of protons, electrons, and neutrons to be able to bind together stably in certain configurations, one of them being the 2-2-2 combination of a He atom. Now as it happens, we can show (perhaps only approximately) that the Schrödinger equation is satisfied by an isolated He atom. What this shows is that, in a highly artificial and idealized situation, if we massage the mathematics just right (and perhaps add a few winks and squints), we can say that He atoms satisfy the Schrödinger equation. Which is to say, the combination of the causal powers or natures of the component parts of the atom is such that, under these constraints and if we squint at things just right, we can see them instantiating a simple and precise mathematical regularity. But this shows little or nothing about whether they instantiate some precise mathematical regularity in *all* contexts, much less that there is *one* regularity, one law that they instantiate in all contexts.

Things carry their natures and capacities with them all over the place, but the results of the interactions of many different natures and capacities will not, presumably, always be describable with a neat mathematical law. Such cases should be regarded as the exception rather than the rule, and Cartwright gives us a name for the contexts where a regularity does result: *nomological machine*. A nomological machine is ". . . a fixed (enough) arrangement of components, or factors, with stable (enough) capacities that in the right sort of stable (enough) environment will, with repeated operation, give rise to the kind of regular behaviour that we represent in our scientific laws" (Cartwright 1999: 50). Many nomological machines are, as the name suggests, man-made: pendulum clocks, lasers, bubble chambers. Other instances of the stable-enough and shielded-enough arrangements occur naturally: the planets instantiating Newton's or Kepler's or Einstein's laws or a population of animals instantiating the Hardy-Weinberg law. The point is that nature's natures and capacities break out into law-respecting behavior only under certain special circumstances; what happens the rest of the time is messier but none the less beautiful (or scientifically tractable) for it.

The above example of a He atom (mine, not Cartwright's) may give the mistaken impression that one can understand the behaviors of all things in a bottom-up, reductionist way (albeit with natures rather than equation-laws at the bottom level). This is certainly not Cartwright's view. She would instead advocate a disunified, pluralistic approach to understanding the natures and behaviors of complex or higher-order systems. Some of the properties of a lasing gas may be traceable to the natures of the protons, electrons, and neutrons that compose the atoms in the gas, but probably not all will be. Some will be best thought of simply as properties of the higher-level kind (Ar-Ne gas mixtures, for example). How much reductionism we should accept, and in what contexts, is again something that Cartwright would say we should judge by looking at what our best science actually does. And it is widely accepted nowadays that, however firm reductionist faith may still be among physical scientists, the successful practices of science argue for a multilevel, pluralistic patchwork of connections between higher and lower levels in the traditional reductionist hierarchy.

Nomological machines often correspond to physical situations for which our theories provide explicit models. The laws of the theory are "true in the model" (or at least approximately true), and thus true in any real-world system that we find we can apply the model to. But our theories—even those with pretensions to universality, such as QM, QED, or General Relativity—simply do not tell us how to make a model to fit each and every situation that occurs in reality. Fundamentalists still have faith that the true, final theory will break this pattern and be (at least, *pace* mathematical complexity) visibly applicable to all situations. Cartwright thinks that the history of physics should push us by induction to the opposite conclusion: *No* mathematical physical theory contains prescriptions for arbitrary situations, and, apparently, the stronger the claim to universality becomes (e.g., super-string theory, or the Standard Model of particle physics), the less applicable the theory seems to be to ordinary real situations.

I have tried to sketch the reasons for Cartwright's rejection of fundamentalism and the alternative metaphysical picture of the dappled world that she offers in its place. But in doing so, I have perhaps created another misleading impression. For Cartwright is not so much concerned to convince us that her alternative picture is *true*, as she is concerned to make us see that

(a) *it is better supported, empirically, than the traditional fundamentalist picture*, and (b) *if we embrace the dappled world picture we may make better choices about how to do science in order to improve the world.*

In DW Cartwright announces that her main motivation for studying science is that of the social engineer: the desire to see science put to good use in the improvement of society and the lives of its people. And Cartwright is concerned that for those goals, the fundamentalism and imperialism characteristic of much physics and economics may be a bad prescription. As genetics has become the hot avenue for cancer research, implicitly supported by an easy-to-believe, but false, gene-fundamentalism, Cartwright fears that public spending on more effective ways of fighting cancer get shunted aside. Or one might equally wonder: how much earlier might nanotechnology and the sciences of exotic new materials have flourished, if science funding agencies had not been spellbound by the promise of the Superconducting Supercollider to help us uncover the true and final particle-physics theory? Or one might ask nowadays whether global warming would have as many skeptics as it does (among *real* scientists), if the mathematical physicists trying to model weather did not assure us that everything weather-related is too hopelessly complex to make predictions more than ten days into the future?

Cartwright argues that if we adapt our aims and our methods to those apt for a dappled world, governed at best by a patchwork of laws, we are likely to make better practical progress and not waste time and money pursuing reductionist/fundamentalist pipe dreams. As a metaphysician with strong fundamentalist faith, I do not agree with all of Cartwright's arguments and conclusions (see "For Fundamentalism", this volume). But as an observer of both science and human nature, I think she is right in thinking that a shake-up in the metaphysical world view we attach to science would be a very good thing. Cartwright deserves thanks and high praise for her attempts to shake the philosophy of science out of its current dogmatic slumbers.

NOTES

1. *How the Laws of Physics Lie* (1983); *Nature's Capacities and Their Measurement* (1989); *The Dappled World: A Study of the Boundaries of Science* (1999).
2. Cartwright has always endorsed a form of entity realism, but to take a skeptical line on physical theories in the early 1980s certainly seemed to be a fairly serious form of scientific antirealism. To the extent that this perception has changed, Cartwright's own work is probably largely responsible.
3. (Cartwright 1983: 26). Condition (iv) is needed to handle problems that would occur if one held fixed causes of E that sometimes are intermediate steps between C and E.
4. See Pearl (2000) and Spirtes, Glymour, and Scheines (1993, 2001).
5. One aspect of these worries that Cartwright highlights, the fact that there may not *be* probabilities for some of the variables needed in a true DAG, is also emphasized in Hoefer (2005).

REFERENCES

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Clarendon Press.

———. (1989) *Nature's Capacities and Their Measurement*, Oxford: Clarendon Press.

Hoefer, C. (2005) "Humean Effective Strategies", In Petr Hajek, Luis Valdes-Villanueva, Dag Westerstahl (eds.), *Logic, Methodology and Philosophy of Science: Proceedings of the Twelfth International Congress*, London: KCL Publications.

———. (1999) *The Dappled World*, Cambridge: Cambridge University Press, p. 1.

Pearl, J. (2000) *Causality*, Cambridge: Cambridge University Press.

Spirtes, P., C. Glymour, and R. Scheines. (1993) *Causation, Prediction, and Search*, New York: Springer-Verlag.

Part I

Models and Representations

2 Standing Up Against Tradition

Models and Theories in Nancy Cartwright’s Philosophy of Science

Daniela Bailer-Jones

INTRODUCTION

Tradition has it that theories are carriers of knowledge telling us what the empirical world is like. Scientific models are thought to be of little consequence in this context. In the early days of twentieth century philosophy of science, it was common to consider models merely as hypothetical and as heuristic devices (famously, Duhem [1914] 1954; also Carnap 1939: 69). When models became a “hotter” topic in the 1960s, their role was delineated as explanatory (Harré 1960; Achinstein 1968) and, less frequently, as creative (Hesse 1966). Those who preferred formal approaches adopted the mathematical model-theory as a guideline, resulting in the so-called “semantic view” of theories (Suppes 1961 and others). These traditions of interpreting models have repercussions and followers to this day (Bailer-Jones 1999).

A number of recent contributions to the philosophy of science, important among them Nancy Cartwright’s, suggest, however, that models, not theories, are the carriers of knowledge about the empirical world. This is an intriguing claim which itself raises two important questions.

1. How do we want to decide the question concerning the dominance of models or of theory in the project of describing phenomena of the empirical world?
2. How can we usefully distinguish between models and theories?

Cartwright has contributed proposals in response to both these questions, but there remain questions, some of which I shall highlight and address in this chapter. I am far from fundamentally disagreeing with Cartwright, though her writings on models are a bit of a patchwork in that they come to the topic from different angles and with different approaches, all ultimately belonging together, while it is not always easy to envisage the entire design in the process. It is my aim in this chapter to join some patches together and to point out how they do or, in some instances, do not fit together. I provide some of the desired elaboration with the goal of strengthening a promising model-theory distinction.

To provide some orientation, let me give a provisional outline of what I take models and theories to be. A model is an interpretative description of a phenomenon. I use the term “phenomenon” very much in the tradition of Bogen and Woodward (Bogen & Woodward 1988). A phenomenon is a fact or event in nature, such as bees dancing, rain falling, or stars radiating light; it is something that is taken to be a subject deserving further research. Interpretative descriptions may rely, for instance, on idealisations or simplifications or on analogies to interpretative descriptions of other phenomena. A model focuses on specific aspects of a phenomenon, sometimes deliberately disregarding others. As a result, models tend to be partial descriptions only. Models can range from being physical objects, such as a toy aeroplane,[1] to being theoretical, abstract entities, such as the Standard Model of the structure of matter and its fundamental particles. The majority of scientific models are, however, a far cry from consisting of anything material like the rods and balls of molecular models used for teaching; they are highly theoretical. Saying that models are theoretical means that they strongly rely on theories for their construction.[2] The means by which scientific models are expressed range from sketches and diagrams to ordinary text, graphs, and mathematical equations—to name just some—all of which serve as description of the phenomenon in question in one way or another. Theories, in turn, are not about the empirical world in the same concrete sense as models. Theories are not formulated specifically for a phenomenon but are expected to be applicable in modelling a whole range of different phenomena. In that sense, they are much more general than most models. Models, by their very constitution, are applied to concrete empirical phenomena, whereas theories are not. Theories, in turn, have the *capacity* of being applied to empirical phenomena when specific constraints belonging to a concrete case are inserted into the more abstract theory. Theory needs to be customized for its use in modelling a specific empirical phenomenon. Classical examples of theories are Newton’s laws or Maxwell’s equations.

In the section ‘Models in Cartwright’s philosophy’, I retrace Cartwright’s understanding of scientific models in contrast to theories. It is notable that her view on models underwent some development over the years. I will highlight some of the changes in order to facilitate understanding Cartwright’s position. Then, in the section ‘The dappled world and scientific practice’, I address the general point that Cartwright’s philosophical claims about science, models, theories, and the world strongly depend on the study of scientific practice. I spell out some difficulties of this approach as a methodology in philosophy. The dilemma of this approach is a reason for Cartwright’s philosophy being both productively provocative and difficult to grasp. Then, in the section ‘Can theories be true?’, I consider some issues arising from Cartwright’s terminology with regard to the truth of models and suggest how to make talk about truth, if one chooses to indulge in it, more coherent. In the section ‘Theories and the empirical world’, I want to endorse the distinction between models and theories, with models being about concrete phenomena and theories not. In this I build on Cartwright's foundation and contribute my explication of the sense in which phenomena that are modelled are concrete and in which theories are not theories of such concrete phenomena. The final section contains the conclusions.

MODELS IN CARTWRIGHT'S PHILOSOPHY

Models are a recurring topic in Cartwright's work. They have their role to play in the larger picture of Cartwright's philosophy of science; they are one component in the explication of how science works. The following sections show how Cartwright's understanding of models has evolved over the years and how she sketches the relationship between models and theory. There are a number of issues arising from her position(s) which I will follow up in the last few sections.

Models as Fictions

The first manifestation of Cartwright's views on scientific models is her simulacrum account of explanation (Cartwright 1983). According to this account, explanatory power does not count as an argument for the factual truth of theories or models. A model may explain a phenomenon and yet not have any claim to truth in virtue of this. Instead,

> [t]o explain a phenomenon is to find a model that fits it into the basic framework of the theory and that allows us to derive analogues for the messy and complicated phenomenological laws which are true of it. (Cartwright 1983: 152)

As we have learned, laws of nature can lie, thus the theoretical framework on which the model may be based does not warrant the truth of the model. Models are prepared specifically that fundamental laws can feature in them: 'For the kind of antecedent situations that fall under the fundamental laws are generally fictional situations of a model, prepared for the needs of the theory, and not the blousy situations of reality' (Cartwright 1983: 160).

Thus laws do not literally apply to the real situations (Cartwright 1983: 161). It is only phenomenological laws that can be true of phenomena. Phenomenological laws are laws as they are observed in phenomena; they are not integrated into a background of theories; they are formulated ad hoc.[3] Then, a model—which is based on certain theoretical laws—is merely an analogue to the messy phenomenological laws. Furthermore, different models have different purposes, with the models having different emphases depending on their purpose (Cartwright 1983: 152). This sheds some further doubts on models being realistic: 'We should not be misled into thinking that the most realistic model will serve all purposes best' (Cartwright 1983: 152).

Honouring such antirealistic tendencies of models, Cartwright proposes a "simulacrum account of explanation". "Simulacrum" is defined, in accordance with the *Oxford English Dictionary*, as 'something having merely the form or appearance of a certain thing, without possessing its substance or proper qualities' (Cartwright 1983: 152–153).

Things are "not literally" what their models say they are. Cartwright, thus, goes on to claim that '[a] model is a work of fiction. Some properties ascribed to objects in the model will be genuine properties of the objects modelled, but others will be merely properties of convenience' (Cartwright 1983: 153).

The aim of such "properties of convenience" is 'to bring the objects modelled into the range of the mathematical theory' (Cartwright 1983: 153). Models render theories, here assumed to be mathematical in character, applicable to phenomena, albeit with models having fictional status. This fictional status presumably has to do with the limitations of theories as not telling us how things "really" are:

> I think that a model—a specially prepared, usually fictional description of the system under study—is employed whenever a mathematical theory is applied to reality, and I use the word "model" deliberately to suggest the failure of exact correspondence [. . .]. (Cartwright 1983: 158–159)

So, models, on the one hand, fail to have "exact correspondence" to the phenomena they represent, and on the other hand, they are needed for theories to establish some kind of relationship to reality: '[O]n the simulacrum account, models are essential to theory. Without them there is just abstract mathematical structure, formulae with holes in them, bearing no relation to reality' (Cartwright 1983: 159).[4] This characterisation of theories as (a) abstract mathematical structure, (b) formulae with holes in them, and as (c) not bearing any relation to reality is the core idea which will provide the skeleton for the characterisation of theory later in this paper. For such a characterisation, we will have to consider what "abstract" means in this context, what the "holes" in the formulae of the theories are, and how the relationship of theories (or models) to reality can best be understood.

Models as Fables

In an article some years later, (Cartwright 1991) compares scientific models to fables. This is not about models being fictions. It is about the contrast between the abstract and the concrete. Fables have a moral which is abstract and they tell a concrete story that instantiates or "fits out" that moral. A moral of a fable may be "the weaker is prey to the stronger", and a way to "fit out" (Cartwright's formulation) this abstract claim is to tell the story of concrete events of the marten eating the grouse, the fox throttling the marten, and so on. Similarly, an abstract physical law, such as Newton's force law, $F = ma$, can be fitted out by different more concrete situations: a block being pulled by a rope across a flat surface, the displacement of a spring from the equilibrium position, the gravitational attraction between two masses. Thus, Newton's law may be fitted out by "different stories of concrete events". Drawing from the analogy between models and fables, models are about concrete things; they are about concrete empirical phenomena. The contrast between models and theories is not that theories are abstract and models are concrete. Rather, models are about concrete phenomena, whereas theories are not about concrete phenomena. If at all, theories are about concrete phenomena only in a very derivative sense. A second claim, beyond the distinction between the abstract and the concrete, has to do with "existence". "Force", which is an abstract notion, does not manifest itself outside concrete empirical situations. Force is a factor in and contributor to empirical phenomena. Cartwright's everyday example for this relationship is "work": The abstract concept of "work" may be filled out by washing the dishes and writing a grant proposal, and this does not mean that a person washed the dishes and wrote a grant proposal, *and* worked—working does not constitute a separate activity—since working consists in just those activities. Cartwright explains:

> *Force*—and various other abstract physics' terms as well—is not a concrete term in the way that a color predicate is. It is, rather, abstract, on the model of *working*, or *being weaker than*; and to say that it is abstract is to point out that it always piggy-backs on more concrete descriptions. In the case of *force*, the more concrete descriptions are ones that use the traditional mechanical concepts, such as *position*, *extension*, *motion*, and *mass*. Force then, on my account, is abstract relative to mechanics; and being abstract, it can only exist in particular mechanical models. (Cartwright 1991: 65)[5]

Cartwright then infers that 'laws are true in the models, literally and perhaps precisely true, just as morals are true in their corresponding fables' (Cartwright 1991: 68). However, this is not to say that the models need to be true of the world, just as the fables may not be true of the world. An abstract concept, such as "force", may not be suitable to modelling *all* aspects of the world, perhaps only certain ones which are carefully constructed (in the laboratory, under exclusion of various other factors). It is for those situations where the abstract concept of force can be applied in a model that, according to Cartwright and in her terms, Newton's law is true of the model.

As I spell out in the section 'Can theories be true?', I take models to be about the world, so the important relation is not one of theories being true of models, as Cartwright sometimes suggests, but one about models being true (or something like that) of the world. In the penultimate section, I will endorse the suggestion that theories are abstract and that models are about concrete phenomena of the empirical world. Making this distinction obviously depends rather a lot on what "abstract" and "concrete" are taken to mean, and I shall examine this below.

The Tool Box of Science

In (Cartwright et al. 1995), Cartwright revises her position on theories and models from her earlier statements. There she criticises what she calls the "theory-dominated" view of science and sees herself as part of the 'movement to undermine the domination of theory' (Cartwright et al. 1995: 138). It is models, rather than theories, that represent phenomena of the physical world (Cartwright et al. 1995: 139). Theories, in turn, are but one of the tools used in model construction. Other such tools are, for instance, scientific instruments or mathematical techniques. The change from her earlier views is that not only do theories no longer represent the world via models; they do not represent it at all. Correspondingly, Cartwright states:

> I want to urge that fundamental theory represents nothing and there is nothing for it to represent. There are only real things and the real ways they behave. And these are represented by models, models constructed with the aid of all the knowledge and techniques and tricks and devices we have. Theory plays its own small important role here. But it is a tool like any other; and you can not build a house with a hammer alone. (Cartwright et al. 1995: 140)

Cartwright's general claims are illustrated by an example which her coauthors Towfic Shomar and Mauricio Suárez elaborate. The example is the 1934 model of superconductivity developed by Fritz and Heinz London, and the claim is that this model did not develop via the theory-driven strategies of approximation and idealisation. A classic example of theory-driven modelling would be gradually modifying an equation to make it more realistic by adding correction terms, e.g., when adding a linear term for mechanical friction to the equation of the simple harmonic oscillator, resulting in an equation for a damped linear oscillator. The superconductivity example is a case in point that not all scientific modelling is a process of de-idealisation. Instead, there can be what may seem to be ad hoc adjustments to the theory that are not theory-driven but phenomenological. In sum, the argument is that there exists phenomenological model building in science that is perfectly valid yet independent of theory in its methods and aims (Cartwright et al. 1995: 148).[6]

This scenario about the relation between theories and models is the extreme point of the demotion of theory: Theory does not represent anything; it is but a tool in model construction. As I spell out below, I think that this conclusion is already a direct consequence of the fables account.

Models in a Dappled World

In accordance with Cartwright's earlier views that the laws of physics lie, it also holds for the dappled world that not everything that happens in the empirical world can be captured by the laws of physics: only those things for which there are models that match them. Cartwright vividly illustrates this point with the example of a thousand dollar bill swept around in St. Stephen's Square (Cartwright 1998: 28). There exists no model of classical mechanics that is capable of describing this complex physical situation. In Cartwright's terms this means that classical mechanics is not universally applicable (in principle), and it means that the laws of mechanics do not determine this particular process. Instead, it may be necessary to switch to another area of physics for a description, e.g., fluid dynamics. It may then be possible to have a model based on fluid dynamics that nearly enough captures what is going on with the thousand dollar bill. The point is again that any theory applies to the world only through its models:

> Fluid dynamics can be both genuinely different from and genuinely irreducible to Newtonian mechanics. Yet both can be true at once because—to put it crudely—both are true only in systems sufficiently like their models, and their models are very different. (Cartwright 1998: 29)

In the same way, quantum mechanics does not replace classical mechanics. Both theories make good predictions in certain real world situations and are frequently employed in cooperation (Cartwright 1998: 29). On Cartwright's account, the world is dappled, which is to say that '[n]ature is not reductive and single minded. She has a rich, and diverse, tolerant imagination and is happily running both classical and quantum mechanics side by side' (Cartwright 1998: 30).

Turning the whole argument around, it is not only the case that laws apply within the limited range of a model only but also that models can serve as blueprints for "nomological machines" which provide the basis for arriving at a law (Cartwright 1997; 1999b: Ch. 3). Models tell us under which specific circumstances certain laws arise, in opposition to a Humean regularity view of laws that portrays laws as universal (Cartwright 1997: 293). So, what is a nomological machine?

> It is a fixed (enough) arrangement of components, or factors, with stable (enough) capacities that in the right sort of stable (enough) environment will, with repeated operation, give rise to the kind of regular behaviour that we represent in our scientific laws. (Cartwright 1999b: 50)

Laws can be formulated under the specialized conditions created by the nomological machine. These conditions are mostly achieved by "shielding", that is by controlling the input into the machine such that anything is prevented from operating which might interfere with the machine functioning as prescribed. The result is *ceteris paribus* laws for the specific situation. Even probabilistic laws can be developed by means of nomological machines, so-called “chance set-ups”. In short, nomological machines produce the orderly and lawful outcome that is so much in contrast to the real, dappled world of the thousand dollar bill on St. Stephen’s Square.

Representative and Interpretative Models

The limits of theory continue to be a topic for Cartwright (Cartwright 1999a; reprinted in 1999b) in her paper on the BCS model of superconductivity, where she puts her modelling view in the historical context of theory dominance in philosophy of science. Cartwright not only rejects the received view of scientific theories as axiom systems in formal languages because they lack expressive power (Cartwright 1999a: 241), she also discounts the semantic view of theories which considers models as constitutive of theories (Cartwright 1999a: 241). Instead, she adopts the view of models as mediators between theory and the real world (Cartwright 1999a: 242; Morrison & Morgan 1999). The models that mediate between theory and world are *representative models* (formerly phenomenological models; Cartwright 1999a: 242). They represent the world not by being part of a theory (in contrast to *interpretative models*), although they may draw from theories. Cartwright takes representative models to be ‘models that we construct with the aid of theory to represent real arrangements and affairs that take place in the world—or could do so under the right circumstances’ (Cartwright 1999a: 242; 1999b: 180).

Representative models can represent specific situations and to do so they may go well beyond theory in the way they are built. This means that theory is not the only tool for model construction; others are scientific instruments, mathematical techniques, or the kind of laboratories, just as proposed in the toolbox approach. Thinking of theories, in turn, as abstract is already familiar from the notion of models as fables:

> I want to argue that the fundamental principles of theories in physics do not represent what happens; rather, the theory gives purely abstract relations between abstract concepts: it tells us the “capacities” or “tendencies” of systems that fall under these concepts. No specific kind of behaviour is fixed until those systems are located in very specific kinds of situations. (Cartwright 1999a: 242)

Interpretative models, in turn, are models that are ‘laid out within the theory itself’ (Cartwright 1999a: 243). Via bridge principles the abstract terms of a theory can be made more concrete in an interpretative model. Interpretative models establish a link between abstract theory and model, whereas representative models establish the link between model and world

(Cartwright 1999a: 262). So it would seem that representative models can be, but do not have to be, interpretative models. This is so insofar as interpretative models make abstract notions that feature in theories more concrete, and in that sense they can serve to represent certain situations that fall under the theory. Interpretative models have the function of representing certain theoretical situations, and these may or may not be similar to real situations. Representative models, in turn, need not and do not have this interpretative function to "fit out" theories.[7]

One problem with the notion of representative models is that Cartwright does not elaborate the concept of representation which she uses to say that theories do not represent the world and that representative models do (as in Cartwright et al. 1995). She does not want representation to be thought of as structural isomorphism (Cartwright 1999a: 261). According to Cartwright, the notion needs to be broader than one 'based on some simple idea of picturing' (Cartwright 1999a: 262). It is a 'loose notion of resemblance' that is instead suggested (Cartwright 1999a: 262). As Cartwright herself acknowledges, this is not much more than pointing to the problem of representation.

THE DAPPLED WORLD AND SCIENTIFIC PRACTICE

It is not enough to lay out Cartwright's claims about models and theories, however. A vital part of her approach to philosophy of science is the methodology with which to arrive at a position in philosophy of science. Philosophy of science, for her, is also about the methodology by which to arrive at certain claims. Cartwright's creed is that of studying the scientific practice. Correspondingly, the dappled world, as Cartwright draws it up, is supposed to be as close to the real world as it can be. However, choosing an approach to philosophy from scientific practice raises a number of more general issues about how we should do philosophy of science. Cartwright, with her dappled world, sees herself as moving, broadly, in a realist framework:

> I take seriously the realists' insistence that where we can use our science to make very precise predictions or to engineer very unnatural outcomes, there must be "something right" about the claims and practices we employ. (Cartwright 1999b: 9)

Although very carefully phrased, this statement of realism still has the character of a general claim about the science we do and its relation to "the world". Cartwright admits that there is no reason per se to believe in a dappled world rather than a world well structured in which unification of theories reigns. While the fact that theories only give us small snippets of accurate representation of the world constitutes a good reason to think of the world as "objectively dappled", Cartwright acknowledges that the evidence is not conclusive. This is probably why Cartwright introduces a pragmatic argument, namely the issue that the beliefs about the structure of the world can influence the methodologies that are adopted to study this world (Cartwright 1999b: 12). The rationale is the following. Believing that one theory is the one and only approach to a certain issue makes it likely for scientists to overlook and disregard other approaches, based on alternative laws and assumptions. This narrow-mindedness can have devastating practical (social, medical . . .) consequences (Cartwright 1999b: 16; see also Teller, this volume). Cartwright simply suspects that a methodology geared towards unification is not the best methodology to capture what is going on in the world and to solve the problems people confront in the world (Cartwright 1999b: 13).

So the argument for Cartwright's particular stance towards the world runs along the following lines: Scientific practice is such that successful models of situations in the empirical world draw from a whole range of sometimes competing scientific theories. The success of this practice is such that it also convinces us that proceeding in this way is the methodology to be preferred when doing science. Consequently, the world is more likely to be such as indicated by piecemeal multimodel constructions from theories than uniform and completely describable by fundamental theory. What counts, according to Cartwright, is 'what image of the material world is most consistent with our experience of it' (Cartwright 1999b: 9). Thus, the kind of realism Cartwright proclaims is tied to the facts of scientific practice. Furthermore, according to Cartwright, modelling is such a fact of scientific practice.[8] If not theories, but certain types of models represent the empirical world, then realism must mean that *models* tell us what the world is like, and not theories, as was traditionally thought. However, what models tell us about the world is not always easy to accept in a realistic framework. The issue of mixed modelling, or multiple models of one and the same phenomenon, notoriously causes headaches with regard to such scientific realism. The puzzle is: how can we tell what the world is really like if there are different "stories", in the form of different models, available? Cartwright would say we will simply have to accept all the stories, individually taking into account what the story is about and in view of the *ceteribus paribus* conditions that apply. Well, although all-too-common practice, this seems hard to swallow for an epistemology. Yet this is where we stand in studying the scientific practice. Interestingly, there is considerable resistance to taking "the easy way out" of becoming instrumentalist and antirealist about the whole project of science.[9] If Cartwright has it her way, then mixed modelling implies that we can be realist with regard to *all* models that are in use. This is also an intuition behind some of the recent literature on models as "representing" (Bailer-Jones 2003).

For me, a more general issue that lies behind these considerations, besides the obvious difficulties, is what to do with scientific practice in philosophy of science. What is the status of arguments from scientific practice? The main dilemma of any claims about the world based on case studies is how the case studies are chosen (Pitt 2001). There is, after all, likely to be a bias in even selecting the cases such that they support a certain view. There is probably a selection effect depending on the philosophical point one wants to make. Even if this were not so, the number of cases looked at is always going to be limited, and it is not clear how such a limited number of cases would warrant some general conclusion. The assumption behind this criticism is that philosophical claims are, almost by their nature, general. So, while Cartwright may admit that the world may be dappled *or* may be unified, she still wants to make the suggestion that, based on scientific practice "generally", it makes more sense to view the world as dappled. This kind of suggestion will always leave copious space for the sceptic: Is the world really dappled for everybody, and would it be so if one considered all the examples? Sometimes the world may look more like this, and sometimes more like that. It is interesting that, although we can never ultimately dispel the sceptic's argument, we constantly seem to try and generalize and put into a pattern what we find. This is what we seem to be set up to do as philosophers, and this is just what Cartwright can be taken to do when she proclaims a dappled world on the basis of scientists' modelling experience. This is not to suggest that Cartwright draws hasty conclusions from a biased sample of examples. Yet, no matter how good and how representative examples are, they are only examples and not a complete set of cases, rather like in the problem of induction.

Another way to avoid the dilemma of how to do philosophy of science is to acknowledge that there is no one way science is (Burian 2001).[10] It is only possible to make generalisations within certain limited contexts, just as Cartwright suggests with regard to the relationship between models, theories, and the world. In other words, there simply is no general pattern to be found in the world. This also means that the whole scope of the project of philosophy is more limited. Burian correspondingly talks about a 'reduction in the ambitions of that discipline' (Burian 2001: 401). However, if we go to the extreme of never daring a generalisation of what we find out there in the world, then philosophy loses its subject. It is part of thinking about the world to try out different patterns and generalisations that may fit that world, so what is needed is a methodology that makes case studies profitable for philosophical purposes (Pinnick & Gale 2000). What case studies give us is more detailed knowledge than we would have if we stayed at the quite general level. This is why case studies in science can prevent us from grave errors in our interpretation of science. If we then go beyond the case in our philosophizing, we must become speculative. But to the same degree to which we become speculative (and philosophically interesting) we lose the safety of the empirical foundation provided by the case study, precisely because it was a case study only. Cartwright's claim that the world is dappled is strangely at odds with the impossibility of concluding anything definitive about the world on the basis of case studies. Perhaps Cartwright is really not making any very big claims about the way the world is, but then this is strangely at odds with what philosophy is set up to do. Perhaps this is one source of confusion that sometimes arises when trying to interpret Cartwright's philosophy. It is certainly a confusion that I do not know how to resolve easily. Let me therefore return to more formal arguments about Cartwright's position concerning the truth of models and theory. The problems discussed in the next section occur precisely because scientific practice is taken seriously despite the oddities and potential logical inconsistencies arising from it.

CAN THEORIES BE TRUE?

According to Cartwright, it is common practice that models may be constructed using competing theories, depending on the situation to be modelled, and they may even combine competing theories in one model. The paradigmatic example for this issue is the use of classical and quantum mechanics: 'In the right kind of situations some systems have quantum states, some have classical states, and some have both' (Cartwright 1999b: 216). A model may be pieced together from suitable components, as seems useful and promising (Cartwright 1999b: 223). If it really is true that a system has both classical and quantum states, this could mean that a certain particle both does and does not have a determinable location or that it is both wave and particle. But how can such different and even competing properties in models coexist? The answer seems startling:

> Let us grant that quantum mechanics is a correct theory and that its state functions provide true descriptions. That does not imply that classical state ascriptions must be false. Both kinds of descriptions can be true at once and of the same system. (Cartwright 1999b: 231)

So, in the extreme this could mean that it is both true that a particle is located in a certain position and that it is not. Logically, however, it is hard to conceive how both A and non-A are supposed to be true because this would result in a contradiction.

The same kind of problem arises when there are multiple models of different aspects of one and the same phenomenon—a topic which Cartwright does not pick out as explicitly as the clash between quantum and classical mechanics in one model. Multiple models are, however, a common occurrence and perhaps even a crucial feature of science as currently practised. The kind of problem arising from such multiple modelling also fits into Cartwright's general framework, I think. Rather than confronting different phenomena for the description of which different theoretical components are employed, one confronts different descriptions of one and the same phenomenon, such as in the different models of the atomic nucleus (Morrison, 1998: 74). One and the same phenomenon may give rise to different models, depending on how the phenomenon is experimented upon, or with which procedure it is examined or observed. Just think of the models of water in Paul Teller's favourite example: If one considers water flow in a pipe, water is treated as an incompressible continuous medium, while considering the diffusion, e.g., of a drop of ink in water, water is treated as consisting of discrete particles in random thermal motion (Teller 2001: 401). Cartwright would indeed probably say that these are different phenomena (water flowing in a pipe and a drop of ink diffusing in water), or at the very least different aspects of a phenomenon (the behaviour of water), therefore requiring different theories in their appropriate models.

My difficulty with 'assigning two different kinds of descriptions to the same system and counting both true' (Cartwright 1999b: 232) is what "truth" is supposed to mean in this context. One obvious strategy would be to argue that Cartwright should not talk about truth in the context of models. Perhaps she should not, and depending on one's concept of truth, this is certainly a defendable strategy to take. On the other hand, Cartwright confronts a tradition that deals in the currency of truth. This is why she formulates her position with reference to this vocabulary. This is also why I go along and consider the implications of talk about truth in the context of scientific modelling.

Truth is something that can be attributed to propositions, and a proposition counts as true in those cases in which things are in the world as the proposition states. Truth of propositions is then nontrivially interesting if these propositions are about the empirical world.[11] In this sense "the earth is the smallest planet around the sun" is false while "the earth orbits around the sun" is true. If there exist two descriptions of one system that are both true, then this may be because the descriptions are independent of each other in that they are about different parts of the system. In other words, the descriptions may be about the same object yet about different aspects or characteristics of the common object of description. It is also possible that the descriptions are about apparently very different phenomena—e.g., a highly accelerated system versus one at rest or a microscopic versus a macroscopic system. If two descriptions are about the same part, aspect, and/or characteristic of a system and they are both true, this may mean that they are equivalent and merely appear to be expressed differently. If the two descriptions are about the same part, aspect, and/or characteristic of the system and yet different, then there is a reasonable chance that certain inconsistencies occur, at least from a logical standpoint.[12] Such inconsistencies do occur in scientific practice. If classical mechanics and quantum theory are *both* employed in modelling SQUIDs (Superconducting Quantum Interference Devices), then this raises the issue of consistency, no matter how successful the resulting model is because classical mechanics and quantum theory are based on different principle. This comes down to the classical puzzles: do particles have a precise location?, etc. Of course, it may be perfectly all right not to be fully consistent in modelling in practice. Successful and useful models are not necessarily in the same category as true models (assuming that such true models even exist and that we know what truth is supposed to mean when said of models). These are instances where taking the cues from scientific practice, as Cartwright rightly requires, confronts us with philosophical puzzles.

Another problem of talking of a model as true is, of course, that a model may tell us a whole range of things about the world some of which may be true, others not.[13] How can one decide about the overall truth of a model? Let me here play through a potential, if fictional, way of how one could interpret talk of truth of models. I do this not because I think it particularly useful to talk of models in terms of truth but because the truth of models has been philosophically discussed. If a model tells us some things about a phenomenon that may count as true and others that do not, then one sometimes has to decide whether the model as a whole can count as true. If there are a number of competing models, then one model may somehow be "more true" than another, e.g., *entailing* more true propositions. I do not mean logical entailment here, where the truth of certain propositions can be deduced from the truth of certain other propositions. To capture the content or the "message" of a model, a range of different means of expression can be employed.[14] Such means of expression can be texts, diagrams, mathematical equations, etc. The idea of entailment is that at least some of the "message" or content of the model can be expressed in terms of propositions, even if it is, in the model, expressed by nonpropositional means. The propositions thus entailed by the model state what the model is taken to state about the phenomenon modelled, and any model entails many different propositions. The overall truth of a model would then somehow consist in the model entailing many true propositions. Of course, it makes no sense to determine a fraction of propositions that need to be true in order for the model to count as true. Rather, if one wants to talk of the truth of models, one would have to say that certain propositions are more central for the "message" captured by the model and that it is therefore more important for these propositions to be true than for others, considering the overall truth of the model (Bailer-Jones 2003). Correspondingly, the majority of models would only ever achieve the predicate "roughly true", which is obviously not the same as true. Notice that the claim is not that models are necessarily propositional. Instead, this is about how one can interpret talk of truth with regard to models. Correspondingly, the assumption is that models, whatever they are, communicate a "message" about phenomena and that some of this at least can be expressed in terms of propositions. Thus the propositions *are not* the model, but they are *entailed* by the model which means that they can be employed to communicate its content.

There is yet another problem with regard to truth. This time it concerns theories. Cartwright says that 'theories are true only of their models and, at best, of real systems that resemble them [the models] closely enough'

(Cartwright 1998: 33–34), and one has to ask what it means for a theory to be true of a model. This can be taken to mean that, in a set-theoretic sense, the model satisfies the theory, which Cartwright may or may not have in mind. However, I would like to reserve talk of truth for propositions about facts or things that belong to the empirical world. Of course, theories are sometimes employed in modelling phenomena. This means that the theories can be useful and appropriate for a certain modelling approach to a phenomenon, and this is why elements of these theories enter into the model. It is then not surprising that a theory is satisfied by a model, given that the theory has been selected to be used in the construction of the model. This is no basis, however, for suggesting that the theory is true of the phenomenon. The model satisfies the theory, and the model is about a phenomenon, but this does not require us to think of the theory of being about that phenomenon. In any case, theories are not directly about phenomena, so they would be true of phenomena only in some mediated (mediated through the model) sense.

I recommend reserving the terminology of truth and falsity for talking about the world, which is why it makes little sense to me to state that theories can be true of models. Models, in turn, may be interpreted as true in a certain sense. Notice that the fables account leads to an ambiguity at this point. Whereas I claim that models are or aim to be at least roughly true of empirical phenomena, fables need not be true of the empirical world. The story told in the fable has never happened, although the implication is that the story still has a lot to do with what the empirical world is like. This makes me think that fables are like interpretative models, whereas I focus on what Cartwright calls representative models. These are models in the context of which one *can* talk of truth to the extent that they are about the empirical world. If they are called true, then it is in virtue of being about the empirical world. Cartwright, in turn, still often considers models as being about some constructed situation (like a fable), which is also why she portrays models as nomological machines.

THEORIES AND THE EMPIRICAL WORLD

Theories are not the kind of statements about which we can find out whether things are in the empirical world as the theory states. Let me go back to the fables to illustrate this point. The moral of a fable, such as "the weaker is prey to the stronger", cannot, in this form, be tested. There may be individual cases for which this moral or "theory" works, such as when the marten eats the grouse, and the fox throttles the marten. In a dappled world there may be *many* instances where the weaker is prey to the stronger, but it is exactly the *universality* with regard to which the moral or "theory" needs to be doubted, if we really live in a dappled world. There is absolutely no reason not to think that there could be other morals, or other theories, that provide a suitable base for a model. Just think of the race between the hedgehog and the hare. There it is not the weaker, or in this case the slower, who loses out. The moral there is that factors other than physical strength can play a role in winning a race. Just as Cartwright asserts for the dappled world, both models have their justification. The world can be like that in the fable of the hedgehog and the hare or like that in the fable of the grouse, the marten, and the fox, hence like both fables. This simply depends on the individual situation that is in question. And indeed, as Cartwright claims elsewhere, theories can form "partnerships" where different theories are applied in different models when real phenomena cross over different areas of physics (Cartwright 1998: 34). The world offers a wide range of instances and cases only some of which fall comfortably under one or the other moral. There are lessons about the world in "the weaker being prey to the stronger", as well as in "winning over sheer physical aptitude by means of wit". In view of the fables analogy and having to apply different morals to different empirical situations, it perhaps becomes easier to see how theories do not tell us anything *directly* about the world.

Cartwright characterizes theories as abstract in the context of the analogies between models and fables ("theories are like morals of fables"). Being abstract is related to why theories cannot be true, as discussed in the last section, but this still leaves open what "abstract" means. Fearing that I cannot come up with a satisfactory answer to what abstractness is, I will concentrate on one aspect only of theories being abstract. This aspect is that theories, being abstract, are not directly about empirical phenomena.[15] Abstractness is opposite to concreteness. The phenomena that are explored by modelling are *concrete* in the sense that they are (or have to do with) real things—things such as stars, genes, electrons, chemical substances, and so on. Of most phenomena we can find many specimens in the world; these phenomena belong to the same class.[16] Modelling a star, there are many different individual stars that could serve as a prototype.[17] One tries to model, however, not any odd specimen of a phenomenon, but a typical one. Often this involves imagining the object of consideration as having "average" or "typical" properties, and this "prototypical" object or phenomenon may not even exist in the real world. The point is that it could typically exist in just this way and that there exist many very much like it. So, the prototype is selected or "distilled" from a class of objects. The prototype has all the properties of the real phenomenon; it is merely that the properties are selected such that they do not deviate from a "typical" case of the phenomenon. It is then this prototype that is addressed in the modelling effort and that may be subject to idealisation. The assumption behind this process of prototype formation is nonetheless that the model is not only a model of the prototype but one of the real phenomenon, including specimens that display a certain amount of deviation from the norm. Correspondingly, modelling the human brain is not about modelling the brain of a specific person but that, roughly, of all "typical" people. For my purposes, the prototype of a phenomenon still counts as concrete, because it has all the properties of the real phenomenon and *could* exist in just this manner. The target of the examination remains an empirical phenomenon, even if members of the class of that phenomenon can come in different shapes and variants. This prototype-forming procedure is often needed in order to grasp and to define a phenomenon and to highlight what it is that one wants to model. The important point here is that despite prototype formation, the phenomenon is not in any way stripped of any of its properties.

Phenomena have properties. Abstraction I take to be a process where properties are taken away from a phenomenon and are not replaced by another property. That which is abstract lacks certain properties that belong to any real phenomenon that is concrete. To put it very crudely, something concrete becomes abstract when certain properties, which belong to the "real thing" (and that make it concrete), are taken away from it.[18] Giere (this volume) rightly points out that there can be different degrees of abstractness, from fairly abstract models (e.g., the model of the ideal pendulum) to fully specific models which model concrete situations, the kind of situation in which actual measurements could be made, but the definition of abstraction that I give can accommodate for these degrees. Not all concepts, principles, or theories that are called "abstract" are abstract in the same way, but, again, I think the notion of taking away some of those properties that make something concrete can still serve as a guideline. It is important to recognize that no theory is conceivable without the concrete instantiations from which the theory has been abstracted. We need to go through different example problems in order to understand how $F = ma$ is instantiated in different models. The theory is that which has been distilled from several more concrete instantiations. In this sense, the abstract theory is not directly about concrete phenomena in the world. The properties that are missing in such an abstract formulation as $F = ma$ are how the force makes itself noticed in different individual situations. Think again of a block being pulled across a flat surface, or the displacement of a spring from the equilibrium position, or the gravitational attraction between two masses. It depends on the situation that the force or the acceleration consists in (deceleration due to friction, the repulsion of a spring, or acceleration due to gravitation). Moreover, for each *concrete* situation one would have to establish what the body is like whose mass features in the physical system. Correspondingly, force, acceleration, and mass can be associated with different properties in different physical systems. Force, abstractly speaking, can be something that applies to an object or system, but force alone, without an object or a system, is not something about which we can say anything nor know the properties of. To establish a theory we need models that tell us how the theory is relevant with regard to the phenomenon or process modelled.

Finally, let me add a brief note on laws and theories. Some laws have the status of theories, but not all do.[19] There can be laws that are merely generalisations of concrete instances, e.g., "the melting point of lead is 327 degrees

Centigrade", which is presumably true of all lead. This is not abstract. An abstract law would be one that told us, for instance, how to infer the melting point of quite different metals. For a law that simply states the melting point of lead, be it right or wrong, i.e. for phenomenological laws, we do not need a model in order to apply it to the world. Such a law does not apply to a range of different instances from which it is abstracted; such a law applies only to one kind of instance generally. This makes the law not theoretical. Correspondingly, for such a law it involves no great difficulty to resolve the issue of its truth empirically.

Theories can become general because they are abstract; they are free of the properties that are typical of certain individual instances where the theory might apply or the properties that are typical of different prototypes. In order to model a phenomenon, abstract theory needs to be made more concrete, taking into account the specifications of the phenomenon that is modelled and inserting the ramifications and boundary conditions of that phenomenon (or the prototype thereof). To see how the theory holds in a model, we need to fill in the concrete detail that is not part of the theory because, being abstract, the theory has been stripped precisely of those details.

To summarize, theory in science is not that which tells us what the world is like but that to which we resort when we try to describe what the world is like by developing models. This is in overall accordance with Cartwright's position, though I have attempted to elaborate the point further. "Abstract", said of theories, means having been stripped of specific properties of concrete phenomena in order to apply to more and different domains. Models, in turn, are about concrete phenomena that have all the properties that real things have. Theories are applied to real phenomena only via models—by filling in the properties of concrete phenomena. Being abstract and therefore not directly about empirical phenomena does not, however, render theories worthless or unimportant. Theories and models have to prove themselves at different levels: models by matching empirical phenomena and theories by being applicable in models of a whole range of different phenomena (or prototypes thereof).

CONCLUSION

What has changed with regard to how models and theories are viewed? Models have moved to the fore when it comes to expressions of scientific knowledge. This is a development to which Nancy Cartwright significantly contributed and which I recapitulated in the second section, 'Models in Cartwright's philosophy'. The argument for a focus on models derives largely from scientific practice. As discussed in the section 'The dappled world and scientific practice', this type of argument carries its own problems. The advantage of this approach is that what we philosophically model is much closer to what science in the real world is like. The philosophical result, the model-theory distinction which is promoted, is abstract like a theory in science. Correspondingly, this result is unlikely to fit all cases of scientific practice, but it will still illuminate those to which it applies. My considerations concentrated less on the problems one may encounter when representing empirical phenomena by means of models and more on exploring the role that is left in this arrangement for scientific theories. I have argued that theories cannot be true or false because they do not directly apply to phenomena, as elaborated in the section 'Can theories be true?' Scientific theories are relevant for widely different phenomena only if details of concrete phenomena are filled in as part of a modelling process. Indeed, theories only apply to empirical phenomena via models, just as Cartwright proposed. A theory crucially depends on its concrete instantiations for having roots in the empirical world. What I have done in this paper is taken a few more steps in showing how this is so.[20]

NOTES

1. I use the term *description* here in a wide sense that is not restricted to descriptions having to be propositional. This is why, in my terminology, a toy aeroplane can be a description of a real aeroplane.
2. In contrast to the toy aeroplane, such theoretical models tend to describe a phenomenon in the form of propositions, propositions that derive their form from theories.
3. One could also think of phenomenological laws as laws that are merely generalisations, but they would be generalisations only with reference to one particular phenomenon, i.e. not "very" general.
4. While she does not make this explicit in 1983, Cartwright later highlights the proximity of her early position on scientific models to the semantic view of theories: '*How the Laws of Physics Lie* supposed, as does the semantic view, that the theory itself in its abstract formulation supplies us with models to represent the world' (Cartwright 1999a: 242). There may be a similarity in how to characterize the relationship between models and theory, but the fact that Cartwright portrays models as descriptions, even if fictional, does not go together with the portrayal of models as nonlinguistic entities, which they are according to the semantic view.
5. Cartwright seems to imply here that position, extension, motion, and mass are concrete concepts or at least more concrete than force. This seems like a claim hard to defend, but I will leave this issue here.
6. For a critique of the treatment of this example, see French and Ladyman (1997).
7. For a detailed discussion of the distinction between representative and interpretative models, see Morrison (this volume).
8. This is not something I want to deny. I and numerous others who study science agree with this thesis, and so do many scientists (Bailer-Jones 2002).
9. Besides by Cartwright, this tendency is also, for instance, resisted by Giere's perspectival realism (1999).
10. Cartwright can also be taken to subscribe to this view.
11. Those propositions that are true regardless of what they are about, i.e. are trivially true because they are tautologous, I consider as uninteresting.

12. Making this distinction itself presumes that whether or not something is about the same subject matter can be straightforwardly decided. Admittedly, this is a crude simplification.
13. Notice that this talk of models as true would obviously not arise in the context of Cartwright's early position portraying models as fictions. But as I have shown in section 'Models in Cartwright's philosophy', Cartwright's position with regard to models has undergone changes, and there is no basis to assume that she still thinks of models as fictions, given that she talks of them in terms of truth and representation.
14. Compare the wide sense of *description* which I introduced earlier.
15. Cartwright discusses idealisation and abstractness in Chapter 5 of *Nature's Capacities and their Measurement* (Cartwright 1989). There the notions of abstractness and idealisation are expected to do work in the context of the concept of capacities and of causality, but this is a somewhat different context from *theories* being abstract.
16. There are exceptions to this. For some phenomena that are modelled there exists only one specimen that is taken into account, e.g., the earth.
17. I am aware that the term *prototype* has some connotations that are counterintuitive to my use of it, but for want of a better alternative I introduce it here as a technical term to be used in the way described in the following.
18. The *Oxford English Dictionary* defines *abstract*: 'Withdrawn or separated from matter, from material embodiment, from practice, or from particular examples. Opposed to *concrete*', besides older uses. Cartwright (1989: 197, 213) identifies this as the Aristotelian notion of abstraction. She recounts: 'For Aristotle we begin with a concrete particular complete with all its properties. We then strip away—in our imagination—all that is irrelevant to the concerns of this moment to focus on some single property or set of properties, "as if they were separate"' (Cartwright 1989: 197).
19. Some sciences may be hard-pressed to formulate theories or principles that are abstract enough to apply quite generally, although an effort is often made. In other words, in the scenario I sketch there can be sciences that only employ models and do not have theories.
20. For illuminating exchanges of what it means to be abstract and for reading a draft of this chapter I sincerely thank Joke Meheus, Jim Bogen, and Peter Machamer. I am grateful to referees Paul Teller, Robert Rynasiewicz, and Margaret Morrison for helpful comments.

REFERENCES

Achinstein, P. (1968) *Concepts of Science*, Baltimore: Johns Hopkins Press.

Bailer-Jones, D. M. (1999) 'Tracing the Development of Models in the Philosophy of Science', in L. Magnani, N. J. Nersessian, and P. Thagard (eds) *Model-Based Reasoning in Scientific Discovery*, New York: Kluwer Academic/Plenum Publishers.

———. (2002) 'Scientists' thoughts on scientific models', *Perspectives on Science*, 10: 275–301.

———. (2003) 'When scientific models represent', *International Studies in the Philosophy of Science*, 17: 59–74.

Bogen, J., and J. Woodward. (1988) 'Saving the phenomena', *The Philosophical Review*, 97: 303–352.

Burian, R. M. (2001) 'The dilemma of case studies resolved: The virtues of using case studies in the history and philosophy of science', *Perspectives on Science*, 9: 383–404.

Carnap, R. (1939) 'Foundations of logic and mathematics', *International Encyclopaedia of Unified Science*, Chicago: Chicago University Press.

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Clarendon Press.

———. (1989) *Nature's Capacities and their Measurement*, Oxford: Clarendon Press.

———. (1991) 'Fables and models', *Proceedings of the Aristotelian Society Suppl.*, 65: 55–68.

———. (1997) 'Models: The blueprint for laws', in L. Darden (ed.) *PSA 96*, *Philosophy of Science* 64 (Proceedings): 292–303.

———. (1998) 'How theories relate: Takeovers or partnerships?', *Philosophia Naturalis*, 35: 23–34.

———. (1999a) 'Models and the limits of theory: Quantum Hamiltonians and the BCS model of superconductivity', in M. Morgan and M. Morrison (eds) *Models as Mediators*, Cambridge: Cambridge University Press.

———. (1999b) *The Dappled World. A Study of the Boundaries of Science*, Cambridge: Cambridge University Press.

Cartwright, N. et al. (1995) 'The tool box of science: Tools for the building of models with a superconductivity example', in W. E. Herfel et al. (eds) *Theories and Models in Scientific Processes*, Amsterdam: Rodopi.

Duhem, P. (1914; 2nd edn 1954) *The Aim and Structure of Physical Theory*, trans. P. P. Wiener, Princeton: Princeton University Press.

French, S. and J. Ladyman. (1997) 'Superconductivity and structures: Revisiting the London account', *Studies in History and Philosophy of Modern Physics* 28: 363–393.

Giere, R. (1999) *Science Without Laws*, Chicago: University of Chicago Press.

Harré, R. (1960) 'Metaphor, model and mechanism', *Proceedings of the Aristotelian Society*, 60: 101–122.

Hesse, M. (1966) *Models and Analogies in Science*, Notre Dame: University of Notre Dame Press.

Morrison, M. C. (1998) 'Modelling nature: Between physics and the physical world', *Philosophia Naturalis*, 35: 65–85.

Morrison, M. C., and M. S. Morgan. (1999) 'Models as mediating instruments', in M. Morgan and M. Morrison (eds) *Models as Mediators*, Cambridge: Cambridge University Press.

Pinnick, C., and G. Gale. (2000) 'Philosophy of science and history of science: A troubling interaction', *Journal for General Philosophy of Science/Zeitschrift für allgemeine Wissenschaftstheorie*, 31:109–125.

Pitt, J. (2001) 'The dilemma of case studies: Toward a Heraclitian philosophy of science', *Perspectives on Science*, 9: 373–382.

Suppes, P. (1961) 'A comparison of the meaning and uses of models in mathematics and the empirical sciences', in H. Freudenthal (ed.) *The Concept and the Role of the Models in Mathematics and Natural and Social Sciences*, Dordrecht: D. Reidel Publishing Company.

Teller, P. (2001) 'Twilight of the perfect model', *Erkenntnis*, 55: 393–415.

Reply to Daniela Bailer-Jones

Daniela Bailer-Jones's paper forces us to confront the questions of what it means to say that a theory is true, what it means to say that a model is true, and what it means to say that a model provides a true representation. For my own part I have no real philosophic views about truth and indeed follow Arthur Fine in his claim that there is—and probably should be—no "theory" of truth. But I do think that scientific models sometimes provide claims about the world, that sometimes these claims are meant to be true or approximately true, that sometimes they might well be true, and that sometimes we have good evidence to suppose them to be true. Often even when models are intended literally, not everything in the model is meant to depict something in the world and certainly not everything in the world—perhaps not even everything relevant to the phenomenon under study—is meant to be depicted in the model.

One thing we should not assume, which Bailer-Jones says in passing but I do not think she believes to be universally the case, is that if a model is supposed to represent the world truly, all its deductive consequences are also true (Bailer-Jones this volume: 17). Clearly, very often only some things depicted in the model or that follow from the model are meant to depict the world correctly. Or, as Mary Morgan urges, often the static deductive consequences do not matter at all, but rather things we learn when we "experiment" on the model in various ways. Morgan has taught us the importance of the stories that go along with, or perhaps even partly constitute, models in fixing what we are to learn from the model. One of the many functions of the story is to make clear what kinds of claims taken in what kinds of ways are supposed to be true of the model and in what kinds of circumstances.

To do its job, then, the story is going to have to tell us what form of claim is being made. It might be a universal generalization or a capacity claim; it might be a claim about what actually happens in target systems (sometimes? always? for the most part? under certain specific conditions? erratically?) or about what might or could happen (or. . . . ?). Clearly this needs to be settled before we can think about conditions for truth or evidence. Take as a simple example Bailer-Jones's discussion of the moral, 'the weaker is prey to the stronger' (Bailer-Jones this volume: 20). This may be meant as a capacity claim: Strength brings the capacity to prey on the weak (where we assume the capacity may be countered by other factors). Even then we must consider, is the model claiming that strength *always* brings this capacity? Sometimes? Always, but in some very specific circumstances? Or . . . ? On the other hand we may take the claim not as ascribing a capacity but rather intended to hold literally. As Bailer-Jones remarks, 'there may be individual cases for which this moral or "theory" works, or there may be '*many* instances', or it may work 'universally'. Or we may mean the claim as having an unstated conditional in front: There are conditions *X* such that it holds. But we must still ask: How often? Always? Some set portion of the time? Or . . . ?

Bailer-Jones remarks that I do not 'elaborate the concept of representation which [I use] to say that theories do not represent the world and that models do' (Bailer-Jones this volume: 12). At least with respect to the claim that representative models represent the world, I should like to follow Bailer-Jones's own lead, which I take to be similar to the views of Arthur Fine. If the story makes clear what claims we are to derive from the model and how they are to be understood—as it ideally should—then we judge whether the model accurately represents the world by judging in the usual ways whether those claims are true, or true enough. As Bailer-Jones says, 'Truth is something that can be attributed to propositions, and a proposition counts as true in those cases in which things are in the world as the proposition states' (Bailer-Jones this volume: 17).

I also agree with Bailer-Jones that we do not want to count two contradictory propositions as true. She raises this issue especially with regard to my claims that we can (and do) assign both quantum and classical descriptions to the same system. In particular I claim that in many cases factors represented in quantum mechanics and factors represented in classical mechanics combine to produce an effect not literally predictable from either theory alone. Nevertheless we are often able to produce a representative model (naturally not falling properly under either theory) that can provide good predictions about the targeted effects. In fact it looks to me as if this is what we usually do when we actually produce predictions that we will judge as true or false.

Bailer-Jones worries that this raises problems of consistency 'because classical mechanics and quantum theory are based on different principles' (Bailer-Jones this volume: 18). Again I agree they are based on different principles but whether that leads to problems of consistency depends on how we read those principles. I find it hard to figure out exactly how other people want to read them. I have made two different kinds of proposals, neither of which at least *prima facie* suggests that inconsistencies need result. First is that they are "tools" for building models where there may be lots of local expertise about how to deploy these tools—and how to deploy the tools provided by different theories together—but no overarching rules to tell us how. The second is to read the principles with a particular kind of *ceteris paribus* clause in front; e.g. 'So long as no factors relevant to the effect occur other than those that can be represented as forces occur, then $f_t = ma$ holds.'

Either of these readings support my claim that theory is not true of the world, once they are coupled with the claim that to get a good representative model whose targeted claims are true (or true enough) we very often have to produce models that are not models of the theory. What about my claim that theory is true in models? Here I take on board Bailer-Jones's claims that theories, or theoretical terms, are general because they are abstract, and I agree that 'To see how theory holds in a model we need to fill in the concrete detail that is not part of the theory because, being abstract, the theory has been stripped precisely of those details' (Bailer-Jones this volume: 24). That is indeed how we get a model that theory is true of. But as I have just explained, it is not enough to get a good representative model whose targeted claims will be true of the world.

I have here been claiming that the story of the model ideally tells us what claims the model is supposed to make, that these claims, as Bailer-Jones maintains, are propositions, and that the propositions are to be judged true or false in the usual ways—there is no special problem about scientific models. This is straightforward when the model's claims are about one specific concrete system. But what happens when they are about, say, "the hydrogen atom"? Here we face a big problem about what it means for a scientific model to be a representation and also what it is to be a "correct" representation. Bailer-Jones is kind enough to mention my work on abstraction in this regard, but it is rather her own work on abstraction that helps here. So I should like to close by remarking how important it is that she has articulated this problem and set it centre-stage. Her own account of prototypes as the target of representations in these cases provides an exciting way to attack the problem.

3 Nancy Cartwright on Theories, Models, and Their Application to Reality

A Case Study[1]

Ulrich Gähde

INTRODUCTION

Most approaches in the philosophy of science focus on the structure and dynamics of theories but have little to say—if anything at all—about how these theories are applied to concrete systems. By contrast, Nancy Cartwright's approach provides a detailed account of how empirical theories are in practice applied to reality and what role models play in this process. These views have developed and changed over the years—a process that is described in depth in Bailer-Jones's contribution to this volume. The aim of the following sections is to analyze certain aspects of these views in more detail.

As both the advantages and snags of positions in the philosophy of science become especially apparent when being applied to concrete case studies, I will not carry out my considerations *in abstracto*. Instead, I illustrate and discuss Cartwright's view by using a case study from the history of astronomy, that of Edmond Halley's discovery of the comet later named after him, as well as the subsequent attempts to obtain an adequate theoretical description of the comet's orbit. I largely follow Cartwright's line of thought as presented in her early paper 'Fitting Facts to Equations', which is especially well suited for the treatment of this special case study (Cartwright 1983: 128–142).[2] However, I shall whenever necessary add remarks concerning more recent developments in her position concerning the relationship between theories and models.

In Cartwright's view as presented in that paper, the application of an empirical theory can be described as a three-step process. The first step provides an unprepared description of the system in question: Any information which is thought to be relevant is collected in whatever form it is available. The theory-observation distinction is irrelevant here: The unprepared description may well use the language and concepts of the theory—without, however, being constrained by its mathematical needs. The unprepared description is chosen solely on the grounds of being empirically adequate.

The second step leads to the prepared description. The main task of this step is to make the theory applicable to the phenomenon in question. For this purpose, the unprepared description is replaced by a prepared description—'a description to which the theory matches an equation' (Cartwright 1983: 133). According to Cartwright, the primary concern during this step is not how well the facts concerning that system are represented within the theory but only how successfully the mathematical apparatus of the theory can be applied to it. It is important to note that this is an informal step: The choice of the prepared description is only guided by rules of thumb. The theory does not provide any explicit, let alone formal, principles for that purpose.

In this respect, the third and last step differs essentially from the first two. It consists in the mathematical treatment of the phenomenon in question. Once the prepared description has been chosen, the theory dictates not only what equations have to hold, but boundary conditions and approximations as well.

Nancy Cartwright substantiates the above view in thorough analysis of numerous informative examples. She is particularly interested in the role that theoretical and phenomenological laws play in the process. The results she obtains are in sharp contrast to a widely held view.

According to this common view, phenomenological laws can be derived from fundamental laws, which hold in each and every application of the theory: The phenomenological laws are true because they can be derived from fundamental laws which are themselves true. Nancy Cartwright holds that this view is profoundly misleading. She argues that it is just the phenomenological laws that provide us with highly precise, detailed descriptions of physical systems and thus carry the main burden of description and explanation. These phenomenological laws can only be derived from the theory's basic principles by a long series of approximations and emendations. The fundamental laws themselves hold only in those rare situations in which but one cause (gravitation, electromagnetic force, etc.) is at work. In all more realistic cases, in which numerous causes act together, these basic laws simply don't state the facts. Cartwright articulates this view in her provocative thesis that '. . . fundamental equations do not govern objects in reality; they only govern objects in models' (Cartwright 1983: 129).

The main task of the following considerations is to evaluate both Nancy Cartwright's three-step view on how empirical theories are applied to concrete systems and her theses concerning the role fundamental and phenomenological laws play in this process. In doing so, I compare her views with theses on how empirical theories are applied to reality put forward within the structuralist framework. Comparing these two approaches reveals some striking similarities, as well as some informative disanalogies.

In the next section I provide some background about my case study in the form of a brief account of the discovery and investigation of Halley's Comet. Then I try to apply Nancy Cartwright's concept of unprepared description to this particular case and compare it to the structuralist concepts of intended applications and data structures. The same procedure is then carried out with respect to Cartwright's concept of prepared description. The penultimate section is devoted to the question of how the different attempts to obtain an adequate mathematical description of this comet and its orbit can be mirrored in both approaches. The final section provides a short summary of my comparison of the two approaches and poses some questions that emerged from the preceding considerations.

AN EXAMPLE FROM THE HISTORY OF SCIENCE: THE DISCOVERY OF HALLEY'S COMET

Until the research of Tycho de Brahe, most astronomers believed Aristotle's view that comets were not celestial but terrestrial phenomena. Accordingly, the study of comets belonged to the realm of meteorology rather than astronomy. Tycho de Brahe's measurements, however, clearly showed that the comet of 1577 was celestial and located among the planets. This discovery led to a growing interest in the precise measurements of the paths of comets, as well as in the physical laws governing their movement. When a bright comet appeared in 1682, Edmond Halley (1656–1742) tried to determine the data of its path and to compare them with the data of comets observed previously, including sightings in ancient times (at least those for which records were available).

Two years later, he visited Newton in Cambridge and asked him 'what he thought the Curve would be that would be described by the Planets supposing the force of attraction towards the Sun to be reciprocal to the square of their distance from it' (ctd. in Hoskin 1997: 157). Newton answered that this curve was an ellipse. Halley asked Newton for proof, which Newton could not provide right away. Several months later, however, he sent Halley a short manuscript, which was mainly concerned with the movements of objects in empty space. Although this manuscript was just nine pages long, Halley immediately noticed its revolutionary content. It became one of the nuclei from which Newton's major work, the *Philosophiae naturalis principia mathematica*, evolved.

In 1687 the *Principia* finally appeared, and it had been Halley who had coaxed Newton into writing it and who paid for its publication. The *Principia* contained a considerably generalized treatment of the gravitational forces between *all* (terrestrial and celestial) bodies. In particular, it contained a chapter on comets in which Newton argued that they would move on elliptical, parabolic, or hyperbolic paths. In the second and third editions of the *Principia*, published in 1713 and 1725, this chapter was replaced with more detailed theoretical treatments.

Newton's insight provided the starting point for Halley's considerations. If all comets followed conic paths, Halley argued, then it seemed likely that at least some of them moved on closed elliptic orbits. In this case some comets should reappear: Some of the presumably different comets observed in the last few decades, as well as those recorded since ancient times, might be reoccurrences of one and the same object. Thus one only had to look for comets which shared characteristic properties. He soon found three comets that fitted the bill for being the same object reappearing: the comets of 1531, 1607, and 1682. All of them had one important feature in common: Their movement around the sun was retrograde, i.e. opposite to the movement of the planets. Closer analysis revealed further striking similarities between them. One important indicator that they were in fact the same object was provided by the time intervals between their appearances: 76 years between the comets of 1531 and 1607 and 75 years between the comets of 1607 and 1682. However, Halley carefully noted that although these time intervals were similar, they were not identical. He predicted that the comet ought to reappear 'about the end of the year 1758, or the beginning of the next' (Hoskin 1997: 173). It should be noted that Halley reached this prediction not simply by forming the mean value of the two time intervals but by theoretical considerations to which I turn later.

On Christmas day, 1758, a farmer near Dresden first observed the reappearance of the comet later named Halley's Comet. This observation was confirmed by a professional astronomer a few weeks later. This discovery was regarded as a scientific sensation: Comets had long been viewed as unpredictable signs of imminent disaster very much in contrast to the planets with their ordered, calculable movements. The successful prediction of the reappearance of Halley's Comet was seen as a prediction of the unpredictable and, thus, a staggering triumph for Newton's theory on which Halley's considerations were based.

In the following sections, I discuss how the history of this discovery and the different attempts to reach an adequate theoretical description of Halley's Comet can be mirrored in Cartwright's considerations concerning how empirical theories are applied to reality. In this context I discuss how successfully the concepts of unprepared description, prepared description, and mathematical description can be applied in my case study.

THE "UNPREPARED DESCRIPTION"

Let us turn to Cartwright's characterization of the unprepared description first. For this purpose, I cite two particularly concise theses put forward by Nancy Cartwright on this issue. I analyze how each of these issues can be interpreted in my case study and whether or not it is confirmed by it. Finally, I discuss what the structuralist approach has to say about the issue addressed in these theses and how well it fits in with Cartwright's view.

First thesis: 'The unprepared description contains any information we think relevant, in whatever form we have available. The unprepared description is chosen solely on the grounds of being empirically adequate' (Cartwright 1983: 133). The first part of this thesis is nicely corroborated by the wide variety of data that have been gathered during the history of research into comets. The second part of the thesis, however, has to be treated with some caution. The discovery and investigation of comets provide a striking example of the fact that the choice of data which are thought to be relevant and are collected is by no means solely guided by the aim to achieve empirical adequacy but depends on massive theoretical background assumptions as well.[3]

From Greek antiquity until the rise of modern natural science in the sixteenth century, at least in the Western tradition, the investigation of comets was dominated by Aristotle's claim that comets were not celestial but terrestrial (atmospheric) phenomena. As a consequence, two types of data were of main interest: data concerning the objects themselves (such as their brightness, shape, and the directions in which they could be observed) and data that related their appearance to terrestrial events (such as disasters, etc.).

As mentioned above, it was only thanks to the research by Tycho de Brahe that it became clear that comets were celestial and not terrestrial objects. This discovery made the increasingly precise measurement of their positions and movements relative to other astronomical objects one of astronomers' primary concerns. A hundred years later, Newton's discovery that *all* bodies interact via gravitational forces did not only remove the widely accepted division between physics and astronomy but made data concerning the physical properties of other celestial objects—such as the masses of the sun and Jupiter—relevant for adequately describing comets and their paths.

More specialized hypotheses put forward within the framework of Newton's theory had an important impact on what data were thought to be relevant as well. Halley's hypothesis readily illustrates this point: The insight that putatively different comets might constitute reappearances of one and the same object led to an increasing interest in all data indicating similarities between the physical properties of these objects.

An interesting aspect of this case study is that not all data used for the evaluation of Halley‘s hypothesis were newly collected for that purpose. By contrast, one of the peculiarities of Halley's research consisted in the fact that he analyzed old data that had been collected many centuries before. It was on the basis of these data that Halley claimed that the comets that had appeared in 1305, 1380, and 1456 were all reoccurrences of "his" comet. Of course, these data were collected in the light of theoretical background assumptions that differed fundamentally from Halley's own scientific beliefs.

It is interesting to note as well which data—at a certain stage of scientific development—were not thought to be relevant and thus were not collected. For example, as far as I am aware neither Newton himself nor any of his contemporaries ever hit on the idea of studying the light of a comet through a prism. The data that could thus have been obtained became relevant only after theories enabling the extraction of information about the surface and the chemical constitution of celestial objects from an analysis of their spectra became available.

These considerations show that some aspects of Nancy Cartwright's view concerning the unprepared description have to be taken with a pinch of salt. According to Cartwright, the unprepared description is chosen solely on the grounds of being empirically adequate. Theory only enters at the stages of the prepared description and the mathematical description. However, my case study shows that, at least in the case of Halley's Comet, in one important sense the theory already enters at the stage of the unprepared description: What data were thought to be relevant and were thus collected was guided by the respective theory which was meant to be applied to this object. As these theories significantly changed over time, a seemingly amorphous variety of data of widely differing types were collected.

Second thesis: 'There is no theory-observation distinction here: We write down whatever information we have. . . . The unprepared description may well use the language and the concepts of the theory, but is not constrained by any of the mathematical needs of the theory' (Cartwright 1983: 133). When discussing this thesis with respect to my case study, it should be borne in mind that the notoriously vague theory-observation distinction is particularly problematic with respect to astronomical research. The main reason is that celestial objects differ from most other physical objects in that they cannot be manipulated in experiments—a fact that significantly restricts the possibilities for observation and increases the relevance of theoretical considerations. Statements which at first sight undoubtedly seem to refer to observable facts turn out to be highly theory-laden when regarded more closely. Let us consider the following statement: 'After its appearance in 1682, Halley's Comet was observed again in 1758'. At first reading, this seems to be a clear case of an observational statement, as suggested by the formulation itself. However, this claim cannot be tested by direct observation: The hypothesis that the comets observed in 1682 and 1758 were in fact one and the same object implies that they had very similar physical properties (by which they can be identified).[4] Most of the relevant properties, however, are not directly accessible to observation: Astronomical investigation always has to start from a very restricted informational basis that consists mainly of data concerning certain segments of the electromagnetic radiation emitted or reflected by celestial objects. From these data, all the other physical properties of these objects (such as their masses, diameters, temperatures, and chemical compositions) have to be obtained by complicated calculations that are based on massive theoretical assumptions.

For the sake of the argument, let us set aside problems in connection with the vagueness of the theory-observation distinction. Under this assumption, it may be said that at least in the beginning of scientific research into comets scientists focused on physical properties such as their apparent brightness,[5] the shapes of their cores and tails, etc., which were comparatively easy to observe. After the emergence of Newton's theory at the latest, however, data were collected which were heavily theory-laden. When the first values for the masses of the sun and the planets, for example, had been obtained, they were unhesitatingly judged to be relevant for an adequate description of the movement of comets and thus incorporated into the corresponding database.

This view is in sharp contrast to the original treatment of intended applications as presented in the structuralist approach, to which I now turn. Here it is assumed that in a first step the phenomena to which a scientific theory is to be applied have to be characterized without making use of the theoretical concepts introduced by that very theory. Then, in a second step, it has to be shown that these intended applications can be extended into models of the theory (in the set-theoretic sense) by adding suitable theoretical functions such that the laws of the theory in question are fulfilled.[6] In this approach the theoretical-observation distinction is substituted by the (highly problematic) theoretical–nontheoretical distinction, which I do not discuss in this chapter in any further detail.[7]

Let me illustrate the basic idea behind the original structuralist approach[8], as well as some of its problems, by means of my case study: Halley's Comet as an intended application of Newton's gravitational theory. As stated before, the characterization of this intended application has to be facilitated by using terms which "do not come from that theory". First, one has to specify the set P of objects involved in this intended application. P may be seen as including the comet itself, the sun, the earth, and the five planets known at Halley's time: the two "inner" planets Mercury and Venus, as well as the three "outer" planets Mars, Jupiter, and Saturn. Second, one has to provide information concerning the observational interval T during which the measurements were taken. Finally, one has to state the position function s, which describes the positions of the objects of P during T. In the original structuralist terminology, the triplet $z = \langle P, T, s\rangle$ was called a partial model and was identified with the corresponding intended application. Note that z constitutes a purely kinematic description of the system in question: It provides information on the positions and movements of the objects involved but no information whatsoever on their masses or on the forces by which they interact. The task of providing a theoretical description of this intended application z (which I examine more closely in the following sections) can now be stated as follows: a mass function m and a force function f have to be found, by which the partial model $z = \langle P, T, s\rangle$ can be extended into a model $x = \langle P, T, s, m, f\rangle$ of Newton's gravitational theory.

Even this sketchy account of the original structuralist concept of intended applications should suffice to show that it is based on several highly idealizing, fictitious assumptions. Let me focus on the two of these assumptions that are especially important with respect to the relation between Nancy Cartwright's view and the original structuralist approach.

First, the application of the concept of partial models presupposes that *all* values of the nontheoretical functions are known. In most situations, however, this will not be the case, for at least two reasons. The first is trivial: In general, these functions are defined for an infinite number of arguments. By contrast, only a finite number of measurement data is ever available. The second reason is nontrivial and concerns the specific situation in which measurement is carried out: In most cases there will be certain circumstances that further restrict the set of data available. This second point can be readily explained by my example. Unlike most planets, comets have orbits with significant eccentricity. Furthermore, their maximum distance from the sun may by far exceed that of the planets, thus making them difficult or even impossible to observe when they move outside Jupiter's orbit. Finally, they cannot be observed at all when they are behind the sun or obscured by some other object (e.g., a planet).

Consequently, the basic assumption underlying the use of the concept of partial models is fictitious, for only a systematically restricted number of values will be known for the functions occurring in these partial models instead of all of them. Therefore, the concept of partial models was substituted by the more realistic concept of data structures. Data structures may be defined as substructures of partial models as follows:

> Def.: $\tilde{z}$ is a (finite) data structure iff there exists a partial model $z \in M_{pp}$ such that $\tilde{z}$ is a (finite) substructure of z ($\tilde{z} \sqsubset z$).[9]

By making use of data structures instead of partial models one can account for two facts: (1) For each intended application only a limited, finite database will be available. (2) This database is constantly changing; new, more precise data will become available, and old data might be modified or discarded. Both features are of crucial importance for an adequate account of how empirical theories are applied to concrete systems.

Second, the original structuralist concept of intended applications is based on the highly problematic theoretical–nontheoretical distinction. According to this concept, intended applications had to be described by nontheoretical functions only. As we have seen in the case of Halley's Comet, however, situations can be easily imagined in which not only the position functions for all objects involved are known, but in which the same holds with respect to the sun's mass (which might have been determined in previous applications of the theory). In this case, only the mass of the comet and the forces acting between this object and the sun would have to be determined. In order to account for situations of this type, it has been proposed to considerably liberalize the concept of intended applications (Balzer et al. 1993). According to this proposal, intended applications are to be represented by data structures that might contain values of both theoretical and nontheoretical functions with respect to the theory in question.

These considerations show that at least one basic idea behind the structuralist approach is completely independent of the problematic theoretical-nontheoretical dichotomy: the idea that the task of providing a theoretical description of some concrete system may be restated as the problem of extending some fragment of a model into a complete model of the theory in question.

This proposed liberalization of what is meant by intended applications brings it closer to Nancy Cartwright's concept of an unprepared description. Just as the theoretical-observational is irrelevant for Cartwright's unprepared description, the theoretical-nontheoretial distinction is irrelevant for the liberalized structuralist concept of intended applications. However, a significant difference between both notions should be kept in mind. According to Cartwright, the unprepared description contains 'any information thought to be relevant'. Exactly what is meant by "relevant", however, is not specified in any further detail. By contrast, within the structuralist framework only those data that are relevant in a very specific sense are collected and used for the characterization of intended applications. Data are relevant if and only if they are part of the model to which the corresponding finite data set is to be extended. In other words, intended applications are described as fragments of models of the theory in question. In the case of Halley's Comet, for example, ephemerides (values of the comet's position function) are relevant because they are part of the model of Newtonian gravitational theory into which the corresponding intended application is to be extended.

One of the basic insights of Nancy Cartwright's approach, however, is that in most cases the theory will not be successfully applicable to the unprepared description. For this purpose, it has to be substituted with a prepared description, to which we now turn.

THE "PREPARED DESCRIPTION"

Here is what Nancy Cartwright has to say about the prepared description:

> We present the phenomenon in a way that will bring it into the theory. The most apparent need is to write down a description to which the theory matches an equation. The check on correctness at this stage is not how well the facts known outside the theory are represented in the theory, but only how successful the ultimate mathematical treatment will be. This first stage of theory entry is informal. There may be better and worse attempts, and a good deal of practical wisdom helps, but no principles of the theory tell us how we are to prepare the description. (Cartwright 1983: 133)

The example of Halley's Comet nicely illustrates these theses. Let us turn to some details. As has been mentioned before, Halley noted that the intervals between the occurrence of the comets of 1531, 1607, and 1682 were similar but not identical. He rightly assumed that these differences were caused by the comet being pulled by one or more planets. In particular, he thought that Jupiter might have an important influence on the comet's path. However, all attempts to account for this influence in the theoretical description of the system led to considerable difficulties. In Halley's time, the available mathematical tools could not provide a sufficiently accurate approximate solution of a three-body-problem. Thus, Halley decided to account for the impact of Jupiter in a somewhat half-hearted way. In his calculations, he considered the pull exerted by Jupiter on the comet when it entered the solar system but ignored the opposite consequences of the planet's pull when the comet moved away from the sun. Of course, this strategy when dealing with the comet's orbit was unsatisfactory, if not inconsistent. It was not chosen because Halley believed it to present a true account of all the details of that system, but for two other reasons. First, it made Newton's theory and the mathematical tools available at that time applicable to the comet's movement without completely neglecting the influence of the planets. Second, it enabled a comparatively precise prediction of when the comet would reappear.

Nevertheless, astronomers struggled to obtain a more refined and accurate theoretical solution of the problem. In 1757, Alexis-Claude Clairaut, together with two coworkers, provided a more ambitious treatment. They proceeded in three steps. In the first step, they analyzed "old" data concerning the orbit of the comet as it left the solar system in 1531. In a second step, they used these data to "predict" (or, to be more precise, to rectrodict) its return in 1607. They compared their prognosis with what actually happened. In a third step, they analyzed the comet's orbit as it left in 1607 and learned from these data in order to predict that the comet would move around the sun in mid-April 1759—a prediction that tallied well with the observational data obtained when the comet in fact reoccurred.

These historical details support Nancy Cartwright's views on the prepared description in many respects: They show that it was neither the only nor even the main concern of the astronomers involved in these investigations to provide a highly accurate theoretical description of each and every fact "known outside the theory". By contrast, comparatively reliable data which were already available when the investigations were carried out were discarded. In the case of Halley's treatment, this holds with respect to data concerning the position and movement of Jupiter during the time interval when the comet left the solar system: Halley simply proceeded as if the comet had vanished—although, of course, its position during that time interval was well known to him. Clairaut and his coworkers used a considerably enriched database. However, it still did not contain all the data available. Instead, Clairaut referred to a restricted informational base that was chosen so Newton's theory and the mathematical tools available then could be used.

So far, my case study nicely illustrates Cartwright's point that in order to make an abstract theory applicable to some concrete phenomenon, a somehow simplified model of this phenomenon will have to be used—a model that is known to rest (at least in part) on fictitious assumptions:

> I think that a model—a specially prepared, usually fictional description of the system under study—is employed whenever a mathematical theory is applied to reality, and I use the word "model" deliberately to suggest the failure of exact correspondence. . . .
>
> (Cartwright 1983: 158–159)

In a later paper, Cartwright (1991) draws an interesting parallel between models and fables: Fables illustrate an abstract moral by telling a concrete story, in which this moral is instantiated. In a very similar way, abstract theories or laws are instantiated—or "fitted out" (to put it in Cartwright's words)—by applying them to more concrete situations. However, they can be applied to these concrete situations only in a derivative sense, via simplified, partly fictitious models—in a similar way to fables, which only tally with certain aspects of reality and are fictitious with respect to others. My case study readily illustrates these points.

Furthermore, it corroborates the thesis that providing a prepared description is an informal step: The theory itself—here, Newton's gravitational theory—does not supply any principles as to how one has to proceed. This fact is illustrated by the variety of different prepared descriptions others later used in their attempts to obtain a theoretical treatment of that system (i.e. comet-sun-Jupiter).

Cartwright's claim that 'the check on correctness at this stage is not how well the facts known outside the theory are represented in the theory, but only how successful the ultimate mathematical treatment will be', is an obvious exaggeration. The following formulation seems to be more adequate: '. . . a model is a work of fiction. Some properties ascribed to objects in the model will be genuine properties of the objects modeled, but others will be merely properties of convenience' (Cartwright 1983: 153).[10] My case study emphasizes this point: At least some facts "known outside the theory" were taken very seriously. If, for example, Halley or Clairaut had not succeeded in providing a fairly accurate description of how and when the comet would pass around the sun in 1759, nobody would have been interested in their calculations, no matter how successfully the mathematical apparatus of the theory could be applied here. At least two criteria guide the choice of the prepared description: First, it should make the theory become applicable to the phenomenon in question; second, although the prepared description will in general lead to a simplified picture of this phenomenon, it should nevertheless enable a theoretical description that adequately mirrors certain selected aspects of the system (such as the movement of the comet when passing near the sun) while neglecting others (such as the influence of the planets on its path). What aspects are selected will depend on pragmatic criteria and may well change in time. The task of the prepared description is to enable theoretical insights into *certain aspects* of the phenomenon in question. By contrast, it will not lead to an empirically adequate description of other aspects, but nor does it claim to do so.

Let me add some remarks concerning more recent developments in Cartwright's work concerning the relationship between theories and models. In (Cartwright 1983), her argumentation was based on the assumption that although theories can be applied to reality only via simplified and partly fictitious models, they nevertheless play *the* decisive role in the construction of these models—no matter how many idealizations and approximations might be involved. The prepared description, in particular, is chosen such as to make a certain theory applicable to the system in question. In more recent publications Cartwright argues against this 'theory-dominated view of science' (Cartwright et al. 1995; Cartwright 1999b). Instead, she insists that theories are but one element among others in the toolbox of science, which also contains mathematical techniques, instruments, etc. According to this modified view, these other tools may be equally important for the choice and construction of models:

> There are only real things and the real ways they behave. And these are represented by models, models constructed with the aid of all the knowledge and techniques and tricks and devices we have. Theory plays its own small important role here. But it is a tool like any other; and you can not build a house with a hammer alone. (Cartwright et al. 1995: 140)

The crucial point in this modified view is that modeling does not necessarily have to be theory-driven. Cartwright et al. (1995) try to illustrate this point by a detailed analysis of the construction of a model for superconductivity in the first half of the twentieth century. There she and her coworkers attempt to demonstrate that there are cases, in which ad hoc adjustments to a theory are carried out that are not guided by this or any other theory but are justified on solely phenomenological grounds.

Cartwright (1999a, b) continues this line of thought and adopts a view of models as mediators, as put forward by Morrison and Morgan (1999) and others. Here she distinguishes between interpretative and representative models. Interpretative models are characterised by the fact that they are laid out within a theory. Abstract terms that occur in a theory are made more concrete in interpretative models via bridge principles. However, interpretative models only refer to specific, simple situations and will not be generally well suited to mirror real systems in their complexity. This is the task of representative models. They can (but they do not have to) draw from theory, and they are constructed in such a way that they are applicable to concrete real systems in a dappled word.

It is not quite immediately obvious how the different attempts to model the movement of Halley's Comet around the sun are to be classified in this scheme. Take Halley's original treatment of the comet's path as an example. On the one hand Newton's gravitational theory did not play "its own little important role" in this attempt. By contrast, it played *the* decisive part in this modeling process. In that respect, Halley's treatment seems to be a typical example of theory-driven modeling. This is the main reason why, for the most part, I here follow the line of thought outlined in Cartwright (1983), in which theory is still attributed central importance in the creation of models—a view which is evidently particularly appropriate for my case study.

On the other hand, the correctional term that was introduced to account for the impact of Jupiter was not derived from that theory in any strict sense but was adapted so as to fit the observational data. However, this procedure was not chosen because the capacity of the theory for dealing with this system was questioned in principle. By contrast, it was chosen because the mathematical techniques available at that time—in particular those that were relevant for the treatment of a many-body system—did not suffice to apply Newton's gravitational theory to this concrete system in full. When mathematical techniques improved, the correctional term was discarded, and Halley's original treatment of this system was substituted with more elaborate approaches that systematized successively more comprehensive, realistic sets of data and which were successively more closely bound to Newton's theory. Maybe an adequate way to describe this process in Cartwright's terms would be to say that the representative models used for the description of Halley's Comet first were gradually transformed into models that were interpretative models of Newton's theory as well.[11]

How can the structuralist approach mirror the process of manipulating and restricting the set of available data so as to make a theory applicable to them? How can it mirror what is meant by Cartwright's notion of prepared description? As discussed in the section 'The "unprepared description"', within the structuralist framework intended applications are represented by finite data structures that contain all the relevant data available.[12] Cartwright's basic insight is that in most cases it will not be possible to successfully apply the theory to this unrestricted data structure. Translated into the terminology of structuralism, this means that the original data structure cannot be extended into a model of the theory in question. For this purpose, it has to be substituted by a modified, restricted data structure that has to fulfill two requirements: (1) It has to be sufficiently similar to the original data structure with respect to certain relevant aspects. What is meant by "sufficiently similar", as well as which aspects are judged to be relevant, cannot be specified in general terms but may well depend on the characteristic features of the special case. (2) For the modified data structure, the theory has to "match an equation".

How can this modified data structure be characterised in the case of my example? First, the set *P* of objects is artificially restricted: All objects except the comet itself, the sun, and Jupiter are removed from *P*. At the same time, all data concerning the position functions *s* of these objects are removed from the original data structure. Furthermore, at least some available data concerning Jupiter's orbit are omitted: Halley's calculations treat the movement of the comet as if Jupiter exerted a considerable influence on the comet when the object entered the solar system but suddenly vanished when it moved away from the sun. At least in this case, the modified data structure thus obtained is a proper substructure of the original data structure as mentioned above.

This restricted data structure provides the starting point for the mathematical description of the system in question, to which we turn in the following section. My discussion will show that Cartwright's claim that the theory "matches an equation" for the prepared description is ambiguous in an important sense.

THE "MATHEMATICAL DESCRIPTION"

According to Nancy Cartwright, the mathematical description is '. . . the second stage of theory entry, where principles of the theory look at the prepared description and dictate equations, boundary conditions, and approximations' (Cartwright 1983: 134). Let us analyze the various claims contained in this thesis in order.

1. In the case of Halley's treatment of the comet, the equations involved can be easily identified: They include Newton's second axiom, as well as a special version of Newton's law of universal gravitation adapted to the case of a two-body problem (plus a correctional term that is meant to account for the influence of Jupiter). How, and the extent to which the use of these equations is dictated by principles of the theory, is discussed in greater detail below.
2. I cannot see in this case how Cartwright's claim that principles of the theory "dictate the boundary conditions" is to be interpreted. But perhaps this claim is only meant to refer to particular types of systems and to be vacuous with respect to others.
3. It is equally unclear to me in what sense—at least in the case of Halley's Comet—principles of the theory dictate approximations. By contrast, the history of successively more ambitious theoretical treatments of this comet carried out within the framework of Newton's theory clearly indicates that this theory is compatible with a large variety of different strategies of approximation. These approximations concern not only the objects considered but also their assumed properties. Thus, for example, approximations may concern the question

> of how the comet's mass is modified when the object approaches the sun, as well as the impact this modification may have on its path. Now Nancy Cartwright might reply that the principles of the theory dictate approximations only after a certain prepared description has been chosen. However, even in this case there may still exist numerous different ways of providing an approximative theoretical treatment of the system that are in accordance with the empirical data considered within the error of measurement available at that time. When, for example, the decision has been made to treat the system as a two-body problem (plus correctional term), numerous special laws proposed within the framework of Newton's theory may still—and in fact have been—used to provide an approximate theoretical treatment of the system. I return to that point below.

How can the mathematical description of a physical system be mirrored within the structuralist approach? If we refer to the original structuralist account of intended applications this task can be reformulated as follows: It has to be shown that the partial model z, which corresponds to the intended application in question, can be successfully extended into a model x of the theory to be applied to it. With respect to Halley's Comet, the starting point would be provided by a partial model containing all kinematic data concerning that system. In other words, the partial model has to supply a complete kinematic description. Then, it is claimed that mass and force functions can be found such that the laws of Newton's theory are fulfilled. This claim can be expressed by the following simplified version of the Ramsey sentence:

$$\exists\, x \left[x \in M \wedge r(x) = z \right].$$

Here z is the partial model in question, x is the model to which it is to be extended, and M denotes the set of models of that theory; r is the so-called restriction function that "cuts off" the two functions m and f.[13]

As discussed earlier, the assumption that a complete kinematic description of a system can be provided is highly fictitious. In my case study, for example, it is impossible to know all the ephemerides of the comet. In order to provide a more realistic account of how empirical theories are applied to reality, the concept of partial models was substituted with the concept of finite data structures, which represent all the relevant data which are available *de facto*. A data structure $\tilde{z}$ is a proper substructure of the corresponding partial model z: $\tilde{z} \sqsubset z$. The modified Ramsey sentence for data structures can be formulated as follows:

$$\exists\, x \left[x \in M \wedge r(x) = z \wedge \tilde{z} \sqsubset z \right].$$

In other words, the data structure $\tilde{z}$ can be embedded in a partial model z, which in turn can be extended into a model of the theory in question.

In most cases, however, it will not be possible to fulfill this existential claim: As Nancy Cartwright points out, the theory will not "match an equation" to the unprepared description to which the original finite data structure $\tilde{z}$ corresponds. Therefore, it is substituted with a modified—and, in general, highly simplified—data structure $\tilde{y}$ which corresponds to the prepared description in Cartwright's terminology. As described before, in my case $\tilde{y}$ is obtained from $\tilde{z}$ by discarding certain data: $\tilde{y} \sqsubset \tilde{z}$. A simplified version of the corresponding Ramsey sentence thus runs as follows:

$$\exists\, x \left[x \in M \wedge r(x) = z \wedge \tilde{y} \sqsubset z \right].$$

In a more realistic account, one would have to substitute the Ramsey sentence stated above by an approximate version of that claim.[14] Within the structuralist framework, an approximate formulation of the Ramsey sentence can be obtained by making use of certain topological concepts. However, I shall not go into any further detail here but instead turn to a problem which is of fundamental importance for Cartwright's approach.

When Cartwright says that 'the theory matches an equation', what is meant by "the theory"? Or, translated into the structuralist framework, what is the set of models *M*? By what laws is *M* distinguished?

A first option consists in identifying the theory with its fundamental principles. If this option is chosen, it should be noted that at least in the case of Newton's theory these fundamental laws are comparatively weak requirements in the following sense: Newton's axioms as regarded in isolation do not suffice in order to provide a theoretical description of most intended applications of classical mechanics but have to be combined with suitable special laws instead.

A second option consists in identifying a theory at a certain stage of its development with a more or less comprehensive set of laws that contains both its fundamental principles as well as special laws, which are adapted to fit certain special types of intended applications. In the case of Newton's theory, this set may, among others, contain a version of Newton's law of gravitation adapted to the two-body problem, Hooke's law, or some law describing frictional forces.[15]

The second option seems to be what Cartwright has in mind. Besides the fundamental principles, the theory contains an arsenal of special laws, which are used for the theoretical treatment of special types of systems. Thus, in order to provide a theoretical treatment of some concrete system, one will try to find a type of phenomena (like the free fall, the harmonic oscillator, etc.) under which this concrete system can be subsumed, and for which the theory—understood in the sense described before—"matches an equation". However, in order to achieve this task, numerous idealizations and approximations will have to be carried out, thus leading from the unprepared to the prepared description. In this context, at first Cartwright treats this arsenal of special laws ("equations") available for this task as being fixed but soon cautiously adds that this '. . . is of course a highly idealized description. Theories are always improving and expanding. . . .' (Cartwright 1983: 134).

However, the fact that the set of (special) laws available within the framework of some empirical theory is constantly changing is not just a minor point. By contrast, it is of crucial importance for Cartwright's approach. Due to the dynamics of empirical theories, there are two ways a theory can be made to fit a set of data:

1. The first way consists in choosing a prepared description to which the theory—at a certain stage of its development—matches an equation. As we have seen in the case of Halley's Comet, the price tag fixed to this procedure may be a loss of empirical adequacy: *fitting facts to equations*.
2. The second way consists in modifying the theory instead: Special laws that are already available may be changed or new ones proposed. Countless special laws have been put forward on a trial basis just because an empirical theory failed to provide a successful theoretical treatment of some system, and the scientists involved refused to manipulate their data to make this happen: *fitting equations to facts*.

Thereby, it seems plausible to assume that in general scientists will choose the former approach in the beginning—especially when they are dealing with a well-established theory with an impressive set of special laws. Only after these attempts have failed will new laws be proposed or old ones modified.

However, there may be cases, in which new laws are already proposed and tested at an earlier stage. The astronomers who were involved in the investigation of Halley's Comet readily illustrate this point. As I mentioned before, in 1757 Alexis-Claude Clairaut provided a refined theoretical description of the orbit of this object. Ten years before, he had already tried to apply Newton's gravitational law to lunar motion. In doing so, he was confronted with a (presumed) anomaly: The available data were not matched by an equation provided by the theory at that time. In order to eliminate this anomaly, and thus to make the theory applicable to this system, he did not change the available database but proposed the following modification of Newton's law instead:

$$F = G \cdot \frac{m_1 \cdot m_2}{r^2} + \frac{a}{r^4}$$

However, just a few years later he succeeded in showing that the presumed anomaly did not exist at all: The conflict between Newton's gravitational theory and the available data was due to an oversimplified mathematical treatment of that system. Clairaut's gravitational law stayed "off duty" for more than 100 years: a special law without one single intended application to which it could be successfully applied. However, it was reactivated in the course of the different attempts to deal with a much more severe anomaly: the anomalous advance of Mercury's perihelion—again without success (Gähde 1997). Similarly, dozens of alternative gravitational laws were proposed during the nineteenth century, amongst them several velocity-dependent gravitational laws that were constructed in strict analogy to electrodynamic force laws. One of these laws was proposed by Zöllner, who in his book *Ueber die Cometen* (1872) tried to use it to obtain a more precise treatment of the comet's paths. In all these cases new theoretical tools were developed in order to deal with systems that had previously refused to be theoretically "harnessed".

Let me point to one additional problem that I have with Cartwright's approach: I do not see how the proposed borderline between theoretical laws and phenomenological laws can be drawn in my case study. Is Newton's gravitational law to be counted as a theoretical law or a phenomenological law? What about Clairaut's gravitational law—which was constructed to fit certain observational data? If special force laws are counted as fundamental laws, I do not see what phenomenological laws are involved here. If, by contrast, special force laws are counted as phenomenological laws, I do not understand Cartwright's claim that phenomenological laws, and not the fundamental principles of the theory in question, bear the main burden with respect to the theoretical description of concrete intended applications. In the case of classical mechanics, it is precisely through the cooperation between fundamental laws and special laws that a detailed description of a wide variety of systems becomes feasible.

In classical mechanics, fundamental laws and special force laws do not by any means contradict each other. This fact is essential for an adequate understanding of the dynamics of this theory: As we have seen, Newton's laws are comparatively weak requirements. This provides the necessary scope for the prepared formulation of new special laws in order to adapt the theory to new or modified sets of data. If one special law fails to do the job, another special law can be put forward on a trial basis, without necessarily having to clash with the theory's basic principles. This fact is mirrored within the structuralist concept of theory-nets to which I now turn.

A theory-net consists of one basic element T_0 and a considerable number of specialized theory-elements T_j. Each theory-element consists of a set of models and a corresponding set of intended applications. Let us turn to the basic element first. Its set M_0 of models is distinguished by a set-theoretic predicate that contains the fundamental laws of the theory in question. In the case of Newtonian mechanics, they consist of Newton's three axioms. The corresponding set I_0 of intended applications contains those systems (represented by finite data structures) to which these fundamental laws are to be applied, namely all intended applications of this theory.

From this basic element, more specialized theory-elements can be obtained by means of strict or approximate specialization. Let us illustrate this procedure by the example of the theory most relevant for my case study, i.e. Newtonian mechanics. As we have seen, the force function is underdetermined by Newton's axioms regarded in isolation: They contain requirements concerning the resultant force only but do not specify how this resultant force is composed of different forces of various types. This provides the scope for the formulation of additional force laws in specialized theory-elements. Examples have been mentioned above: Newton's gravitational law, Hooke's law, laws describing frictional forces, as well as combinations of different force laws. By adding these requirements, the set-theoretic predicate that distinguishes the models of the net's basic element is reinforced. Thus, the corresponding set of model M_j will be a proper subset of M_0. Trivially, the more specialized theory-element will only apply to a restricted set of intended applications $I_j \subset I_0$. This nicely fits with Cartwright's view that phenomenological laws, which provide a detailed account of certain phenomena, will only hold in a rather limited number of systems.

The process of specialization may be reiterated, thus leading to more and more complex theory-nets. These nets may well contain dozens of theory-elements, thus enabling refined and differentiated theoretical descriptions of the corresponding sets of intended applications.

Let us assume that some concrete system is to be described by means of an empirical theory. By making use of the concept of theory-nets, this task can be reformulated as follows. Some specialized theory-element has to be found such that the intended application in question—or the finite data structure by which it is represented, respectively—can successfully be extended into the corresponding set of models. As claimed by Cartwright, in most cases numerous approximations and idealizations will be necessary for this purpose, which will depend on the choice of this special theory-element. If this task can be successfully fulfilled, in Cartwright's terminology, the 'theory matches an equation' for this system (within the error of measurement).

Let us now assume that attempts to describe a certain intended application by means of some specialized theory-element have failed. In that case one will not react by stating that the theory as a whole has failed. By contrast, other specializations of the net's basic element which seem to be more suitable for that task will be considered instead. Some of these specializations might already have been part of the theory-net before these attempts, while others are newly created for that purpose (such as in the case of Clairaut's law). In other words, the problematic application starts moving through the net. We have analyzed this process in detail with respect to the various attempts to obtain a theoretical description of the anomalous advance of Mercury's perihelion (Gähde 1997, 2002).

Which approximations and idealizations are used strongly depends on the precision of measurement. If the error of measurement is large, many theoretical descriptions that are compatible with the available data will be found. At this stage, the theory by no means unambiguously dictates what equations are to be used. However, if the measuring precision increases, the set of acceptable options for theoretically describing the system in question will decrease: One alternative after the other will be eliminated. At the same time, the standards for providing a prepared description will become increasingly vigorous: Only minor idealizations and approximations will be allowed.

The history of the discovery and investigation of Mercury's anomaly readily illustrates this point: In the mid-nineteenth century, highly precise values concerning the ephemerides of the inner planets were available, and it was only due to this fact that the anomaly could be discovered in the first place. Consequently, astronomers were unwilling to accept any mathematical description of this phenomenon, which was based on a "prepared description" that was not in accordance with the available measurements within the (small) error of measurement. This was the case even though this would have enabled the phenomenon "to be brought into the theory", or, to put it in structuralist term, to find a theory-element already available within the net such that the modified data structure could be extended into the corresponding set of models.

CONCLUDING REMARKS, WITH SOME QUESTIONS FOR NANCY CARTWRIGHT

The considerations presented in the preceding sections should suffice to show that there are numerous parallels to as well as some interesting differences between Cartwright's view on the relationship between theories, models, and their application to reality on the one hand and the structuralist view of these issues on the other.

Let me summarize some basic parallels: Both approaches start from the assumption that for an adequate conception in the philosophy of science it does not suffice to focus exclusively on the logical structure and formal apparatus of empirical theories. By contrast, they emphasize that it is crucial to analyze how in scientific practice concrete, complex segments of reality are described and which role empirical theories play in this enterprise. For this purpose, both approaches make use of elaborate concepts of "models"—albeit ones which are laid out in very different ways. Furthermore, both approaches agree that for the task of modeling concrete systems, the fundamental principles of an empirical theory (alone) either do not suffice (structuralism) or are even inadequate (Cartwright). By contrast, they insist that more specialized laws, which generally cannot be (strictly) derived from the theory's fundamental principles, must be used. These laws are especially tailored to fit certain types of intended applications, thus enabling a detailed, realistic description of these systems, but they will not hold in others. Furthermore, both approaches agree that in modeling real, complex systems more than one theory may—and in many cases will—be involved.

Let us now turn to some major differences between the two approaches. One of the main differences refers to the underlying concepts of "models". The structuralist approach can be seen as a special variant of the so-called semantic view, according to which models are constitutive for theories, or to be more precise, make a crucial contribution to the constitution of theories.[16]

By contrast, at least in her more recent papers, Cartwright does not regard models as constituting theories but as 'mediating between theories and the world' (Cartwright 1999b: 179). A second major difference consists in the fact that only cases of theory-driven modeling can be handled by the structuralist framework. By contrast, Cartwright fights this "theory-dominant" view. She insist that theories are but one tool among others used in the construction of models, and that there may even be cases of modeling in which no theory at all is involved. These cases—if they really exist—cannot be mirrored in the structuralist concept as it stands. Another major difference refers to the way in which both approaches try to account for the fact that in modeling concrete systems numerous empirical theories may be involved in highly complex ways. In Cartwright's approach this is expressed by the fact that representative models—which mirror real systems—will not generally constitute interpretative models of some empirical theory but have to be seen as complex entities, in which ingredients from different empirical theories, mathematical techniques, as well as other objects from the toolbox of science, may be involved. Although I cannot go into any technical detail here, it should be mentioned that the structuralist concept tries to account for this fact in a different way: Here models are described as (interpretative) models of an empirical theory, which, however, is closely interrelated with other empirical theories via constraints and links. They provide a detailed account of intertheoretical relations, as well as of the role these relations play in modeling concrete systems. Finally, another important difference between the two concepts consists in the fact that the structuralist view supplies metatheoretical tools for a very detailed description of the internal logical structure of empirical theories. As we have seen in my case study, this may become crucially important when dealing with the dynamics of empirical theories.

In spite of these differences, the parallels between the two approaches mentioned above suggest that both views can learn a lot from each other. Some questions for Nancy Cartwright conclude this section; they may help to illustrate some of her most basic concepts by means of my case study and help to clarify their relationship to related concepts as put forward in the structuralist view.

1. How is Halley's original model of the comet's path to be classified with respect to the distinction between representative and interpretative models as outlined in Cartwright (1999b)? Is it to be classified as a representative model that is not an interpretative model at the same time? Or is it to be classified as an interpretative model (in spite of the correctional term that occurred in it)?
2. If the former is the case: can the development of successively more refined models for the description of this system be seen as a (gradual) transition to interpretative models—at least as long as no other theories (like thermodynamics etc.) became involved?

3. Where exactly is the line between fundamental principles of physical theories and phenomenological laws (as used in representative models) to be drawn? Are the numerous modified gravitational laws to be interpreted as fundamental principles or as phenomenological laws?
4. If they are to be interpreted as fundamental principles: what are the phenomenological laws used in the treatment of this system?
5. If they are to be interpreted as phenomenological laws: what does it mean to say that these phenomenological laws come closer to the truth than the fundamental laws—if they can only be applied in combination with the fundamental laws?
6. Cartwright (1983) claimed that 'The most apparent need is to write down a description to which the theory matches an equation.' What was meant by "the theory" in this context? What would be the analogue in structuralist terms? The basic element of a theory-net? Or a specialized theory-element? Or the whole net at a certain stage of its development?
7. In 'For phenomenological laws' (1983: 100–127), Cartwright claimed that the approximations are not dictated by the facts. At first glance this claim seems to be corroborated by the case study of Halley's Comet: Numerous alternative approximations could be successfully applied within the error of measurement. However, the history of the different attempts to describe the movement of comets seems to suggest that the available measuring precision determines the admissible standard of approximation. When improved measurement procedures become available, the set of admissible approximations may shrink. May not—in this sense—approximations become determined by the facts?

NOTES

1. I would like to thank all the participants of the workshop on 'Nancy Cartwright's philosophy of science', above all Nancy Cartwright herself, for stimulating discussion. My thanks are also due to Roman Frick, Ron Giere, Stephan Hartmann, and Iain Martel for their helpful comments on earlier drafts of this chapter.
2. In this early publication, Cartwright still starts from the assumption that theories play *the* central role in modeling certain segments or aspects of reality—a view which seems especially appropriate with respect to my case study, in which Newton's gravitational theory is applied to a concrete astronomical system. However, in more recent publications Cartwright stresses the point that theories are but one tool among others in the construction of models and that, in some cases, no theories at all might be involved in that task. I return to this point in the section on The "Prepared Description".
3. In order to be fair, it should be noted that Cartwright herself describes her characterization of the unprepared description as a "gross exaggeration", (Cartwright 1983: 133).
4. However, their properties will not be *strictly* identical. An example: the comet will lose part of its mass when passing near the sun.

5. The apparent brightness is the luminosity of an object when observed from the earth. By contrast, the absolute brightness is the luminosity the object would have when observed from a standard distance of 1 parsec (= 3.26 light-years).
6. For a more detailed account of this process compare the section on The "Mathematical Description".
7. For an analysis of the theoretical–nontheoretical distinction (Gähde 1990).
8. (Sneed 1979).
9. Let z, $\tilde{z}$ be two tuples with the same number of components. Then $\tilde{z}$ is a (finite) substructure of z ($\tilde{z} \sqsubset z$) iff each component of $\tilde{z}$ is a (finite) subset of the corresponding component of z. M_{pp} is the set of all partial models of the theory in question.
10. Similar formulations can be found in Cartwright 1999b: Ch. 2.
11. However, one might argue that this process was reversed in later attempts to model the movement of Halley's Comet. When measuring precision improved, successively more theories—thermodynamics, in particular—became involved in the construction of these models. One may try to interpret this as an illustration of Cartwright's view on modeling in a "dappled world".
12. With respect to the meaning of "relevant data" cf. the section on The "Unprepared Description".
13. Which, according to the original structuralist credo, are believed to be theoretical with respect to that theory.
14. Furthermore, one would have to account for the bridge structures (constraints and links) that connect the theoretical description of this system with the theoretical description of other intended applications of this, as well as of other, empirical theories (Gähde 2002: 77).
15. Note that in this account Newton's law of universal gravitation is not regarded as a fundamental law but a special law. The reason for this classification is that there are numerous intended applications of classical mechanics that are theoretically described without making reference to this law (e.g., elastic and inelastic collision processes, etc.).
16. According to the structuralist view, empirical theories cannot simply be identified with their set of models. By contrast, the explication of the concepts of *theory-element* and *theory-net*—which serve as substitutes for the informal and vague term *theory*—refers to numerous other entities as well, among them data structures, constraints, and links.

REFERENCES

Bailer-Jones, D. (this volume) 'Standing up against traditions: Models and theories in Nancy Cartwright's philosophy of science'.

Balzer, W. et al. (1993) 'A model for science kinematics', *Studia Logica*, 52: 448–519.

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Oxford University Press.

———. (1991) 'Fables and models', *Proceedings of the Aristotelian Society Suppl.*, 65: 55–68.

Cartwright, N. et al. (1995) 'The tool box of science: Tools for the building of models with a superconductivity example', in W. E. Herfel et al. (eds) *Theories and Models in Scientific Processes*, Amsterdam: Rodopi.

Cartwright, N. (1999a) 'Models and the limits of theory: Quantum Hamiltonians and the BCS model of superconductivity', in M. Morgan and M. Morrison (eds) (1999) *Models as Mediators*, Cambridge: Cambridge University Press.

———. (1999b) *The Dappled World. A Study of the Boundaries of Science*, Cambridge: Cambridge University Press.

Gähde, U. (1990) 'On inner theoretical conditions for theoretical terms', *Erkenntnis*, 32: 215–233.

———. (1997) 'Anomalies and the revision of theory-nets. Notes on the advance of Mercury's perihelion', in M. L. Dalla Chiara et al. (Hg.) *Structures and Norms in Science*, Dordrecht: Kluwer.

———. (2002) 'Holism, underdetermination, and the dynamics of empirical theories', *Synthese*, 130: 69–90.

Hoskin, M. (ed.) (1997) *The Cambridge Illustrated History of Astronomy*, Cambridge: Cambridge University Press.

Sneed, J. D. (1979) *The Logical Structure of Mathematical Physics*, 2nd edition, Dordrecht: Kluwer.

Zöllner, F. (1872) *Über die Natur der Cometen*, Leipzig: W. Engelmann.

Reply to Ulrich Gähde

The German structuralists undoubtedly offer the most satisfactory detailed and well illustrated account of the structure of scientific theories on offer, and Gähde's work on the relations between theories has added considerably to it. Their account of theory structure also has a place for the relations of theory to the world, and it is here that one of my chief concerns intersects theirs. As Gähde notes, the significant difference between us is that I suppose that the starting "unprepared" description to which we will match a representative model contains 'any information thought to be relevant', whereas for the structuralist it contains only data that are 'part of the model to which the corresponding finite data set is to be extended' (Gähde this volume: 7).

The difference depends on what the models look like, to which these starting descriptions are to be extended. Consider my controversial example of Neurath's bill, which is blown by the wind as well as attracted by gravity. Both the structuralist and I will have to include mention of the wind if the bill is to get treated by Newtonian mechanics.[1] Assigning a force function to the wind as well as to the pull of gravity will allow us to treat the acceleration of the bill using $f_t = ma$.[2] But I claim that for many cases any such assignment for the wind will be ad hoc, and if it is ad hoc, then the *form* will be that of Newton's law, but $f_t = ma$ will no longer be the very law that has such a vast army of empirical success to support it. For these successes, I maintain, arise from a law in which f_t has more constraints on its application: Its form must be licensed from the features of the prepared description by bridge principles of the theory.

From the point of view of a structuralist formalization of Newtonian theory, the issue hinges on what kinds of models we take to be in our reconstruction when we are careful to ensure that the formalized theory is something that is indeed well confirmed. My claim from sampling the kinds of treatments that provide this high degree of confirmation is that this theory should not include models with mathematical force functions that contain ad hoc terms. So it must be possible to break down the force function in any model into components, where for each component there is a further description in the model that is associated to that mathematical form by a bridge principle. If the structuralists' models satisfy this constraint, then the theory they comprise can indeed be regarded as well confirmed. But then it may not include a model that covers the motion of the bill.

Turning now to Gähde's questions:

1. Few adequate representative models *are* interpretive models. Rather, when the force functions are assigned in a non-ad-hoc manner, the representative model will, for each component term in the force function, *contain* an interpretive model that is linked to that component by a bridge principle. Halley's treatment of Jupiter was ad hoc; this is reflected in the fact that the representative model he constructed for the comet had terms in the force function not linked by accepted bridge principles to interpretative models that were part of the overall model representing the situation.
2. Yes, the more refined models can be seen as a transition to a representative model that includes descriptions (interpretative models) that license each term in the force function.
3. There is no exact line. My basic point is that the less we constrain our vocabulary the more we are likely to be able to describe accurately what happens. The unamended law "The force of attraction due to gravity between two bodies of mass *m* and *M* separated by *r* is GmM/r^2" is a well-confirmed bridge principle. The amended claim that the force of attraction due to gravity between *m* and *M* is $GmM/r^2 + a/r^4$ is an unconfirmed hypothetical bridge principle.

4&5.I am not sure I would describe anything Halley used as "phenomenological". I think it would be more insightful to describe his equations as fundamental laws with ad hoc amendments to allow calculation.

6. For proper, non-ad-hoc, coverage the description must be matched by an equation from any bit of the theory net. What matters is that the theory-net be constructed so that we have good reason to take it to be well confirmed.
7. Yes. By definition *ad verum* approximations are not licensed by the facts. But the facts to be accounted for, and especially their precision, will certainly constrain what kinds of approximation are acceptable.

NOTES

[1] Because of this I shall stick with my claim that the chief criterion for acceptability of the unprepared description is empirical adequacy: It must include everything "relevant" and by that I mean everything that actually matters to the phenomenon under study. Theory will of course inform our judgements about what the relevant factors are, but the fact that a factor is not describable in the theory or relevant by its lights must not eliminate relevant factors.

[2] Here f_t is total force.

4 Models as Representational Structures

Margaret Morrison

Many of us who were involved in the LSE modelling project with Nancy Cartwright share more or less similar views regarding the position models occupy within the theoretical hierarchy and the role they play within that framework. Taking the notion of a model as 'mediator' between theory and the world as a starting point, several of us have tried to articulate specific details of our own views about various aspects of modelling in the natural and social sciences. One of the things I want to do in this chapter is flesh out, in a bit more detail, ideas presented in Morrison (1999) and Morrison and Morgan (1999) about the relation between a model that represents a physical system (a representative model) and its role as a mediator.[1] There are many different ways that models can function as a mediator. It can mediate between theory and the world in the sense of being an abstract representation of a physical system governed by one or more theories, or it can be a concrete representation of some feature of an abstract theory. The pendulum is an example that covers both of these cases. In the case of theory application we have the ideal pendulum which represents harmonic motion, and we also have the physical pendulum which is modelled by making various corrections to the ideal case.

A model can also function as a mediator in its role as the 'object' of inquiry. In other words, the model itself rather than the physical system becomes the thing being investigated. In that context it serves as a source of 'mediated' knowledge either because our knowledge of the physical system is limited, or the system is inaccessible. Hence we know only how the model behaves in certain circumstances. The quark model of elementary particles or various kinds of cosmological models are cases in point.

In order to flesh out these ideas about how models represent physical systems and how they can act as sources of mediated knowledge I want to examine Cartwright's account of representational and interpretive models as outlined in her paper "Models and the Limits of Theory: Quantum Hamiltonians and the BCS Model of Superconductivity"(1999). My disagreements with her are not about the *function* of interpretive models, but rather with the lack of what I see as a crucial role for representative models in the development of the BCS account of superconductivity.[2] Since it is sometimes said that the devil is in the details it is important to point out how the details buried in the BCS paper can be put to work to highlight the significance of representative models for the process of theory construction.[3] To that extent my disagreement centres on the details of the case which, in turn, reveals methodological disagreements about the role and importance of different kinds of models. I want to claim that before interpretive models can do their job there first needs to be a representative model in place that provides a physical/causal account of the phenomenon in question.[4] In this case such an account would tell us something about the causal aspects of superconductivity, what the important mechanisms are in its production, etc., etc. In summary then, the differences between my account and Cartwright's centre on the importance of representative models, how representation should be understood within the modelling context, and the role representative models play in the context of the BCS theory.

I begin with a short description of Cartwright's views on this issue, my disagreements with it and follow with a discussion of how I want to approach the issues of representation and mediation and why representation is crucial for our understanding of how scientific models deliver information.

CARTWRIGHT ON HOW MODELS REPRESENT

Cartwright claims that theories in physics do not generally represent what happens in the world; only models represent in this way and the models that do so are not already part of any theory.[5] Fundamental principles of theories in physics do not represent what happens because theories give purely abstract relations between abstract concepts like force; they tell us the capacities of systems that fall under these concepts. Those systems need to be "located in very specific kinds of situations" (1999, 242) in order for their behaviour to be fixed; and when we want to represent what happens in these situations we need to go beyond theory and construct a representative model. However, many theories (e.g. QM, QED) require more than this; the abstract concepts need "fitting out" in more concrete form before representative models can be built in a principled or systematic way. This fitting out is done by interpretive models that are laid out within the theory itself—in its bridge principles. An example of such an interpretive model for classical mechanics is 'two compact masses separated by a distance *r*' (1999b, 3). It is the job of these models to ensure that abstract concepts like force have a precise content.

In contrast to this, I want to claim that the story is exactly the reverse. While it is true that the interpretive models are prior in the sense of being already part of a background theoretical framework (quantum mechanics, in the case of BSC superconductivity), one needs a representative model before we can determine how the abstract concepts/theory are going to be applied in a specific situation. In other words, we need a representation of the physical system to which we can then apply the abstract concepts via the interpretive models. My justification for this story will emerge in the discussion of the BCS theory but first in order to fully understand Cartwright's distinction between representative and interpretive models we need to look at her characterization of models and theory.

Many characterizations of model construction (e.g. Giere 1988; McMullin 1985) describe the process as involving the addition of refinements and corrections to the laws of the theory that we use to represent, in an idealised way, certain aspects of a physical system. In the case of the pendulum we can add, among other things, frictional forces in order to incorporate more details about the behaviour of real pendulums. Moreover, many view the addition of corrections as involving a cumulative process with the model coming closer and closer to an accurate representation of the real system. By contrast, Cartwright claims that the corrections needed to turn models provided by theory into models that accurately represent physical systems are rarely, if ever, consistent with theory, let alone suggested by it (1999, 251). For example, she claims that only pendulums in really nice environments will satisfy Galileo's law even approximately; real pendulums are subject to all types of perturbing influences that do not appear to fit the models available in Newtonian theory. Instead when we attempt to correct the model so that it better fits with the real system we produce a blueprint for a nomological machine that generates trajectories satisfying some complex law together with Newton's force law.

What exactly is a nomological machine? Cartwright gives the following definition: "It is a fixed (enough) arrangement of components, or factors, with stable (enough) capacities that in the right sort of stable (enough) environment will, with repeated operation, give rise to the kind of regular behaviour that we represent in our scientific laws (1999b, 50). But, in order for the theory, in this case Newtonian mechanics, to function as a nomological machine it requires a number of components with "fixed capacities arranged appropriately to give rise to regular behaviour"(253). These components and their arrangements are given by the interpretive models of the theory as in the mass point bob of the pendulum, a constraint that keeps it swinging through a small angle along a single axis, the earth to exert a gravitational pull, and other factors used to customize the model. In that sense then the model is the blueprint, the theory is the machine.

As I understand Cartwright the general picture is something like this. The quantum Hamiltonian and the classical forces are abstract concepts whose application requires a more concrete description that tells us how to understand them in specific cases. When we assign a gravitational force to a model we describe a certain mass *m* as subject to a gravitational force GMm/r^2 located a distance *r* from a second mass *M*. In other words, we apply the notion of gravitational forces using interpretive models that provide more concrete descriptions that involve distances and masses.[6] Although Cartwright doesn't describe the situation in exactly this way her point seems to be that we can think of the concrete description as spelling out what the gravitational force law means, that is, providing a kind of *representation* of a system (a model) where the law is applied. Although the theory gives us interpretive models that can be applied to concrete systems, this applicability often does not extend very far. In most cases there are factors relevant to the real world situations that cannot be had from theory via its interpretive models, so we need to go beyond them to get more precise descriptions and predictions. In other words, the interpretive models provide concrete descriptions for more abstract concepts but their domain of application with respect to realistic situations is limited, more concrete details are needed if the situation is to be described in a reasonably accurate way. At this point the representative model takes over and furnishes those details.

Hence, interpretive models can also be seen as having a limited representational role; they can represent some aspects of a physical system, but only up to a point. But, what about the representational models themselves; how do we take account of their function (except to say that they provide a more realistic picture of the phenomenon)? Cartwright herself claims that she has little to say about how representative models represent, except that we should not think in terms of analogy with structural isomorphism; something I wholeheartedly agree with. Again, Cartwright takes us to the domain of the interpretive model as a way of illustrating what the notion of representation *isn't*. She warns against thinking of the models linked to Hamiltonians as picturing individually isolatable physical mechanisms associated with the kinetic energy term plus the Coulomb interaction; in other words, we don't explain Hamiltonians by citing physical mechanisms that supposedly give rise to them. Instead Hamiltonians are assigned via quantum bridge principles for each of the concrete interpretive models available in quantum theory such as the central potential, harmonic oscillator, scattering events and Coulomb interactions. In that sense the theory extends to all and only those situations that can be described by the interpretive models/bridge principles. In some cases ad hoc Hamiltonians are used but these are not assigned in a principled way from the theory and hence the theory receives no confirmation from the derived predictions.

The issue of representation here is not simply one of picturing. In the case of quantum field theory we model the field as a collection of harmonic oscillators in order to get Hamiltonians that give the correct structure to the allowed energies. But, Cartwright points out, this does not commit us to the existence of a set of objects behaving like springs. Nevertheless she appeals to a loose notion of resemblance, claiming that models resemble the situations they represent where the idea of a correct representation is not defined simply in terms of successful prediction but requires independent ways of identifying the representation as correct.

This, of course, is the crux of the problem of representation—in virtue of *what* do models represent and how do we identify what constitutes a correct representation? Cartwright doesn't elaborate on how this identification might proceed. So, in order to sketch an answer I think we need to first look at some different *ways* in which models represent. As I mentioned at the outset, what I mean when I say that models function as mediators between theory and the world is that part of their task is to represent each domain. In other words, sometimes models represent theory by providing a more or less concrete instantiation of one of its laws, as in the case of the pendulum considered as a representation of a Newtonian force law. But the model pendulum can also represent a real physical pendulum. The degree of approximation in the latter case (i.e. how realistically the models represents the physical apparatus) will depend, in part, on what we need the model to do for us (see Morrison, 1999). We also use models to represent physical situations that we are uncertain about or have no access to. An example would be models of stellar structure.

As a way of illustrating some aspects of this representational feature of models I want to focus on the BCS theory of superconductivity. I do so not only because it will help to illustrate how my views diverge from Cartwright's, but also because it provides a nice example of how models and theories interact in the representation of concrete systems. I draw particular attention to what I take to be the representative model[s] that were prominent in the construction and development of the BCS theory, particularly the model of electrons as Cooper pairs.[7] My claim is that the physical account of Cooper pairs functions as *the* representative model that ultimately forms the foundation for the BCS account of superconductivity. It does this by providing an account of how electrons need to be *represented* in order for superconductivity to be understood.[8]

Before going on to discuss the specifics of the case let me recap the important philosophical points. As I mentioned above, part of my disagreement with Cartwright's account is that it tells only half the story, much more needs to be said about how representative models provide the framework for establishing a physical 'picture' that can then be treated mathematically. To that extent both representative and interpretive models have a representative function. As I see it, and taking account of Cartwright's discussion of interpretive models, the latter represent theory while the former purport to represent physical systems or aspects thereof. The question that naturally arises here is whether there is anything specific about the structure of a representative model that differentiates it from other types of models. My own view on this is that all models represent in one way or another. Some provide structural representations of physical systems, as is the case with certain kinds of nuclear models (see Morrison, 1998); others, like the pendulum model (Morrison, 1999) provide more fleshed out versions of how physical systems are constituted, as well as providing a 'representation' in terms of an application of Newtonian mechanics.

Since the representative/interpretive distinction is Cartwright's I will stick to this way of characterizing the models involved in the development of BCS. However, as I said above, an important difference between her account and mine is that I see the representative rather than the interpretive models as prior. One reason for this is because representative models act as the mediator between the theory and the world and function as the source of "mediated" knowledge. That is to say, representative models typically supplant the physical system as the object treated by the theory's interpretive models. Hence, our knowledge in this context is "mediated" because it comes via the representation we have constructed; it gives us a physical picture of how the system we are interested in might be constituted. With this in hand we can then see whether the interpretive models of the theory can be applied in the appropriate kinds of ways. The types of interpretive models that Cartwright mentions (harmonic oscillator, Coulomb potential, etc.) function across all contexts where quantum mechanics is applied. In that sense there is nothing specific about their application in particular instances, except in the construction of the appropriate Hamiltonian. Although this is a significant aspect of the function of interpretive models I want to claim that even this is highly dependent on the kind of representative model we have constructed.

ON THE PRIORITY OF REPRESENTATIVE MODELS

Cartwright begins her discussion of the BCS paper (Bardeen et al. 1957) by highlighting two important steps in its development.The first is the attractive potential due to electron interactions via lattice vibrations and the second is the notion of Cooper pairs. Together these ideas suggested that there was a state of lower energy at absolute zero than the one in which all the levels in the Fermi sea are filled; a state that could be identified as the superconducting state. Cartwright claims that the first job of the BCS paper was to produce a Hamiltonian for which such a state will be the solution of lowest energy and then to calculate the state. The important derivations in the BCS paper were based on a reduced Hamiltonian with only three terms, two for the energies of the electrons moving in a distorted periodic potential and one for a simple scattering interaction. The longer Hamiltonian introduced earlier in the paper also uses only the basic models referred to above (kinetic energy of moving particles, the harmonic oscillator, the Coulomb interaction and scattering between electrons) plus one more, the 'Bloch' Hamiltonian for particles in a periodic potential (which is closely related to the basic model of the central potential). Cartwright claims that "superconductivity is a quantum phenomenon precisely because superconducting materials can be represented by the special models that quantum theory supplies" (266).

But what exactly is being represented here and what kind of representation is it? From Cartwright's discussion it would seem that the answer is relatively straightforward—we use the quantum models as a way of representing the processes responsible for superconductivity. While I agree with this as a minimal claim, it leaves out an important part of the story about how superconductivity is represented. The situation Cartwright describes is a quantum theoretical representation that involves an application of theoretical (interpretive) models mentioned above to a particular kind of process. What licences this kind of representation is theory (QM); but my point is that in order to apply these models we first need some fundamental idea about *how* the phenomenon of superconductivity occurs. This we obtain from a representative model that describes the causal mechanism required for producing superconductivity. In other words, we need an initial representation that explains how the electrons in the metal can give rise to the energy gap characteristic of superconductivity, and how their behaviour can be accounted for within the constraints imposed by quantum mechanics, specifically the exclusion principle. Once this representation is in place we can then go on to use Cartwright's interpretive models to give a full theoretical representation. Very briefly, the main difference between us at this point is the order of priority of representative vs. interpretive models. I claim that the former are necessary before the latter can even be applied while Cartwright seems to suggest that the interpretive models are where one ought to locate the crux of the BCS account. I develop the details of my story below but before doing that it is important to see how Cartwright approaches the problem of constructing the BCS Hamiltonian, a crucial feature for both her account and mine.

A good deal of Cartwright's discussion is directed at the construction of the BCS Hamiltonian, which she claims uses the collection of stock models/Hamiltonians provided by quantum theory, and hence, has "four very familiar terms" (1999, 268). The first two represent the energy of a fixed number of electron-like particles with well-defined momenta. The third term represents the pairwise Coulomb interaction among these particles and the fourth, the interactions occurring between pairs of electrons through exchange of a virtual phonon. But, in addition to knowing the form of the Hamiltonian one also needs to know, with respect to the first two terms, the allowed values for the momenta of the electrons in the model. To justify a particular choice requires that we fill in more details about the model. The structure of the model, however, only becomes clear as the third and fourth terms are developed. Cartwright goes on to describe how each of these terms is justified showing how the BCS Hamiltonian is both theoretically principled and phenomenological or *ad hoc* (269–78). In other words, it is not always the case that we have an established bridge principle that links a given Hamiltonian with a specific model that licenses its use. She claims that because the BCS Hamiltonian is not simply assigned in a principled way there are additional assumptions required that are not built in as explicit features of the model. But these features cannot then be used in a principled way to put further restrictions on the Hamiltonian. What, then, is the origin of these assumptions?

Crucial to the construction of the Hamiltonian is an assumption involving restrictions on which states will interact with each other in a significant way. As Cartwright points out the choice here is motivated by physical ideas associated with Cooper's notion of paired electrons (269). BCS assume that scattering interactions are dramatically more significant for pairs of electrons with equal and opposite momenta. Consequently the assumptions about what states will interact significantly are imposed as an Ansatz, "motivated but not justified and ultimately judged by the success of the theory at accounting for the peculiar features associated with superconductivity" (269). In relation to this Cartwright claims (in a footnote) that because of these ad hoc features it makes sense to talk of both the BCS theory and separately of the BCS model since the assumptions made in the theory go beyond what can be justified using "acceptable quantum principles from the model that BCS offer to represent superconducting phenomena" (ibid.). Indeed she claims that the principled-ad hoc distinction "depends on having an established bridge principle that links a given Hamiltonian with a specific model that licenses the use of that Hamiltonian" (271).

The story I want to tell is, in many ways, at odds with this characterization of the principled/ad hoc distinction. My disagreement with Cartwright's account stems primarily from the role she assigns to the "physical ideas" related to Cooper pairing. What the above quote suggests is that the only principled constraints placed on the construction of the Hamiltonian are those that come directly from theory (QM) while the ideas that appeal to Cooper pairing are ad hoc in the sense that they lack this 'theoretical' justification. However, to my mind such a distinction undermines the role of representative models by suggesting that they are of secondary importance in the development of the BCS theory/model.[9] Instead, I want to argue for their *primacy* by showing that the success of the BCS paper lay in its ability to demonstrate that the basic interaction responsible for superconductivity is the pairing of electrons by means of an interchange of virtual phonons; a demonstration made possible in virtue of the representative model.[10] In fact, one of the important features of the 1957 paper was the construction of the "pairing" Hamiltonian. One can summarize the fundamental postulate of the BCS theory as follows: superconductivity occurs when an attractive interaction between two electrons via phonon exchange dominates the usual repulsive screened Coulomb interaction. The fact that assumptions about Cooper pairing and accompanying assumptions about interacting states are not are not derived directly from quantum theory does not make them ad hoc, unless of course, one classifies the construction of representative models as itself an ad hoc process. Had BCS been unable to provide a quantum theoretical treatment for the pairing hypothesis/model then one could legitimately claim it to be ad hoc. However, it was firmly established as a quantum phenomenon and as such played a vital and systematic role in the BCS paper.

THE EVOLUTION OF AN IDEA

In order to show exactly how the notion of pairing developed by Cooper can be considered a representative model and not just an ad hoc assumption one must first show how it provides a framework for incorporating several of the ideas about superconductivity that were in place at the time. By 1953 phenomenological approaches grounded in experiment had provided an impressive account of many of the phenomena associated with superconductivity. One important result was the verification, using microwave techniques, of the Londons' ideas of rigidity of the wavefunction and long-range order of the average momentum. In addition to the assumptions about the diamagnetic properties of superconductors, London also assumed that the superconducting state could be modelled by representing the superconductor as a giant atom made up of individual atoms, with electrons whirling around its periphery and producing the shielding currents responsible for the Meissner effect. In other words, the superconductor behaved as a single object rather than as a collection of atoms. In order to produce currents there needed to be some order or correlation among all the electrons throughout the 'atom', an order that could be described by a wavefunction.

The Londons were interested in how the diamagnetic property of the equation relating the electrical current density carried by a superfluid to the magnetic vector potential might be related to quantum mechanics. They succeeded in showing that if there was a 'rigidity' of the wave function then their equation could be derived. This rigidity, required to explain the Meissner effect, meant that the wavefunction was essentially unchanged by the presence of an eternally applied magnetic field. In the case of atoms, rigidity arises because of the energy required for the excitations of the system which causes a large diamagnetism in a magnetic field. This led the Londons to suggest that the rigidity of the wavefunction may be due to a separation between the ground state and the excited states; in other words, an energy gap may be present. The idea that a macroscopic piece of matter could have a macroscopic wavefunction was certainly controversial since the application of quantum mechanical ideas of this sort typically involved microscopic objects. No correlation among atoms of the type supposed by the Londons seemed applicable in the macroscopic domain. However, the quantum mechanical picture imposed on the model enabled them to account for the coherent behaviour required for supercurrents. In fact, in his book published in 1950 Fritz London suggested that a superconductor is a quantum structure on a macroscopic scale—a kind of solidification or condensation of the average momentum distribution of the electrons.

We can see that even in this early stage in the development of superconductivity we have a representative model of the superconducting state that involves interaction between electrons in a giant atom. And, it was the attempt to account for this interaction that led to ideas about the rigid wavefunction and a connection with quantum mechanics. As Schrieffer (1973, 24) notes, the momentum space condensation associated with the quantum macroscopic idea was crucial to the BCS paper; indeed "many of the important general concepts were correctly conceived before the microscopic theory was developed". But, and perhaps most importantly, Londons' representative model spawned the idea of an energy gap that would eventually prove crucial in the development of the microscopic theory. My point however is not to delve into the many different lines of thought that went into the early representation of a superconductor; instead I simply want to call attention to the main ideas and how they became incorporated into the notion of Cooper pairing, *the* fundamental feature of the representative model from which BCS arises. In order to see how these ideas fit together we need to briefly explain how the energy gap functions in the larger theoretical picture.

Several theorists including Ginzburg, Frolich and Bardeen himself had all developed models incorporating an energy gap to describe thermal properties. It now became the task of a microscopic theory to explain this gap. That was the problem that stimulated the work of Cooper who investigated whether one could explain, in the context of general theories of quantum mechanics, why an energy gap arises. In ordinary metals one of the basic mechanisms of electrical resistance is the interaction between moving electrons (i.e. electric current) and vibrations of the crystal lattice. However, if there is a gap in the energy spectrum, quantum transitions in the electron fluid will not always be possible; the electrons will not be excited when they are moving slowly. What this implies is the possibility of movement without friction.[11] Work by Frohlich (1950) and Bardeen (1951) pointed out that an electron moving through a crystal lattice has a self energy by being 'clothed' in virtual phonons. This distorts the lattice which then acts on the electron by virtue of the electrostatic forces between them. Because the oscillatory distortion of the lattice is quantized in terms of phonons one can think of the interaction between the lattice and electron as the constant emission and reabsorption of phonons by the latter. The problem however is that the phonon induced interaction must be strong enough to overcome the repulsive Coulomb interaction, otherwise the former will be swamped and superconductivity would be impossible.[12] The problem then was how to account for the strength of the phonon-induced interaction.[13] The solution was Cooper pairs.

We can see here how the representative model initially suggested by the Londons gradually evolved into a more complex account of how one might describe, in qualitative terms, the superconducting state. Although there were many twists and turns along the way, the basic ideas remained in tact (i.e. the energy gap, electron interaction and the connection with quantum mechanics); only the details about how they might be implemented varied. For example, one of the main problems that needed to be taken account of was whether electron-phonon interactions were based on self-energy rather than true interaction between electrons. Let me now go on to show how the notion of Cooper pairs provided a coherent picture—a representative model—from which a quantitative microscopic account could be developed.

Cooper's (1956) paper involves an attempt to determine whether the electron-phonon interactions could give rise to a gap in the one-electron energy spectrum and in particular how one could show this primarily as a result of the exclusion principle. His approach consisted of adding two electrons to the system of electrons filling the Fermi sea. He further assumed that the electrons in the sea are held rigidly in their states so that these states are forbidden by the exclusion principle to the two extra electrons. The problem is somewhat artificial since the electrons in the sea would be scattered above the Fermi surface; however, if this possibility was allowed then one would immediately have to solve a many-body rather than just a two-body problem. Cooper's wavefunction was essentially a solution to a problem that neglected all terms involving operators referring to states within the sea.[14] He worked out the problem of two electrons interacting via an attractive potential $-V$ above a quiescent Fermi sea, i.e. the electrons in the sea were not influenced by V and the extra pair was restricted to states within an energy $\hbar\omega$ above the Fermi surface.[15] He found that two electrons with the same velocity moving in opposite directions with opposite spins had an attractive part that was stronger than the normal Coulomb repulsion. It was this net attractive interaction, resulting in a pairing process of the two electrons, that came to be known as Cooper pairing. As long as the net force is attractive, no matter how weak, the two electrons will form a bound state separated by an energy gap below the continuum states. So, Cooper's account suggested that if one formed the superconducting state out of normal excitations, the matrix elements of the attractive interaction would contribute negatively to the energy (i.e. it would be lowered below that of the normal state) provided the electrons were associated in pairs.

The qualitative physical picture now seemed to be in place; the phonon induced interaction gives rise to Cooper pairing which is responsible for the energy gap. These ideas formed the basis for the representative model of electron-electron interactions and Cooper had shown that the relevant part of the Hamiltonian was that which coupled together the pairs.[16] Even at this point we can see that, far from being *ad hoc*, the pairing hypothesis had a firm quantum mechanical foundation and explained many of the features required for superconductivity. In the remaining part of the paper I want to look at the ways in which this representative model influenced the quantitative account presented in the BCS paper, specifically the construction of the ground state wave equation. I conclude with some remarks about the nature of representation and the import of my disagreement with Cartwright.

FROM REPRESENTATIVE MODEL TO QUANTATIVE THEORY: CONSTRUCTING THE BCS GROUND STATE

Recall that the reduced Hamiltonian expressed only terms for interactions among pairs; the question was: what kind of wavefunction composed of pairs would solve the Hamiltonian. Although the idea that led to the formulation of the wave equation came primarily from Shreiffer, Bardeen also played a significant role. He had been strongly influenced by London's notion of a macroscopic quantum state and by the idea that there would be a type of condensation in momentum space, that is, a "kind of solidification or condensation of the average momentum distribution" [1973a, 41].[17] However, because of the thermodynamic properties of the transition between the normal and superconducting state, Bardeen was also convinced that the condensation was not of the Bose-Einstein type. The strategy was to focus on the states near the Fermi surface (since those were the ones important for the superconducting transition) and set up a linear combination of those that would give the lowest energy. In doing that certain considerations about the nature of Cooper pairs figured prominently.

The electrons that form the bound state lie in a thin shell of width ≈ $\hbar\omega_q$ where $\hbar\omega_q$ is of the order of the average phonon energy of the metal. If one looks at the matrix elements for all possible interactions which take a pair of electrons from any two **k** values in this shell to any two others, one finds that, due to Fermi statistics for the electron, the matrix elements alternate in sign. Because they are also roughly of equal magnitude they give a negligible total interaction energy, that is, a vanishingly small lowering of the energy relative to the normal situation of unpaired electrons. One can however impose a restriction to matrix elements of a single sign by associating all possible k values in pairs ($\mathbf{k}_1$ and $\mathbf{k}_2$) and requiring that either none or both of the members of the pair be occupied. Because the lowest energy is obtained by having the largest number of transitions it is desirable to choose the pairs such that from any one set of values ($\mathbf{k}_1$,$\mathbf{k}_2$) transitions are possible into all other pairs ($\mathbf{k}'_1$,$\mathbf{k}'_2$). Also, because the Hamiltonian conserves momentum it would connect only those pairs that had the same total momentum **K**. In other words, $\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}'_1 + \mathbf{k}'_2 = \mathbf{K}$. The largest number of possible transitions yielding the most appreciable lowering of energy is obtained by pairing all possible states such that their total momentum vanishes.[18] It is also energetically most favourable to restrict the pairs to those of opposite spin. In other words, BCS made the assumption that bound Cooper pairs would still result when all the electrons interacted with each other. We can understand the situation as follows: *At 0°K the superconducting ground state is a highly correlated one where in momentum space the normal electron states in a thin shell near the Fermi surface are, to the fullest extent possible, occupied by pairs of opposite spin and momentum*. Hence their focus on the 'reduced' problem involving only those single electron states that had paired states filled.[19]

The ‘reduced’ Hamiltonian has the following form:

$$H_{red} = \Sigma_{ks}\ \varepsilon_k\ n_{ks} - \Sigma_{kk'}\ V_{k'k} b_{k'}{}^{+} b_k$$

The first term gives the unperturbed energy of the quasi-particles forming the pairs while the second term is the pairing interaction in which a pair of quasi-particles in $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$ scatter to $(\mathbf{k}'\uparrow, -\mathbf{k}'\downarrow)$.[20] An important feature of the Hamiltonian is that the operators $b_k+ = c_k\uparrow^{+}\ c_{-k}\downarrow^{+}$ being a product of two fermion (quasi-particle) creation operators do not satisfy Bose statistics since $b_k{}^{+2} = 0$.[21] The ground state is actually a linear superposition of the pair states, but the question is which one.

In addition to assuming that superconducting ground state energy is due uniquely to the correlation between Cooper pairs, BCS also presuppose that all interactions except for the crucial ones are the same for the superconducting as for the normal ground state at 0°K. The main problem in constructing the ground state wave equation is that one could not use a wavefunction where each pair state is definitely occupied or definitely empty because then the pairs could not scatter and lower the energy. In other words, there had to be an amplitude, say v_k, that $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$ is occupied in Ψ_0 and consequently an amplitude $u_k = (1-v_k{}^2)^{1/2}$ that the pair state is empty. Because a large number, roughly 10^{19}, of pair states $(\mathbf{k}'\uparrow, -\mathbf{k}'\downarrow)$ are involved in scattering into and out of a given pair state $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$, the ‘instantaneous’ occupancy of the other pair states at that ‘instant’ should be essentially uncorrelated with the occupancy of the other pair states at that ‘instant’. In other words, how the pairs interact couldn’t be important; instead what was important was some kind of statistical average; that is, only the average occupancy of the pair states was related. Hence, the wavefunction was a kind of statistical ensemble where pairs were allowed to interact but weren’t strongly correlated. The ground state was a product of operators—one for each pair state—acting on the vacuum (state of no electrons):

$$\Psi_0 = \Pi(u_k + v_k b_k{}^{+})\,|\,0>$$

Where $u_k = (1-v_k{}^2)^{1/2}$. Because the pair creation operators $b_k{}^{+}$ commute for different k’s, Ψ_0 represents the uncorrelated occupancy of the various pair states.

The problem with Ψ_0 was that it was an admixture of states with different numbers of electrons, a problem Schrieffer got round by using a Lagrange multiplier to ensure that the mean number of electrons represented by the wavefunction was always the desired number N. The interaction leading to the transition of a pair of electrons from the state $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$ to $\mathbf{k}'\uparrow, -\mathbf{k}'\downarrow)$ is characterized by a matrix element

$$-V_{kk'} = 2(-\mathbf{k}'\downarrow, \mathbf{k}'\uparrow\ |H_{red}|\ -\mathbf{k}\downarrow, \mathbf{k}\uparrow)$$

where H_{red} is the reduced Hamiltonian from which all terms common to the normal and superconducting states have been removed. $V_{kk'}$ is the difference between one term describing the interaction between two electrons by means of a phonon, and a second one giving their screened Coulomb interaction. The basic similarity of the superconducting characteristics of widely different metals implies that the responsible interaction cannot crucially depend on details specific to individual substances. BCS therefore make the further assumption that $V_{kk'}$ is isotropic and constant for all electrons in a narrow shell straddling the Fermi surface and that $V_{kk'}$ vanishes elsewhere. The fundamental BCS criterion for superconductivity is equivalent to the condition $V < 0$.[22]

As I mentioned above one of the novel features of the BCS ground state was that it did not have a definite number of electrons; a rather odd situation since there clearly could be no creation processes going on in a superconductor. What made this novel was not the notion of an indefinite number of particles itself but that this constraint was used to describe Fermi as opposed to Bose particles; that the creation and annihilation operators referred to electrons. The *form* of the wavefunction was not novel; others including Pines and Frolich had used it in conjunction with Bose particles since these (e.g. phonons, photons and mesons) clearly could be created. Similarly, in other high-energy contexts involving scattering phenomena it was common to write down wavefunctions that had an indefinite number of particles. However, this was not the case for the low energy phenomena where it was assumed that the number of particles should be definite. How did BCS justify this rather bold step? Essentially they appealed to a fundamental idea from statistical mechanics. Given that there were so many pairs spread over such a large volume it made sense to think of them as not being completely correlated with one another but correlated only in a statistical sense.[23] In other words the wavefunction represented a kind of statistical ensemble where the pairs were partly independent, constituting a superposition of states with different numbers of particles. A Hartree type approximation (which does not conserve the number of particles) was used where the probability distribution of a particular state does not depend (at the level of description that is given) on the distribution of the others, something that had never been applied to electrons. This was justified by arguing that the occupancy of some one state was basically independent of whether other states were occupied.[24]

In summary then the basic picture that comprises the representative model can be given in the following sentence: The foundation of superconductivity is the attractive interaction (Cooper pairing) between electrons that results from their coupling to phonons. Once this physical picture was in place, so to speak, BCS then focussed providing a full quantum mechanical treatment that involved solving the "pairing Hamiltonian". While this involved making use of the interpretive models that Cartwright discusses, much more was required to flesh out the complete picture, specifically details regarding the electrons described by the representative model. In the presence of this interaction the system forms a coherent superconducting ground state, characterized by occupation of the individual particle states in pairs such that if one member of the pair is occupied the other is also. BCS then went on to calculate the energy difference between the normal and the superconducting phase at zero temperature and found it to be proportional to the square of the number of electrons (n_c) virtually excited in coherent pairs above the Fermi surface. They also showed that the electron-hole spectrum contains a gap proportional to n_c.

PHILOSOPHICAL CONCLUSIONS

What can we conclude about the role of the representative model in the development of the quantitative theory of BCS? Are there any assumptions comprising specific features of the model that might be considered ad hoc (in the sense that Cartwright suggests) and what clues, if any, can we elicit from the foregoing discussion about the nature of representation? First, we can see that many of the fundamental ideas present in the Londons' work find a place, albeit in a different form, in the BCS theory. In the latter context the 'coherence' and constancy of the momentum vector came about as a result of the energy gained through the interaction of the electrons; it was, in other words, an energetic effect. Fritz London's ideas about waves coupling and a macroscopic quantum wave were absent. BCS incorporated London's suggestions into the context of modern field theory by transforming the understanding of these ideas and leaving behind much of what he relied on in formulating the notion of coherence. But perhaps the most important issue is the focus on the reduced or pairing Hamiltonian which is a direct consequence of how the causal role of Cooper pairing was understood. Rather than being *ad hoc*, this emphasis on the reduced Hamiltonian emerges naturally from the representative model used to describe what BCS took to be the essential features of the superconducting state.

However, once it came to finding a solution for the pairing Hamiltonian a seemingly ad hoc strategy emerged. The fact that the occupation of one state was independent of the occupation of another was the essence of the Hartree approximation. In order to avoid the specific *physical* assumption of an indefinite number of particles BCS took as the groundstate wavefunction the projection of Ψ onto the space of exactly N pairs; in other words, a projection onto a space that had a definite number of pairs. Put differently, it is possible to distinguish particular *ad hoc* assumptions from those that resulted from a systematic development of earlier work, or those that are simply idealizing, such as the neglect of anisotropic effects which result in superconducting properties being dependent only on gross features rather than details of the band structure. The attempt to formulate a microscopic theory required more than just the interpretive models—more than simply describing electrons by assigning them to Bloch states and allowing for their interaction through the Coloumb potential and other interactions. While these ideas were an important part of the BCS picture the development of the theory required a representative model explaining how superconductivity was produced; a model that could be set in a quantum mechanical framework.

Further details about how the pairing takes place and how one should interpret the wavefunction extended the picture provided by the representative model. While some of these details, like the indefinitness assumption, seemed *ad hoc* at the time, the latter soon emerged as an essential feature of superconductivity relating to the phase of the wavefunction and the Josephson effect. Although there have been many different accounts of electron pairing put forward, accounts that differed from the one provided by BCS, the basic fact of pairing which constituted the representative model was and remains the fundamental mechanism at the foundation of superconductivity. Indeed, measurements on superconductors are now used to derive detailed quantitative information about electron-phonon interaction and its energy dependence.

What then are the points of departure between my story and the one told by Cartwright? Cartwright herself notes that the important derivations in the BCS paper are based on the reduced Hamiltonian that I discussed above. She points out that the three terms (two for energies of electrons moving in a distorted periodic potential and one for a simple scattering interaction) use only the basic models provided by quantum theory and goes on to claim that "superconductivity is a quantum phenomenon precisely because superconducting materials can be represented by the special models that quantum theory supplies" (265). My claim is that while these interpretive models do represent, in the general sense of showing how superconducting phenomena exhibit quantum features, that sense of representation is *secondary*—we don't get an understanding of *how* superconductivity takes place from these models. Moreover, BCS simply could not have developed their 1957 account based on these models alone. Not only does one need the representative model that refers to Cooper pairing but we need detailed information about how the electron pairs behave in the context of a quantum representation. None of the crucial information specific to superconductivity comes via the interpretive models. By contrast, the representative model furnishes the fundamental causal mechanism responsible for superconductivity and provides the justification for focusing on the reduced Hamiltonian as well as informing particular constraints imposed on the BCS wavefunction.

Cartwright of course discusses some of these ideas regarding representation in what she calls the 'full underlying model' (275). She describes it in the following way: "There is a sea of loose electrons of well-defined momenta moving through a periodic lattice of positive ions whose natural behaviour, discounting interactions with the electron sea, is represented as a lattice of coupled harmonic oscillators, subject to Born-Karman boundary conditions" (275). But her point is that this is *not* a literal presentation but rather a representation of the structure of some given sample of superconducting material. It is itself a model and not the "real thing" truncated (276). A primary model of this kind aims to be explanatory by not only representing the essential features of superconductivity but also by bringing these elements under the umbrella of the theory (ibid.). Cartwright's description of the underlying model involves a number of assumptions: 1) the particles in the model called 'electrons' are fermions and hence obey the exclusion principle; 2) those referred to as positive ions are bosons; and 3) that the particles in the model affect each other through Coulomb interactions, occurring pairwise between the electrons, positive ions and the electrons and ions. She goes on to claim that the only kind of Hamiltonian used to describe this underlying model is the one for the Coulomb potential. Moerover, it is impossible to solve the equation for the ground state of a Hamiltonian like this. Instead BCS substitute the new interpretive models (like the harmonic model) with their corresponding Hamiltonians together with other constraints that one hopes will result in the full underlying model agreeing with the results from the BCS model (276).

I am unclear as to what exactly this full underlying model is, but, it certainly isn't what I have called the "representative model". Moreover, the steps in Cartwright's story seem to present an account that differs significantly from the development of the 1957 BCS paper. As I understand it once the representative model was in place BCS could then focus on the reduced or pairing Hamiltonian which connects the pairs with zero net momentum and supplies the interaction terms.[25] While the representative model embodies some of the ideas outlined in Cartwright's underlying model, the construction of the wave equation involves the extension of the model to include conditions specific to the pairing mechanism. As she presents it, Cartwright's "full underlying model" doesn't really tell us much about how we get an account of superconductivity within a quantum framework. Her reconstruction does tell us about the importance of "Bloch states" and the "Coulomb potential", but we need much more than this to properly understand how the qualitative and quantitative aspects of superconductivity are brought together in the BCS account. So, my disagreement with Cartwright about the details of the BCS paper emerges as a methodological disagreement about the role and importance of representative models. By focusing on the physical ideas that form the basis of the representative model we acquire not only a richer account of the role of representation than that which arises from the interpretive models, but also a more comprehensive story about how the qualitative and quantitative ideas interact to produce the 1957 BCS account.

The notion of representation present in the BCS model is straightforward in the sense that it provides an account of how a superconducting system might be constituted. The representative model serves as the source of mediated knowledge in the sense that one does not have direct access to the pairing process itself nor to the other aspects of the system that are neglected in dealing only with a reduced Hamiltonian. For example, there are many terms in the complete interaction that connect pairs with total momentum $\mathbf{q} \neq 0$. These have little effect on the energy and can be treated as a perturbation. The kind of mediated knowledge furnished by representative models is characteristic of the practice of scientific modelling in general, with the differences across contexts accounted for by the specific nature of the representation required in each case. In other words, representative models can 'represent' in a variety of ways and the adequacy of the representation will depend, to a great extent, on what we want the model to do for us.

In Morrison (1998) and (1999) I discussed various kinds of representation ranging from the attempts to accurately depict the pendulum in its use as a measuring instrument to the structural dependencies that exist in the model used in the construction of boundary layer theory. In each case the legitimacy of the representation is a function of the model's intended domain and use. That is to say, in some contexts we only need a partial description of the system in question in order to use the model to make predictions and/or explain some specific phenomenon. In other contexts, such as superconductivity, we want to know how (possibly) the superconducting state arises and for that we need a reasonably well worked out model of the basic processes that go on in superconducting metals generally. That function is fulfilled by the representative model described above; a model that provides an explanation of the fundamental features that give rise to superconductivity. The model, by nature, leaves out certain elements deemed to be inessential parts of the real system and in doing so offers us a 'mediated' account of how the system is constituted. Again, I use the term 'mediated' here to indicate that the model functions as a kind of 'stand-in' or replacement for the system under investigation and that it furnishes only a partial representation; it is, in essence, one step removed from the real system.

Because scientific modelling typically embodies this kind of partial representation the natural question to ask concerns when there is sufficient detail for the model to count as a credible source of knowledge. This, however, cannot be answered in advance, nor is there an algorithm for determining the appropriate methods for model construction and legitimation. This fact is perhaps best illustrated in Bohr's response to BCS where he claims that while they had the essential answer, the understanding of what the pairs *really were* and why the other terms were unimportant was completely obscure. It was an interesting idea but nature wasn't that simple. Clearly for Bohr more explanatory principles or facts were necessary to justify the BCS account. For others the model was sufficient as an explanation of superconductivity but the lack of gauge invariance spoke against its status as a general theory. But, these kinds of debates are part of the practice of model construction and acceptance, as well as general feature of scientific investigation. By focusing on the role representative models play in understanding scientific phenomena we can, among other things, learn more about the relation between theories and models in the production of knowledge.[26]

NOTES

1. That said, I do intend my discussion to be read as a general contribution to philosophical issues related to modelling. I mention the LSE project primarily because I do not want to reiterate in detail the position in Morrison and Morgan (1999) and Morrison (1999), a position that grew, in part, out of work done in conjunction with that project.
2. Cartwright uses the terms ‘interpretive’ and ‘representative’ models so in the spirit of consistency I will do so as well. I see no difference between her use of representative models and what some people call representational models. Both allegedly represent, to a greater or lesser extent, some aspect of the world. I would extend that use and say that models can also represent theory insofar as they provide a concrete instantiation of some formal aspects of a theory, as in the case of the pendulum model representing laws of motion. This is the role Cartwright assigns to interpretive models and I assume this is what she means when she says that interpretive models can also represent. I understand function here to mean, literally, what the model ‘does’ rather than the role it plays in developing the theory. Cartwright’s account seems to imply that interpretive models are the crucial ones for the development of the BCS account. I want to claim that while their role is important in filling out the theoretical story it is secondary and depends, ultimately, on having a representative model in place before the interpretive models can be put to work in the appropriate sort of way. While I am not prepared to say that this is universally the case it certainly is so in the case of BCS.
3. I was reminded of this phrase by a recent book by Robert Batterman (2002).
4. Just as a point of clarification, I am not making any metaphysical claims here about the nature of causation or that the story must have some kind of truth/necessity attached to it. Instead the notion of ‘cause’ is meant only in the intuitive, pre-theoretical sense.
5. This depends on what one takes to be a theory or a part of theory and what one takes to be a model. This is a complicated issue and one that requires more than a few lines to clarify the differences. In Morrison and Morgan (1999) we focussed on various aspects of models and modelling practices in an attempt to lay some groundwork for what models do and how they function. We did this specifically because it wasn’t clear to us that one could give a straightforward definition of what constitutes a model (leaving aside the way model is defined in model theory or by the semantic view, a view we wanted to distance ourselves from). That said, I, at least, think it is possible to differentiate a model from a theory in specific contexts but I am still not convinced that there are general criteria for doing so that are applicable across the board. To that extent I want to here resist the temptation, so irresistible to many philosophers, to ‘define’ what a model is.
6. Of course this new description may be still abstract but the point is that it more closely approximates the real system than the abstract concepts do.
7. I say ‘models’ here instead of ‘model’ because I want to show how the development of the BCS account of superconductivity relied on a variety of representational models, culminating in the model that represents the electrons as Cooper pairs. It is the historical evolution of that representation that makes use of many different representational models along the way.
8. Some ideas associated with pairing come from QM while others are motivated in a more indirect way.
9. The 1957 paper is sometimes referred to in connection with the BCS model, with certain assumptions retained in the later and more comprehensive BCS

theory. Distinguishing between the two is not important for my purposes here so I will simply refer to the BCS theory/model. Ultimately I would want to distinguish between the BCS theory as a *generic* theory in which Cooper pairs are formed and the BCS model to include *specific* assumptions made in the 1957 paper about the form of the attractive potential etc. See Morrison [2007].

10. Essentially the lattice is distorted by a moving electron and this distortion gives rise to a phonon. A second distant electron is in turn affected when it is reached by the propagating fluctuation in the lattice charge distribution. The nature of the resulting electron-electron interaction depends on the relative magnitudes of the electronic energy change and the phonon energy. When the latter exceeds the former the interaction is attractive.
11. But, at temperatures near absolute zero the interaction between the electrons and the lattice is very weak and hence not sufficient inter-electron attraction to overcome the Coulomb repulsion. Consequently there is no transition into the superconducting state.
12. This is what happens in the case of semiconductors, i.e. solids which also have an energy band gap but yet don't show superconducting properties. The key difference between these two types of metals is of course the presence of Cooper pairs.
13. I should mention here that the discovery of the isotope effect was crucial to both the development (in Bardeen's case) and the confirmation (in Frohlich's case) of the account of electron-phonon interactions. The discovery in 1950 confirmed that the critical temperature varies inversely with the square root of the isotopic mass. Bardeen took this to suggest that the energy gap might arise from dynamic interactions with the lattice vibrations or phonons rather than from static lattice distortions. Both had concluded that the amount of self-energy was proportional to the square of the average phonon energy. In turn this was inversely proportional to the lattice mass, so that a condensation energy equal to this self-energy would have the correct mass dependence indicated by the isotope effect. (The isotope effect showed that the critical temperature is related to the mass of the atoms of the solid.) Unfortunately the size turns out to be three to four orders of magnitude too large.
14. For a technical discussion of the construction of the wave equation see Schrieffer [1973] and of course, Cooper [1956] and Bardeen, Cooper and Schrieffer [1957].
15. The concept of a Fermi sphere or sea refers to the idea that all the particles (nucleons or electrons) ar distributed mostly into the lowest energy states which (in the simplest case) form a spherical shape in k-space.
16. However, the conceptual centrepiece of the model for a superconductor, notion of Cooper pairs, was not without its problems. As I noted above, Cooper's model for pairing was essentially a two-body problem; a more realistic account of the superconducting state required that one be able to write down a many-body wavefunction taking the pairing into account. But, in order to do that some fundamental problems needed to be resolved. One such problem was that if all the electrons near the Fermi surface were paired in the way that Cooper described, the pairs would strongly overlap. This results from the determination of the binding energy, which in turn determines the size of the pair wavefunction at 10^{-4}cm. However, if all the electrons take part in pairing the average spacing between pairs would be only about 10^{-6}cm, a distance much smaller than the size of the pair. The other difficulty concerned the subtlety of the energy change in the transition from the normal to the superconducting state which was of the order of 10^{-8}eV per electron, far smaller than the accuracy with which one could hope to calculate the energy of either the

normal or the superconducting state. The root of these problems was the fact that the situation described by the model was highly artificial. More specifically, the two electrons whose interaction was considered were treated differently from the others whose role was only to occupy the Fermi sea and prevent certain states within from being occupied by the principle actors. A satisfactory account would require that all electrons be treated equally, especially with respect to the Fermi statistics. In other words, the many-body wavefunction must be antisymmetric under interchange of the coordinates and spin indicies of any two electrons.

17. For a discussion of the influence of these ideas see Bardeen's own articles (1973a & b).
18. This is because resonance between each two pairs would lower the energy, so some of the possible energy lowering would be lost if all the pairs did not have the same momentum.
19. Moreover, it produced an intuitive explanation of the energy gap—each pair was interacting with many others, hence breaking one of the pairs meant losing all the negative energy that had been derived through those many interactions.
20. This is actually a truncated version of the three term Hamiltonian that appears in their 1957 paper. The interaction terms are defined with a negative sign so that $V_{kk'}$ will be predominately positive for a superconductor.
21. This point is essential to the theory and leads to the energy gap being present not only for dissociating a pair but for making a pair move with a total momentum different from the common momentum of the rest of the pairs.
22. The simplification of the interaction parameter *V* leads to what can be called a law of corresponding states for all superconductors, that is, virtually identical predictions for the magnitudes of all characteristic quantities in terms of reduced co-ordinates. Any empirical deviation from this complete similarity would not constitute an invalidation of the basic premise of BCS but would simply indicate the idealisation built into their account of *V*.
23. Bardeen later pointed out however, that the form of the wavefunction, with the all-important common momentum for paired states, is determined by energetic rather than purely statistical considerations. See his 1957 reply to Dyson quoted in his paper in Kursunoglu and Perlmutter [1973b].
24. This aspect of the model definitely had the appearance of an *ad hoc* assumption and despite its apparent efficacy, BCS were not happy with the idea of an indefinite number of states. They attempted to downplay the idea by claiming that the spread of particle number in their state would be small, or by claiming that the superconducting wavefunction could be taken to be the projection of the state that they had introduced onto the space that did have a definite number of particles. Later developments in the theory would reveal that this indefiniteness was an essential feature of the superconducting state since it allowed for the introduction of the quantum phase defined over the whole of the macroscopic superconductor. However, it was still not possible to think of electrons being created or destroyed since we were dealing with low temperature solid-state phenomena. To that extent BCS emphasized that their system really did have a definite number of electrons despite the form of the wavefunction.
25. While Cartwright mentions the 'reduced Hamiltonian' in her paper (256) much of her discussion centres on the longer Hamiltonian that contains four terms.
26. I would like to thank Ron Giere, Paul Humphreys and Paul Teller for helpful suggestions and comments on an earlier draft. Support of research by the SSHRC is gratefully acknowledged.

REFERENCES

Bardeen, J. (1951) Reviews of Modern Physics, 23, 261.

——— (1973a) "Electron-phonon interactions and superconductivity" *Physics Today*, 41–46.

——— (1973b) "History of Superconductivity" in *Impact of Basic Research on Technology*, Kursunoglu, B. and Perlmutter, A. (eds.) New York: Plenum Press.

Bardeen, J., Cooper, L. and Schrieffer, J.R. (1957) "Theory of Superconductivity" *Physical Review*, 108, 1175–1204.

Batterman, Robert (2002) *The Devil in the Details*. New York: Oxford University Press.

Cartwright, N. (1999a) "Models and the limits of theory: Quantum Hamiltonians and the BCS models of superconductivity" in Morgan, M. and Morrison, M. (eds.) *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press, 241–81.

——— (1999b) *The Dappled World*. Cambridge: Cambridge University Press.

Cartwright, N., Shomar, T. and Suarez, M.(1995) "The Toolbox of Science" in *Theories and Models of Scientific Processes* Poznan Studies in the History and Philosophy of the Sciences and the Humanities. Vol. 44 Rodopi, 137–49.

Cooper, L. (1956) "Bound Electron Pairs in a Degenerate Fermi Gas" *Physical Review*, 104, 1189–90.

——— (1973) "Microscopic quantum interference in the theory of superconductivity" *Physics Today*, 31–39.

Frohlich, H. (1950) "Theory of Superconducting State 1. The Ground State at the Absolute Zero of Temperature" *Physical Review*, 79, 845–56.

Giere, R. (1988) *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

Kuhn, T. (1977) "Objectivity, Value Judgements and Theory Choice" in *The Essential Tension* Chicago: University of Chicago Press.

Kuper, C.G. (1959) "The Theory of Superconductivity" *Advances in Physics*, 8, 1–44.

London, F. (1950) *Superfluids* New York: Wiley

London, F. and London H. (1934) "The Electromagnetic Equations of the Supraconductor" *Proceedings of the Royal Society* A149, 71–88.

Leggett, A.J. (1997) "The Paired Electron" in *Electron: A centenary volume* M. Springford (ed.) Cambridge: Cambridge University Press, 148–181.

McMullin, E. (1985) "Galilean Idealization" *Studies in History and Philosophy of Science*,16, 247–73.

Morrison, M. (1998) "Modelling Nature: Between Physics and the Physical World" *Philosophia Naturalis*, 38, 65–85.

——— (1999) "Models as Autonomous Agents" in Morgan and Morrison (eds.) *Models as Mediators: Essays on the Philosophy of the Natural and Social Sciences*, Cambridge: CUP, 38–65.

——— (2007) "Where Have All the Theories Gone?" Philosophy of Science, 74, 195–227.

Morrison, M. and Morgan, M. (1999) "Models as Mediating Instruments" in Morgan and Morrison, op.cit., 10–37.

Rickayzan, G. (1965) *Theory of Superconductivity*, New York: Wiley.

Schrieffer, J.R. (1973) "Macroscopic quantum phenomena from pairing in superconductors" *Physics Today*, 23–28.

Reply to Margaret Morrison

The following reply to Margaret Morrison's paper was written by LSE Ph.D. student Gabrielle Contessa; I think I can do no better than to offer you his remarks.

Margaret Morrison and Nancy Cartwright share many views about scientific models. Far from revealing a disagreement between them about the role and importance of interpretive and representative models, their analyses of the superconductivity case seem to be not only compatible but largely complementary. According to Morrison, Cartwright's discussion of the BSC model of superconductivity case in 'Models and the Limits of Theory' does not do justice to the role that representative models played in the development of a model of superconductivity. Morrison highlights the continuities among successive representative models of superconductors and, in particular, the recurring assumption that in superconductors there is an energy gap between the ground state and the excited state. According to Morrison, this continuity shows that the development of a theory of superconductivity was driven by the representative models culminating in the BCS model, which by employing the concept of Cooper pairs, provides a concrete underpinning for the mathematical BCS theory.

If Cartwright does not focus on representative models in the paper in question, however, it is not because she thinks that the development of the BCS model was driven by interpretive models or because representative models do not play any role. Rather, it is because in that paper she is concerned with the limits of theory; the case of the BCS model serves to illustrate her thesis that theory, in this case Quantum Mechanics, does not stretch beyond those successful applications that use its interpretive models in a principled way. The questions of whether interpretive or representative models are prior or of whether the construction of the model was driven by interpretive or representative models seem to rest on the assumption that those two kinds of models serve the same functions. But they do not, and this is why, to apply the theory to any real world situations, we need both kinds of models. We need interpretive models to exemplify how the abstract concepts of the theory apply to more concrete situations. We need representative models to represent the systems we come across in the world.

In *How The Laws of Physics Lie*, Cartwright imagined that, when modelling a real world system, we start by writing down an unprepared description of the system. However, in order to apply the concepts of the theory to the situation we need to "prepare" the description—we need to redescribe it in a way that allows us to apply the equations of the theory in a principled way via the bridge principles of the theory. So, the preparation of the description is done with an eye on the interpretive models of our theories. But this is often not enough. To turn the models obtained by combining the basic blocks of the interpretive models into representative models, we often need to make ad hoc corrections—corrections that ultimately can only be justified by the empirical success of the model. Cartwright's contention in 'Models and the Limits of Theory', as well as in many other places, is that when these ad hoc corrections are needed, the success of the model does not count as a successful application of the theory, as these corrections not only are often not suggested by the theory, but are sometimes at odds with it.

When we look at the broader picture, the questions of whether interpretive or representative models are prior or of whether the construction of the BCS model was driven by interpretive or representative models seem to dissolve. And so does the supposed disagreement between Morrison and Cartwright.

5 The Finewright Theory[1]

Paul Teller

INTRODUCTION

In *How the Laws of Physics Lie* (Cartwright 1983) Nancy Cartwright launched contemporary attention to the fact that science describes the world with the use of models that are always limited in scope and never completely accurate. One year later Arthur Fine introduced his Natural Ontological Attitude, or NOA, (here cited in Fine 1986a, b) which steers a middle ground between what he argues are problematic extremes of scientific realism and various contrary forms of antirealism. I will explore a certain confluence between Fine's and Cartwright's views. Seeing Fine's view through Cartwright's lens will bring to light an interesting way to see certain aspects of his view. On the other hand, approaching Cartwright through the issues that concern Fine will call our attention to the idea that Cartwright's observations apply much more broadly than just to science, indeed that they raise issues about how to think about truth.

What follows will vary greatly in what it accomplishes. On the one hand I hope to present some well-developed clarifications and alternative ways of thinking about some of the views of both Fine and Cartwright. On the other hand, I can be no more than tentative about the most important facet of this investigation: the potential repercussions for our thinking about truth. The latter is a vast project, and I can here do no more than show how the material motivates alternatives, illustrate the tentative ideas with examples, and explore in a preliminary way the kinds of alternatives that one might try to develop in more detail. But even so much should demonstrate the richness and importance of the views under consideration.

FINE'S NATURAL ONTOLOGICAL ATTITUDE (NOA)

An effective way to understand NOA is by contrast with the views that Fine finds problematic. Fine certainly does not want to deny that, for example, there are electrons. If the evidence has not misled us, electrons exist in the same straightforward sense that chairs and tables exist (Fine 1986a:

126–127, 130; 1986b: 176–177). Rather, the problem with scientific realism, as Fine interprets this view, is that it adds a problematic correspondence theory of truth (Fine 1986a: 116, 133, 139; 1986b: 150). On the other hand, where realism asks more of truth by way of some kind of "outer" connection, the various forms of antirealism look for a human based recharacterization of truth (Fine 1986a: 129, 133, 139). After critically examining some special cases of the views he finds problematic, Fine asks more generally: why do we need any interpretation of truth claims? Rejecting the need for any interpretation, such as might be proposed by various forms of realism and antirealism, Fine recommends what he calls "the homely line" which underlies what he calls the "core position" that must be common to all parties:

> I certainly trust the evidence of my senses, on the whole, with regard to the existence and features of everyday objects. And I have similar confidence in the system of "check, double–check, check, triple check" of scientific investigation, as well as the other safeguards built into the institutions of science. So, if the scientists tell me that there really are molecules and atoms, and psi/J particles, and, who knows, maybe even quarks, then so be it. I trust them and thus, must accept that there really are such things with their attendant properties and relations. (Fine 1986a: 126–127, 147–148; 1986b, 176–177)
>
> Then it seems to me that both the realist and the antirealist must toe what I have been calling the "homely line". That is, they must both accept the certified results of science as on par with more homely and familiarly supported claims . . . let us say then, that both realist and antirealist accept the results of scientific investigations as "true", on par with more homely truths . . . and call this acceptance of scientific truths the "core position". (Fine 1986a: 128, 149–150; 1986b: 172–173, 176–177)

As I initially suggested, the way to grasp the content of the core position is to contrast it with what the alternatives want to add:

> The antirealist may add onto the core position a particular analysis of the concept of truth, as in the pragmatic and instrumentalist and conventionalist conceptions of truth. Or the antirealist may add on a special analysis of concepts, as in idealism, constructivism, phenomenalism, and in some varieties of empiricism. . . . or the antirealist may add on certain methodological strictures . . .
>
> (Fine 1986a: 128–129, 137; 1986b: 150, 157, 171–172)
>
> . . . the realist wants to explain the robust sense in which *he* takes these claims to truth or existence; namely as claims about reality—what is

> really, really the case. The full-blown version of this involves the conception of truth as correspondence with the world, and the surrogate use of approximate truth as near correspondence. (Fine 1986a, 129, 136–137; 1986b: 171–172, 176)

Most important for our comparison with Cartwright is NOA's concomitant attitude towards truth and theories of truth:

> . . . realism differs from various antirealisms in this way: realism adds an outer direction to NOA, that is, the external world and the correspondence relation of approximate truth; antirealisms (typically) add an inner direction, that is, human-oriented reductions of truth or concepts, or explanations . . . NOA suggests that the legitimate features of these additions are already contained in the presumed equal status of everyday truths with scientific ones, and in our accepting them both as *truths* . . . [Thus] a distinctive feature of NOA, one that separates it from similar views currently in the air, is NOA's stubborn refusal to amplify the concept of truth by providing a theory or analysis (or even a metaphorical picture). Rather, NOA recognizes in "truth" a concept already in use and agrees to abide by the standard rules of usage. (Fine 1986a: 133, 149–150; 1986b: 170)

LIES AND THE HOMELY LINE

But what are we to make of Fine's homely line if science gives us only lies?[2,3] Let us review Cartwright's conclusions (1983). Physics is analogized to theater (Cartwright 1983: 139–142) in adhering to the facts only as closely as is practicable and then only to those aspects of the situation that are currently of interest. In staging a representation of, say, a historical episode, in many respects one would fail to communicate to the audience clearly what happened if one insisted on too much realism:

> We need only adhere [as Thucydides describes writing a good history] 'as closely as possible to the general sense of what was actually said.' Physics is like that. It is important that the models we construct allow us to draw the right conclusions about the behavior of the phenomena and their causes. But it is not essential that the models accurately describe everything that actually happens; and in general it will not be possible for them to do so . . . (Cartwright 1983: 140)

Cartwright sees theories and laws as tools in the model building 'Toolbox of science' (Cartwright et al. 1995). Laws are true only of things in the models that the laws are used to design: 'My basic view is that fundamental equations do not govern objects in reality; they govern only objects in models' (Cartwright 1983: 129). In turn, the models built are themselves caricatures, works of fiction (Cartwright 1983: 150):

> A model is a work of fiction. Some properties ascribed to objects in the model will be genuine properties of the objects modeled, but others will be merely properties of convenience . . . Not all properties of convenience will be real ones. There are the obvious idealizations of physics . . . [and] some properties are not even approached in reality. They are pure fictions. (Cartwright 1983: 153)

Models are simulacra:

> I propose . . . a "simulacrum" account . . . The fundamental laws of the theory are true of the objects in the model, and they are used to derive a specific account of how these objects behave. But the objects of the model have [following the second entry for "simulacrum" in the Oxford English Dictionary] only "the form or appearance of things". (Cartwright 1983: 17, 143)

Even a superficial acquaintance with any science shows that science in fact delivers only the sorts of things that Cartwright reports. Some maintain that nonetheless science "aims" to deliver real truths. I will examine this issue in detail below. For the moment it suffices to note that insofar as the issue is understanding the scientific knowledge we currently have, and anything remotely like it, what different sort of thing science "aims" to deliver is quite beside the point.

If we accept, as I certainly do, Cartwright's observations about what science delivers, how are we to understand Fine's "homely line" that we 'accept the results of scientific investigations as "true" on par with more homely truths'? How are we, at once, to accept such deliverances as truths but also as caricatures, fictions, simulacra, and lies? Yet such IS the way science tells us about the world. To maximize the pain of this tension let me quickly review a salient example, the hydrodynamic model of water. To understand the fluid properties and behavior of water we describe it as a continuous medium. ONLY in this way can we get a characterization that provides a humanly accessible understanding of water's fluid properties. Even if we grant fundamentalists' claims of in-principle reducibility to, say a quantum description (which in any case is still only a model), we would need to set up and solve a Schrödinger equation with 10^{25}–10^{27} variables and then make sense of the results.[4] And it is not as if we could "factor out" true from false "conjuncts". Nor does the hydrodynamic model serve merely as a "useful fiction" to give no more than "observational predictions". The model provides real understanding of how water behaves in respects of intellectual as well as practical interest.

RECONCILIATION WITH NOA

The problem is that NOA admonishes us to accept as true, at least provisionally, the well-confirmed conclusions of science.[5] But science rarely, if ever, gives us better than false idealizations. This conflict between the views can be reconciled by recasting one aspect of the statement of NOA, and this in turn by looking in more detail at the pitfalls of scientific realism, as Fine understands it. The examination will, in turn, pave the way to an alternative way of seeing "false" idealizations in science and elsewhere.

The scientific realism to which Fine objects adds to his core position a correspondence understanding of truth. On this correspondence account, that which is supposed to correspond to a statement operates as a standard of correctness, the "way things really (REALLY!) are". But it is idle to suppose that our deployment of our representations could actively make use of any such correspondence. It is just a fantasy that we could hold our representations up to "the way things really are" to see how our representations fall short (Fine 1986a: 131; 1986b: 151). We can "hold up" our representations only to other representations. "Access" means access via representations, so "direct access" is a contradiction in terms.

After this critique, the best to which a realist-minded philosopher could, in good logical conscience, aspire is a standard of exact representations, that is, representations that are without any inaccuracies. Let us call these "surrogate realist" representations. That science delivers, or at least aspires to, such surrogate realist representations is exactly what Cartwright calls "fundamentalism", which we will examine in more detail below. For the moment it is enough to note: Cartwright has pointed out that we don't have these surrogate realist, fundamentalist representations. It is implausible that we shall ever get them. Most importantly, we don't need them. We get on just fine without them.

I suggest that it is in lock step with the spirit of NOA that we get on just fine without any surrogate realist representations. What, on Fine's reading, realism wants to add to the core position is exact correspondence. But exact correspondence is, logically, not available to us as something that we can incorporate as part of our representational practices. Thus (and this now goes beyond anything Fine explicitly says) what realism in good logical conscience could try to add to the core position is exact, surrogate realist representations. So, in rejecting the need for any addition to the core position, in effect, NOA rejects the exact, surrogate realist descriptions and any need for them. In other words, by reinterpreting the rejection of realism as the rejection of exact descriptions and any need for them, we have the statement of a position that coincides with Cartwright's conclusions about what science actually provides and splendidly serves our practical and intellectual needs.

However, this reconciliation brings with it a possibility of a different kind of tension. Above we saw that NOA incorporates a 'stubborn refusal to amplify the concept of truth by providing a theory or analysis (or even a metaphysical picture)'. Fine's NOA is committed to a "no-theory" view of truth (Fine 1986a: 133, 149–150; 1986b: 175–177). Fine advocates that we accept our use of "true" as a primitive with no interpretation; and in so doing he makes remarks that might appear to conflict with Cartwright's observation that at best science gives us idealizations, not completely accurate descriptions. For Fine appears to take science, at least sometimes, to give us genuine truths. Furthermore, in discussing refinement of scientific conclusions Fine talks about 'conserv[ing] the true parts' (Fine 1986a: 132). He talks about respecting the 'customary logic and grammar' of 'truth' (Fine 1986a: 133, 149; 1986b: 175). And at one point at least he appears to distance himself from the notion of approximate truth (Fine 1986a: 133).

We must remember that, in the first instance, talk of truth is a way of characterizing our representations. It is part of the traditional 'customary logic and grammar' of 'truth' to take sentential or propositional representations to be evaluated in terms of satisfaction or failure of what are thought of as exact truth conditions. Already analogue representations are not so evaluated.[6] The foregoing drives us towards drawing a similar conclusion about evaluation of the representations traditionally evaluated as "true" or "false".

In short, we need alternative ways of thinking about truth: accept the results of scientific investigation and theorizing as "true" on par with more homely truths, but be cognizant about what it is to do so. It is not just that when we evaluate a representation as true we do so provisionally, subject to revision. That is, our attributions of truth are not, or not just, provisional in the sense that we expect to discover the deliverances of science to have some true "conjuncts" and some false ones. Rather we are well advised to look for ways of thinking about how we represent and know about the world, alternative to something characterizable in terms of exact truth conditions.

Such an admonition may appear to be a call for a global interpretation of truth of the kind Fine rejects. Below I will offer some preliminary thoughts about relevant alternative ways of thinking about truth and address the worry of whether so doing flouts Fine's proscription. But first a clear understanding of the foregoing reinterpretation of Fine and comparison with Cartwright requires sorting through what Cartwright has called "fundamentalism".

FUNDAMENTALISM

In the foregoing I have reinterpreted Fine as rejecting what Cartwright calls "fundamentalism".[7] Cartwright herself has in various places appeared to take somewhat different stands on this issue. It will be useful to sort through the positions. This in turn will be aided by further refining the characterization of what gets called "fundamentalism", breaking it down into what I will call "epistemic" and "ontic" versions:

> *Epistemic-fundamentalism:* We have, or can reasonably expect to have, or even just hope to find—the exact, universally applicable generalizations that have been the traditional objective of science.

I entirely agree with Cartwright's rejection of epistemic-fundamentalism, a rejection that is always implicit and often explicit in her overall critique.

But should we nonetheless think of the world we live in as being governed by such laws, even if we don't have them and expect never to have them? To endorse this view would be to endorse

> *Ontic-fundamentalism:* The world is governed, in every respect, by exact laws that apply universally even though we have not yet found these laws and the world may, in fact, be too complicated for us ever to find them.

Now, by "fundamentalism" does Cartwright have in mind the epistemic or the ontic version? For example, she writes:

> [One can make a] rough division of the concrete facts we know into two categories: (1) those that are legitimately regimented into theoretical schemes, these generally, though not always, being about behavior in highly structured, manufactured environments like a spark chamber; (2) those that are not. There is a tendency to think that all facts must belong to one grand scheme, and moreover that this is a scheme in which the facts in the first category have a special and privileged status. They are exemplary of the way nature is supposed to work. The others must be made to conform to them. This is the kind of fundamentalist doctrine that I think we must resist. (Cartwright 1999: 24–25)

I think that this passage, as well as others in Cartwright (1999) can be read either way.

As a convenient terminological handle for what is here at issue, let us use *theism* as shorthand for ontic-fundamentalism—the belief that the world is as it would be described by a successful epistemic-fundamentalism. And let us use *atheism* as shorthand for the positive rejection of theism.

Cartwright is clearly no theist. But is she an atheist? Although she does not appear anywhere explicitly to link the term "dappled world" to the term "fundamentalism" the positive thesis of a dappled world seems clearly to be an expression of atheism:

> This book supposes that, as appearances suggest, we live in a dappled world, a world rich in different things, with different natures, behaving in different ways. That laws that describe this world are a patchwork, not a pyramid. They do not take after the simple, elegant and abstract structure of a system of axioms and theorems. Rather they

> look like—and steadfastly stick to looking like—science as we know it: apportioned into disciplines, apparently arbitrarily grown up; governing different sets of properties at different levels of abstraction; pockets of great precision; large parcels of qualitative maxims resisting precise formulation; erratic overlaps; here and there, once in a while, corners that line up, but mostly ragged edges; and always the cover of law just loosely attached to the jumbled world of material things. For all we know, most of what occurs in nature occurs by hap, subject to no law at all. What happens is more like an outcome of negotiation between domains than the logical consequence of a system of order. The dappled world is what, for the most part, comes naturally; regimented behavior results from good engineering. (Cartwright 1999: 1)

In various places Cartwright offers explicit arguments for atheism, the conclusion that the real world is dappled:

> My belief in the dappled world is based in large part on the failures of [physics and economics] to succeed in [their aspirations to account for almost everything, the first in the natural world, the second in the social]. (Cartwright 1999: 1)

> . . . I conclude that even our best theories are severely limited in their scope. For, to all appearances, not many of the situations that occur naturally in our world fall under the concepts of these theories . . . I want to consider what image of the material world is most consistent with our experiences of it. . . .
>
> (Cartwright 1999: 9)

> The point is that the claims to knowledge we can defend by our impressive scientific successes do not argue for a unified world of universal order, but rather for a dappled world of mottled objects.
>
> (Cartwright 1999: 10)

This kind of argument also appears in (Cartwright 1998: 98) and (Cartwright 2000: 220–221).

I take the form of the argument to be this: our best efforts to describe the world, taken as a whole, describe it as dappled (this is just anti-epistemic-fundamentalism). Sound methodology requires, as the most reasonable conclusion about how the world operates, that it is as so described by our most careful and vigorous descriptive efforts—anything else would be unwarranted metaphysical trappings. So we must conclude the world to be dappled—atheism as opposed to theism—as the best-supported hypothesis. The most specific presentation of this kind of argument comes in Cartwright's discussion of the $1,000 bill blown about by the wind in St. Stephen's Square. (Cartwright 1999: 26)

I have cited a number of passages that appear to present Cartwright, as I think many have read her, as an uncompromising atheist. However, in Cartwright (1999) there is also a significantly different approach to this issue. On page 11 she develops an analogy to the problem of evil: Are our observations of the world's pain and misery compatible with an infinitely powerful and good Deity (1999: 11)? Yes, but the evidence hardly favors belief in such a Deity. Similarly:

> Complication and limitation in the truest laws we have available are compatible with simplicity and universality in the unknown ultimate laws. But what is advanced by this concession? Just as we know a set of standard moves to handle the problem of evil, so too are we well rehearsed in the problem of unruliness in nature—and a good number of the replies have the same form in both discourses: the problem is not in nature but, rather an artifact of our post lapsarian frailties. I follow Philo [From Hume's *Dialogues Concerning Natural Religion*] in my reply: guarantee nothing *a priori*, and gather our beliefs about laws, if we must have them at all, from the appearance of things. (Cartwright 1999: 11–12)

She continues:

> The dappled world that I describe is best supported by the evidence, but it is clearly not compelled by it. (Cartwright 1999: 12)
>
> Why then choose at all? Or, why not choose the risky option, the world of unity, simplicity, and universality? If nothing further were at stake, I should not be particularly concerned about whether we believe in a rule-governed world or in a unruly one, for, not prizing the purity of our affirmations, I am not afraid that we might hold false beliefs. The problem is that our beliefs about the structure of the world go hand-in-hand with the methodologies we adopt to study it. The worry is not so much that we adopt wrong images with which to represent the world, but rather we will choose wrong tools with which to change it. We yearn for a better, cleaner, more orderly world than the one that, to all appearances, we inhabit. But it will not do to base our methods on our wishes. We had better choose the most probable option and wherever possible hedge our bets (Cartwright 1999: 12–13; cf. also Cartwright 1998: 96).

Cartwright then gives three examples of 'how belief in the unity of the world and the completeness of theory can lead to poor methodology . . .' (Cartwright 1999: 13). Economists who believe they have an all-encompassing theory tend to reject, on principle, what otherwise might well be seen as important relevant information (Cartwright 1999: 13–16). In physics advocates of research on "fundamental" theory can draw attention—and funds—from more practical work (Cartwright 1999: 16–17). In medicine research can be hypnotized by the holy grail of genetic reductionism (Cartwright 1999: 17–18). Cartwright explains the damage that can ensue:

> Is the level of effort and funding that goes into the gene programme versus the others warranted by the promise of these programmes for understanding and controlling [e.g.,] breast cancer or does the gene programme get a substantial edge because it is the gene programme; because it is our best shot right now at a theory of everything? I care about our ill supported beliefs that nature is governed by some universal theories because I am afraid that women are dying of breast cancer when they need not do so because other programmes with good empirical support for their proposals are ignored or underfunded. Cartwright 1999: 18)

Here is how I read these passages: The atheism we are considering is a piece of metaphysics. And on the face of it Cartwright hears metaphysics with a pejorative ear: Insofar as it hardly matters what the world beyond our practical access is "really like", a piece of metaphysics isn't worth the worry one way or the other (see also Cartwright 2002a: 271). However, insofar as our metaphysics prejudices our methodology it DOES matter. Consequently Cartwright urges that we take a stand on this issue. Even if the evidence is not conclusive for atheism, at the very least for such important practical affairs we had better hedge our bets (Cartwright 1999: 12–13; 1998: 96).

It is crucial not to misconstrue the form of the argument. This is NOT an argument of the form: Theism has bad consequences. Therefore we should reject theism and be atheists. Rather the argument goes: the evidence is for atheism. If the subject had no practical import, it would be acceptable not to pay too much attention to this evidence. But there ARE vitally important practical consequences pursuant on being a theist. So we had better pay attention to the evidence, which favors atheism. (This argument is also given briefly in Cartwright 2000: 221–222).

Does the evidence really support atheism over theism? The argument for that conclusion was: our best efforts at description yield a picture of a dappled world. Any supposition that there is a more orderly world "beyond" what we seem to be able to describe would be an idle hypothesis, which sensible methodology excludes.

This is an incomplete account of the evidence. We have copious evidence that the world is a fiercely complicated place, far beyond what human capacities can encompass. So we really have two ways of accounting for the evident failure of epistemic-fundamentalism: The failure might be due to failure of ontic-fundamentalism. Or, ontic-fundamentalism might be true but far from humanly accessible to comprehensive characterization. Lipton likewise makes this point (Lipton 2002: 258). He characterizes the argument from failure of epistemic-fundamentalism to failure of ontic-fundamentalism as an inference to the best explanation. He then criticizes the argument with the observation that known human cognitive limitations comprise an equally good explanation of the failure of epistemic-fundamentalism.[8] And in (Cartwright 2002b) Cartwright agrees with Lipton's evaluation: 'Occasionally I overstate the case for the dappled world. That's because the vision of a dappled world delights me.' (Cartwright 2002b: 271). And, 'The evidence against fundamentalism in physics or economics or elsewhere is not compelling. Nor is the evidence in its favour. The world may be dappled after all, or it may not be' (Cartwright 2002b: 274; 1998: 90).

What about the pernicious methodological consequences? In *The Dappled World* the argument was that the evidence favors atheism over theism, and since the consequences matter we should not ignore the evidence and rest content with a neutral agnosticism. How should this issue be reevaluated now that we agree that the evidence is a draw? All that is required to avoid the pernicious methodology is to acknowledge failure of epistemic-fundamentalism. The problematic methodology is one that presupposes that ontic-fundamentalism obtains AND that we have reasonable prospects of characterizing it—epistemic-fundamentalism is humanly accessible. But deny the human accessibility of epistemic-fundamentalism—even deny its accessibility in the relevant short run—and the methodology to which Cartwright is objecting is undermined. It matters not whether the underlying failure is that of ontic-fundamentalism or of the reach of human cognitive powers. Failure of epistemic-fundamentalism suffices.[9]

GENERALIZING THE ISSUES

Cartwright rejects epistemic-fundamentalism and is agnostic about ontic-fundamentalism, with all of which I heartily agree. Insofar her position would appear to be less radical than, I think, it is often taken to be. But one does not have to be an atheist about the relevant metaphysics to see radical implications in this material, implications for very general traditional ways of thinking about representation, epistemology, and truth. Cartwright calls our attention to the fact that, with few or no exceptions, science reveals the world to us through caricatures, fictions, simulacra, and lies. But if science is carefully applied common sense, how much more must this be true of human representation and knowledge generally? Full exploration of this hunch will be a vast undertaking. In the remainder of this chapter I can do no more than offer some tentative suggestions of things to consider and avenues to explore.

So far the relevant literature has made very little mention of something that I suspect is of central importance to the issues: imprecision, inexactness, and vagueness. To begin with, these characteristics apply to models. No model is completely characterized. In the end what a model tells us, in important part, turns on the practice of application, in theoretical as well as practical respects and contexts, in a network of interconnected model applications.[10]

Once the imprecision/inexactness/vagueness of models is recognized we see that models in science and our representations much more generally share shortcomings. Like models, representations more generally are, with few or no exceptions, to some extent, imprecise, inexact, and/or vague. And as with models, our trusted representations outside of science, the ones we routinely characterize as "true", are also often tinged with the false. I say that my car gets 30 miles to the gallon (Well, not exactly. . . .), and that Peter is a good-natured soul (Yes, of course, he does have his bad moments . . .). Below I will argue that the shortcomings of imprecision and of falling short of "exact truth" are intimately connected.

Here is a general statement of the problem. Our representations, in and out of science, are usually, and perhaps always, in some way imprecise, inexact, and/or vague. And insofar as we take ourselves to grasp what they give us about the world, our representations are, at least frequently, not exactly correct. YET such representations are our way—our only way—of presenting information about the world. If they do not give us truths, in any straightforward sense in either science, or "at home", then how do they provide knowledge and understanding? Earlier I illustrated this dilemma with the example of the hydrodynamical account of the fluid properties of water. I want now to press the point that this problem is extremely general.

CETERIS PARIBUS GENERALIZATIONS, OPEN ENDED CAPACITIES, AND THE DUAL NATURE OF IDEALIZATION AND INEXACT REPRESENTATION

My examples of the fuel-efficient car and good-natured Peter will immediately bring to mind Cartwright's recurrent themes of *ceteris paribus* generalizations and open-ended capacities. On the one hand, we never have exactly true generalizations. At best generalizations hold *ceteris paribus* (Cartwright 1999: 4, 25, 28–29, 49–50, 57,147–148, 151, 175–176, 188). On the other hand, objects have capacities—real characteristics of objects, understood as, or in terms of, natures or tendencies to respond in various circumstances, but not with fixed reliability. Nor is there a "deeper" layer of description that will give exact conditions (exceptionless laws) for activation of a capacity (Cartwright 1989: 3, 158–159, 206, 227; 1999: 28–29, 41, 59, 64–67, 69, 80, 82, 84, 90; 2002a, 430; 2002b: 272–273, 276–277).[11] Furthermore, the points about capacities and *ceteris paribus* laws can be seen as very closely connected, if not just different ways of bringing out the same point (Cartwright 1999: 28–29, Ch. 3; 2002a passim).

I want to look at the phenomena that I think Cartwright has in mind in a somewhat different—and I hope compatible—way. Where Cartwright sees *ceteris paribus* generalizations and open-ended capacities, I see a very general duality between idealizations and imprecise, or vague, descriptions. Suppose, for example, that I say that the floor here is flat. There are two ways that this statement could be understood.

> *Idealization:* "Flat" means FLAT, exactly flat. It is false that the floor is exactly flat, in the sense of a Euclidian plane. But it is close enough to exactly flat so that treating it as if it were flat serves for present purposes.
>
> *Inexact or vague statement:* No, no—of course I didn't mean EXACTLY flat! I meant "pretty flat", or "close enough as makes no difference to exactly flat". We accept the statement as true just in case the departure from exactly flat is irrelevant for present purposes.

In this and in many similar cases, we get the same descriptive or communicative work done with false idealized or with "true" vague statements. These are, semantically, different statements. But they are intimately connected in how they function to describe the world, so that, in their slightly different ways, they get us to the same descriptive place.

The example of the "flat" floor is relatively clean but not representative. We have a good enough idea of what one has in mind by the ideal, limiting case of a perfectly flat surface, and similarly in numerous other examples. But not so for a much larger range of cases. For example, suppose we agree that Alice is smart. This is clearly a vague, inexact statement. But does it have, in analogy to the case of the flat floor, an alternative form that works in terms of an "ideal case" of "smartness", the "Platonic form of smartness", as it were?

If taken too literally, such a supposition is farfetched. Nonetheless, thinking in terms of an ideal gives a pretty good description (or model, if you will) of how we operate when we characterize such statements as "true". In what we take to be clear cases we treat "smart" as applying unproblematically, not allowing ourselves to be distracted by complications of the vagueness of "smart". Insofar, we treat "smart" as if we had some ideal in mind.

Such considerations incline me to speculate that ALL human representation works in terms of something like the foregoing duality of imprecise/inexact/vague descriptions that, in suitable circumstances, we characterize as "true"; and idealizations/Platonic ideal elements/determinate characteristics that we know do not truly apply but are, in certain circumstances, usefully treated as if they did. I have done no more, of course, than to illustrate the line of thinking, with a few (clearly generalizable) examples and not yet extracted from the examples any sort of precise account, let alone argued for such an account in detail. I hope, nonetheless, that the discussion will have attested to the worth of perusing this line of thought.

We can also learn more about what such an account might look like by examining three different ways in which the foregoing line of thinking will be challenged.

EXACT CAPACITIES?

Cartwright has argued at great length that capacities are real; and so, presumably, that they are perfectly determinate characteristics of things; and so, presumably, that attributions of capacities are simply true or false and not subject to either the failings of imprecision/inexactness/vagueness or of the kind of idealization we find in models.

The argument for capacities is that to make sense of scientific practice we must suppose that things have capacities, understood as determinate characteristics of objects or situations (Cartwright 1989: sections 4.3, 4.4: 227; 1999: 59, 64, 77). In Cartwright's view there is latitude or indeterminateness involved, but this comes in through indeterminateness in whether something does what it has the capacity to do. Even when the activating conditions obtain, or when at least all the activating conditions that can be described in the relevant theory obtain, the capacity may or may not successfully operate to produce (one of) its characteristic resultants. For example aspirin has the capacity to relieve headaches in spite of the fact that it does not always achieve this result (Cartwright 1989: 3, 136, 141; 2002: 427). The indeterminateness is in whether the aspirin's capacity to relieve a headache will successfully operate, not in whether aspirin has this capacity or in the determinateness of the capacity itself.

But is there any such thing as a determinate capacity that aspirins have to relieve headaches? For this to be the case, what is involved in having a headache would have to be completely determinate and likewise for what counts as successful relief. I submit that the example of the capacity to relieve headaches is relevantly similar to my example of being smart. We can think of the expression 'capacity to relieve headaches' as vague, as taking it to apply in cases that are clear, sufficiently for purposes to hand. Or we can think of it as naming a determinate capacity, operating as an idealization, supposing that there is a completely determinate capacity that is in question (the Platonic ideal!!). But attribution of any such completely determinate characteristic is not, strictly speaking, correct; any more than it is strictly speaking correct that the floor is (literally) flat; or, though the reasons are somewhat different, any more than that there is a completely determinate characteristic of being smart. ("Smart" is, after all, also a capacity term!)

Readers of Cartwright will also know that if what I claim for being smart and the capacity to relieve headaches is accepted there will be a difference in degree but no elimination of the problems if we retreat to the less imprecise realms of the exact sciences.

Cartwright may disagree with these suggestions about capacities, insisting that there are completely determinate capacities and that the indeterminateness arises only in how they are exercised. But even if so, I don't think that we are as far apart as it might seem. What I say is consistent with taking Cartwright's arguments to show, not that there are (completely determinate) capacities, but that to make sense of scientific practice we must acknowledge the role of open-ended capacity terms. In fact we now have three characteristics to which a more thorough investigation will have to pay attention: (a) false idealizations that are (at least taken to be) exact; (b) inexact/indeterminate/vague terms and statements using such terms that we judge to be "true"; and (c) even insofar as we have a case in which we need not take the first two problems into account there is often, and perhaps always, indeterminateness in the exercise of capacities.

I have speculated that there are close connections between (a) and (b). It might seem that (c) is clearly distinct from (a) and (b). However, in a more thorough investigation this question bears closer examination, as does the ways in which all three of these considerations are in strikingly close step with the open-ended character of *ceteris paribus* generalizations and, more generally, with the inexact and open-ended character of models in their use in representation.[12]

STRAIGHT-OUT TRUTHS?

Many will reject the claim that we are rarely, if ever, possessed of statements that we can confidently endorse as straightforwardly "straight-out true". The challenge is useful because it can be met only by acknowledging that, in many cases, failure of "straight-out true" arises in ways other than simply being in some way false. There is an important and interesting range of ways in which statements fail of exact truth.

At the workshop Carl Hoefer[13] offered the example: the mass of the earth is greater than the mass of the moon. Short of Cartesian epistemic catastrophes, how could this statement be anything other than true, with no qualifications whatsoever?

But what does the statement that the mass of the Earth is greater than the mass of moon come to? Depending on the details of your semantics, it is either asserted or presupposed that there is an M_e, the mass of the earth, and an M_m, the mass of the moon, and then asserted that $M_e > M_m$. But there is no such thing as the utterly determinate mass of the earth or any utterly determinate mass of the moon. The Earth and the Moon are always acquiring and losing bits of stuff. For that matter, there is nothing completely determinate about what the Earth and the Moon are.

The standard response to this worry is supervaluation. “The mass of the Earth” and “the mass of the Moon” are acknowledged to be vague expressions. But the statement, ‘The mass of the Earth is greater than the mass of the Moon’ is said to be determinately true if all acceptable ways of making the statement precise result in a truth. That is, we have a determinate truth if, for any acceptable values that could be assigned, respectively, to “the mass of the Earth” and “the mass of the Moon” (any “admissible precisification”), the first is larger than the second. I don’t see that this move helps with the problem at all. The vagueness of terms in the initial statement have simply been foisted off on the vagueness of “acceptable” or of “admissible precisification”.[14]

But the example is useful, for it illustrates the circumstance that there can be a variety of ways in which our statements fall short of being “straight-out true”. In the present example, we see that role of idealization may operate in the presuppositions of a statement rather than in the statement itself.

Here is another test case. Carl is holding a white mug, looking at it in a good light. The object to hand is a clear case of a white mug, and Carl says, ‘The mug I am holding in my hand is white’. Surely what Carl has just said is unproblematically, “straight-out” true.

But WHAT has been said? Let’s try the option of saying that a (true) proposition has been affirmed. And let’s suppose that a proposition either is a range of possible cases or functions to pick out a determinate range of possible cases. But then, since there is no sharply determinate range of cases that count as ones in which what Carl is holding in his hand is a white mug, what Carl has asserted cannot be reconstructed as assertion of a determinate true proposition.

Here is a somewhat different way to bring out the problem. Suppose I am talking to Carl on the telephone and I hear him say, ‘The mug I am holding in my hand is white’. How should we understand the information I have just obtained? I know that Carl is ruling in certain possibilities and ruling out others. But as “white” and “mug” are vague, the range of possibilities ruled in and out is indeterminate. So there is nothing perfectly determinate that I have learned that I can say is unproblematically true, and so also nothing perfectly determinate and true has been asserted.

Perhaps the context eliminates the imprecision by supplying a perfectly determinate standard for what is to count as white and as a mug. This suggestion would not seem very plausible: How does—how could—the context eliminate all imprecision? At the very least the needed complete precision would not be something that is humanly accessible and so not relevant to what can be grasped and understood by language users.[15]

Perhaps so much the worse for propositions. What about simply taking Carl’s utterance to be unproblematically true? I know Carl to be careful, even circumspect—he does not make such assertions unless the case, here of being a case of a white mug, is clear. So, it will be concluded, the utterance can be taken to be unproblematically true.

This approach yields an understanding of unproblematically true utterances that is only as clear as "clear case" is itself clear; and of course, "clear case" is itself indeterminate and vague. Most important to the indeterminate status of "clear case" is that it is relative: clear enough for WHAT? Change what is at issue and the case may no longer count as clear. This is so, and relevantly, even if the case is "so clear" that it will count as a clear case of a white mug for any issue that might, in practice, come up—what is sometimes called a case of a "moral certainty" (another vague, indeterminate notion, of course . . .).

Both the examples of this section illustrate an alternative way in which statements or utterances can somehow fail to be "straight-out true": they fail, at least in the first instance, not because they assert something that is completely determinate but somehow, "in some little way" false; rather they fail because they fail to say anything completely determinate. It will prove a challenging—and interesting—part of the larger project to better understand how this kind of failure relates to others. For the present I note only that it is easy to multiply such examples, examples that illustrate another way in which most, if not all, of our representations are characterizable as unequivocally true or false only as an idealization.[16]

REQUIREMENT OF AN EXACT STANDARD?

In response to the claim that all our representations are in some way inexact it will be said that, even so, this shows nothing about our conception of what it is for a statement to be true because we must have the conception on which to be true is to be exactly true, with no qualifications, to give any sense to what it is for a representation to be inexact (the "contrast-requires-an-opposite" argument).

To begin with, I do not take any of the foregoing considerations to show that any conception of truth as exact, exact in every sense, is unintelligible. Quite the contrary: we can model exact conceptions of truth, which I take to be sufficient for its intelligibility. Indeed, we do this all the time: You model me as using "Robert Redford" to refer to a perfectly determinate individual and "is male" to cover a perfectly determinate property and so my use of "Robert Redford is male" to express a truth that is not qualified in any way. Since we have some idea of what it would be for the world to be as characterized by these models, we have some idea of what exact (correspondence!) truth would be. Indeed, I suspect that it is because we so ubiquitously think in terms of such models that the conception of truth as exact has such a grip on us. Let's call this the "ideal model conception of exact truth".

The trouble is, as I have argued in previous sections, we rarely, if ever, actually have such truths.

I cannot hope here to unravel all the questions surrounding the contrast-requires-an-opposite argument in its application to truth: Is the argument a good one?[17] If so, is the ideal model conception of exact truth sufficient to meet its requirement? If not, what does this show about how we should think about the "truths" that, in practice, we have? But let me get the discussion started.

We can make sense of exact truth as an idealized model. And, like so many idealizations, it is extremely useful for many purposes. But I am skeptical as to whether this model is of much good when it comes to worries about how to understand what is involved in inexact truth. If one is worried that to understand the import of an inexact truth one needs to understand it in reference to exact truths, how is one's understanding aided by mandated comparisons that can never be made because the comparison points are so hopelessly out of reach? Or, if one persists in the hope that the needed exact truths will one day be found, how are we to understand our inexact truths in the interim? What is to be made of our current understanding if the hoped for utopia never arrives?

I also don't think that the demand for a comparison point of exact truths is sound. For a quick way to see this, imagine that we were to banish all talk of (literal, exact, not qualified any way) truth from our discourse and thinking. Instead we would use expressions such as "accurate", as in 'that's plenty accurate enough for present purposes' and similarly, "precise", "correct", "exact", and "good enough for government work". We could retain use of "true", but only when understood in senses such as 'That's true enough', 'That's very true', 'How true is that?' (with "correct", "exact", etc. understood similarly, of course).

I have heard some protest that one can talk this way, of course, but only against the background of an understanding of exactly true, exactly correct, accurate, etc. But I don't know what the argument for this claim is. It seems very easy to conduct the thought experiment of imagining a linguistic community arising with practices of evaluating as to accuracy, but not in any way thinking in terms of any ultimate standard. Once this thought experiment is carried out, I no longer see why we should not take the community in question to be ours, with the idea of exact truth tacked on because it is a useful idealization when we can put aside the possible need for further refinement.

Let's dig into this dispute in a little more detail. "Inexact", as I have used it in the foregoing paragraphs, can be broadly understood in two kinds of ways, reflecting the duality I have discussed in the section 'Exact capacities'. First, one may mean that a representation is vague, not providing a precise content. On such a reading, the contrast-requires-an-opposite objection comes to the same as saying that in order to provide meaning to the claim that a statement is inexact one must suppose there to be some characterization that is exact in the sense of being completely precise as to its content: I take the claim to be that the way in which the inexact representation falls short must be understood with reference to some kind of comparison with the presupposed content-exact standard (it not being clear whether the demand is for an exact standard that is in practice accessible or only supposed available "in principle" or "in concept").

On the second kind of reading of "inexact", a representation is treated as exact in its content but taken to fail in some way in how it describes the way things are. The representation may characterize something as 5.72 centimeters long when the true (note the use of "true"!) value is 5.73 centimeters. Or the representation may be an idealization, as in the hydrodynamic example. For such cases it will be said that any understanding of a mistake in describing the facts presupposes a standard of correctness, the way things in fact are, and so a conception of an error-free characterization, and so a conception of exact truth that is free of error of any kind. (Again, it is not clear whether the demand is for an exact standard that is in practice accessible or whether it is claimed that we must understand the idea of inexact representation in terms of reference to such characterizations "in concept" whether or not any such characterizations are actually accessible to us.)

Let's start with inexactness in the sense of vagueness. When the vagueness doesn't matter, we treat and think of a description as if it were exact. When we run into trouble we drop the form of description in favor of an alternative relative to which the problematic imprecision of the former description can be characterized. For example, when a description of people in terms of "tall" and "short" breaks down, we drop "tall" and "short" and start giving people's height in centimeters. Generalizing, to say that a description fails to be completely content-precise can be understood simply in terms of the possibility of an alternative description being more precise, being a description from the point of view of which the vagueness of the first can be clarified.

The objection will be that, short of some conception of description that is without any imprecision, any alternative description is no more than just that: an alternative. Relative to the second, the first description may be less exact, but that does not make the second more exact than the first in any absolute or context-independent way.

But why is that a problem? The objection concedes that there is a relative sense to "less vague" and "more vague" (always, it is crucial to remember, in specified respects): that from the point of view of one description another can be described as giving a less determinate characterization (in specified respects). What the objection does not make clear is why such relative characterizations are not sufficient. We use one pattern of description when this method works well enough. When it breaks down we replace or supplement the first with another, from the point of view of which the failings in precision of content of the first can be described. We know that relative to other possible descriptions the new description will also count as inexact; but that is of no present concern as long as the new description meets our current practical and intellectual needs. In such circumstances, it is appropriate to treat the new description as exact—always meaning adequate to present concerns. In practice we don't ever need, and in fact—I would contend—never have descriptions that are completely precise in content. Why then are we said nonetheless to need them "in concept"?

The argument for the other sense of "inexact" follows a similar pattern. Taking a description for which we disregard any possible inexactness in content, to say that it is inexact in the sense of somehow misrepresenting the facts can be understood in terms of the possibility of some alternative description that is more factually accurate than the first, and from the point of view of which inaccuracies of the first can be illuminated. The objection will take the same form as before: How is "more factually accurate" to be understood? Short of a conception of an ultimate standard, free from all inaccuracies, such an alternative is no more than just an alternative, a different descriptive form with no substance to the claim that it is "more" factually accurate than the first.

The response to this objection is that not all alternatives are of equal standing. Some are better justified than others by our familiar standards of justification that include better fit with both our practical and intellectual needs. In the familiar cases—I'll give an example in a moment—the justified alternatives are not themselves free of all factual inaccuracies, so these justifications don't work by justifying alternatives as being exactly correct. We take such justifications to support the claim that a new description constitutes an improvement in the sense of being free of certain errors to which the prior description was liable and that can be described from the point of view of the new description. Why, to make sense of this process of successive refinement, do we have to think of it in terms of some limit point of completely error-free description, any more than we need to think of the concept of moving faster in terms of some limit point of infinite speed?[18]

Let's work through the example of Newtonian mechanics. For more than two centuries Newtonian mechanics was taken to provide an exactly correct description. We have revised that evaluation—and in a variety of ways. First, with hindsight we now appreciate that there was not one thing or form of description, "Newtonian mechanics". There was latitude in what was included; evolving differences of substance, not just formulation; and at any one moment points of variable interpretation within the theory.[19] These facts illustrate our situation with respect to inexactness in the first sense. Second, with the development of quantum mechanics and both special and general relativity, we also have come to see Newtonian mechanics as not completely accurate in the sense that it misrepresents the way things are.

What makes us think that Newtonian mechanics is inaccurate? Observations that resisted all efforts to be fitted into the Newtonian framework were never enough by themselves to shake physics out of its confidence that Newtonian physics offered completely error-free descriptions. What was required were new theories, quantum and relativistic mechanics, that provide points of view from the vantage point of which we can say, in great detail, in what ways Newtonian mechanics falls short and why. All of which makes perfectly good sense even when keeping clearly in mind that the new theories are themselves not perfectly precise and accurate.[20] I suggest that we can take the relation between Newtonian mechanics and its successors as an exemplar. It is not clear why we need more, when we say that a theory falls short of complete precision and accuracy, than that it is always possible, at least in principle, to get a new descriptive vantage point that bears a similar sort of relation to our current theories that they bear to their predecessors.

It is a frequently mentioned worry about the Peircian conception of truth in terms of the end of inquiry that we will never get to the "final theory". The present raises the question: why do we need to think in terms of a limit point at all?

To summarize the suggestions of the last half of this chapter: Cartwright observes that in science we always deal in idealizations and in models that are not completely precise in their characterization; and insofar as they are precise as representations, not completely accurate; and insofar as accurate, only accurate enough within some limited domain. She notes that the generalizations that science offers are open-ended *ceteris paribus* generalizations. And she argues that understanding science requires characterization in terms of capacities that are open-ended in their operation, and, on my further interpretation, involve use of open-ended capacity language. I want to conclude that this kind of characterization applies much more broadly to all human representation and knowledge.

Again, I am not suggesting that a conception of truth as exact, in both senses of "exact", is unintelligible. Quite the contrary. Thinking of our representations as exact makes perfectly good sense. But it is itself an idealization—a model. Rather, I am claiming that this idealization, like all models, has serious shortcomings when taken to apply to the representations that we actually have and that there are other models worth developing and considering. Reflexivity of course requires that we expect that these new models will have shortcomings of their own, which provides the immediate response to the complaint: but won't you insist that the account you are developing will be the one that is TRUE! What I am claiming is, rather, that few, and perhaps none of our representations are completely free of inexactness in both senses; that exact truth is not the only viable model of the characterization of the objective status of our representations; and that we should be developing alternative models that do better with the kinds of considerations that I have been addressing throughout this chapter. Exact truth is not something we actually have. That fact alone should lend some plausibility to the claim that we should be able to make good sense of our representational situation in the world without appealing to exact truth in hand at any point.

A NEW THEORY OF TRUTH AFTER ALL?

The present effort has mined both examples and general considerations to suggest that in application to the representations we actually have, the way we commonly think about truth is problematic. And it is hoped that the exposition of the difficulties will also begin to suggest some of the kinds of things we might consider by way of alternatives. By this point it will have become clear that the proposed program has much in common with traditional forms of pragmatism. We are inclined to attribute truth when and insofar as our representations serve a wide range of practical and intellectual human purposes.

But just as the affinity to pragmatism emerges it may look more and more as if I am flouting Fine's admonition that truth needs no interpretation. In particular, Fine argues specifically against pragmatic interpretations of truth. He takes pragmatic accounts to be a variety of acceptance accounts that characterize truth in terms of what some specified class of (real or ideal) agents would accept under specified kinds of circumstances (Fine 1986a: 137–138). Alternatively, he characterizes 'the pragmatic conception of truth [as one that] confounds truth with reliability' (Fine 1986b: 154–157).

Fine advances pressing problems with pragmatism as he construes it. Something can be accepted, in whatever sense specified, or can be reliable, and fail to be true for all that (Fine 1986a: 140; 1986b: 175). (Indeed, the import of Cartwright's observations, and I urge they be generalized, is that our accounts, the ones we accept and find so reliable, ARE false.) Furthermore, acceptance and reliability must be understood counterfactually, and it is hard to see how that is to be understood without relying on some prior understanding of truth (Fine 1986a: 140–141; 1986b: 170). More generally, Fine inveighs against anything that would seek to give a theory or account of truth (Fine 1986a: 133, 149; 1986b: 175–176) or something giving the "essence of truth" (Fine 1986a: 142, 149; 1986b: 174–175): for example, accounts that insist that for a statement to be true is for it to correspond to some independent reality or for it to serve certain human ends.

There may indeed be tension between such admonitions and my call for new models. But not insofar as Fine meant to counsel against efforts to find fixed, context-independent ways of thinking about truth. I am urging an examination, or at least an awareness, of certain aspects of our practices of accepting things as true. I suggest that all our representations are flawed but in a variety of ways that can be expected to bear complex relations to one another.[21] And I am urging development of alternative models that address these circumstances. Such a program in no way commits one to searching for any context-independent "essence of truth".

These considerations are shot through with pragmatic elements but not ones that would constitute anything like some kind of "essence giving" account of truth that Fine abjures. We treat our current representations as satisfactory; we treat them as if they were true in the traditional sense. When we run into trouble we look for better representations, not "better" in an unqualified or context-independent sense, but representations that do not give rise to the problems of what went before, and, at least often, from the point of view of which we can understand the pitfalls of the predecessor.

Human interests—both practical and intellectual—guide these choices. But that they guide the choices does not make pragmatic considerations anything deserving the epithet, "essence of truth". We drop the pragmatist slogan, 'To be true is to work' in favor of 'To be true enough is to work well enough.' The latter is still no more than a slogan, a guide, a rule of thumb for the kind of considerations that are likely to be relevant in individual cases. What is it "to work"? To work "well enough"? Individual cases will provide local, idiosyncratic refinements of these very general and imprecise ideas. This accords exactly with Fine's view of truth as open-ended and evolving: 'If pressed to answer the question of what, then, does it mean to say that something is true . . . NOA will reply by pointing out the logical relations engendered by the specific claim and by focusing, then, on the concrete historical circumstances that ground that particular judgment of truth' (Fine 1986a: 134, 149; 1986b: 173–175). What counts as true—objectively, not just in our evaluations—is a confluence of the interaction between us and something else, something that we cannot exactly describe, something that we can always understand only partially and only from one or another point of view, these points of view themselves being, in part, products of our past voyage of discovery.

CONCLUDING THOUGHTS

The foregoing encourages very different ways of thinking about truth than what many of us have had in mind. It probably counts as some kind of pragmatism, but it is not the pragmatism with which Fine finds fault. These are ways of thinking about truth that study of Cartwright's work forces us to take seriously.

But, surely, you may insist, there REALLY IS an exact ways things are! Well, to so insist is, if not (ontic)-fundamentalism, something fundamentalism must presuppose. I am an agnostic. So don't ask me.

NOTES

1. Many thanks for very useful comments from Daniela Bailer-Jones, Ron Giere, Carl Hoefer, Wolfgang Spohn, and Mauricio Suárez, and for discussion at the workshop from Nancy Cartwright and many other participants.
2. In *How the Laws of Physics Lie* (especially Ch. 6) Cartwright takes phenomenological laws to be much more accurate or truthful than fundamental theoretical laws. But phenomenological laws still work in terms of "prepared descriptions", that is in terms of models of the raw data, or what are often called "data models" (Cartwright 1983: especially Ch. 7). In 'The tool box of science' (Cartwright et al. 1995: 139–140) Cartwright distances herself somewhat from the application of "lies" in application to basic theory, emphasizing the circumstance that theory does not apply to the world directly but functions as one among a number of tools in fashioning models that do the representing. Of course models are never completely accurate—they provide

a "simulacrum" of the real world (Cartwright 1983: Ch. 8), and this is as true of the highly theoretical models fashioned with fundamental theory as it is of the data models to which they apply. I will nonetheless continue exposition in terms of "lies": it being understood that this is a perhaps exaggerated way to emphasize the fact that scientific representations are rarely, if ever, exactly correct.

3. For much more on Cartwright on models, see the contributions to this volume by Bailer-Jones, by Giere, and by Morrison.
4. Readers will recognize this as making the same point as Putnam's example of explaining why the round pegs won't go through the square hole (Putnam 1975: 295–298).
5. Or at least, Fine's (1986 a, b) would appear to present a conception of NOA that is committed to such an attitude towards well-confirmed scientific conclusions—this is just the "core position", as quoted above from (Fine 1986a: 128). But in 'Fictionalism' Fine appears to construe NOA as compatible, at least in spirit, with Vaihinger's fictionalism:

   > [T]he industry devoted to modeling natural phenomena, in every area of science, involves fictions in Vaihinger's sense. If you want to see what treating something "as if" it were something else amounts to, just look at most of what any scientist does in any hour of any working day. (Fine 1993: 16)

   The problem is, what is it to 'accept the results of scientific investigations as "true," on par with more homely truths'? By addressing this problem the present essay should, by its end, suggest a reconciliation of these apparent contrasts in Fine's discussion of NOA.
6. See Elijah Millgram's (to appear) screamingly funny, because devastatingly accurate, parody of the practice of attributing truth and falsity to sentences or propositions in terms of a practice of attributing truth or falsity to portraits.
7. Cartwright writes about fundamentalism in two different but closely related ways. On the first characterization fundamentalism concerns exact, universally applicable laws, what above I called surrogate realist representations (Cartwright 1999: 24, 34, 37). On the second characterization fundamentalism urges an exact and comprehensive description by science (Cartwright 1999: 27, 31—some of these citations can be read in terms of both characterizations). The two characterizations are also by the contrasts that Cartwright offers, on the one hand with *ceteris paribus* laws and the open-ended operation of capacities; on the other hand with Neurath's conception of "the system" as 'the one great scientific theory into which all the intelligible phenomena of nature can be fitted. . . .' (Cartwright 1999: 6) The two conceptions are, of course intimately related, in particular, the second presupposes the first. My considerations will be exclusively concerned with fundamentalism understood in terms of exact, universally applicable laws.
8. Lipton also uses the term "agnostic dappling" 'according to which we do not know whether laws rule' when we are unable to model the phenomena (Lipton 2002: 257–258).
9. Again, Lipton makes essentially the same point (Lipton 2002: 259–260). It has also to be acknowledged that the politics of this issue complicate the intellectual purity that I am essentially urging in the text. Consider, for example Weinberg's congressional testimony in support of the SSC (Weinberg 1987: 437)
10. See Teller (2004b), where I explore this idea and its relation to *ceteris paribus* generalizations. See also, Bailer-Jones (2003). There is a question as to whether, or to what extent, the imprecision/inexactness/vagueness is a characteristic of models as opposed to their application. Nothing that follows should

turn on this question, so I won't try to sort it out here and write simply as if these limitations are characteristic of the models themselves.

11. At least no humanly accessible such level of description—pursuant to Cartwright characterizing herself as an agnostic rather than an atheist.
12. Some of these considerations are examined in my (Teller 2004b)
13. I'm very appreciative of Hoefer's further correspondence with me, which helped in formulating the material in this section.
14. Keefe argued this verdict in detail (2000: Chs 7, 8). For reasons detailed by Keefe, I regard other existing efforts at dealing with such examples of vagueness as even less sound.
15. The suggestion is an instance of the general "epistemic" approach to vagueness, and the brief objections are instances of commonly voiced difficulties with this account. Williamson (1994) provides a spirited exposition and defense.
16. At the workshop Stathis Psillos was quick to remind us that such failures to provide exact truths need not count as failures in a more general sense. From many points of view, the flexibility that such indeterminateness provides can be a tremendous advantage in meeting human descriptive and communicative goals.
17. For some general discussion of the principle that a contrast requires an opposite and examination of why many of its applications are fallacious, see Passmore (1961: Ch. 6).
18. The idea that all speeds are finite and there is no infinite velocity, as described in the Newtonian framework, makes perfectly good sense, which is all the analogy needs. So the point is not compromised by thinking of the speed of light in relativistic physics as an actual limiting velocity. Furthermore, thinking of the speed of light as a velocity is troubled. It is not the velocity of any massive object. On the light cone, "time stands still". And some will argue that *c* is better thought of, not as a velocity, but as a conversion factor between measures of space and time.
19. See Wilson (1998) and van Fraassen (2002: 145–151)
20. Nor will it work to suppose that the newer theories are further way stations on a linear progression down some "Royal road to the truth". This is argued in detail in Teller (2004a).
21. But keeping Psillos's admonition in mind.

REFERENCES

Bailer-Jones, D. (2003) 'When scientific models represent', *International Studies in the Philosophy of Science*, 17: 59–74.

———. (this volume) 'Standing up against traditions: Models and theories in Nancy Cartwright's philosophy of science'.

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Oxford University Press.

———. (1989) *Nature's Capacities and their Measurement*, Oxford: Oxford University Press.

Cartwright, N. (1998) 'Comments and replies', in M. Paul (ed.) *Nancy Cartwright: Laws, Capacities and Science*, Munich: Lit Verlag.

———. (1999) *The Dappled World*, Cambridge: Cambridge University Press.

———. (2000) 'Against the completability of science', in M. W. F. Stone et al. (eds) *The Proper Ambition of Science*, London: Routledge.

———. (2002a) 'In favor of laws that are not *ceteris paribus* after all', *Erkenntnis*, 57: 425–439.

———. (2002b) 'Reply', *Philosophical Books*, XLII: 271–278.

Cartwright, N. et al. (1995) 'The tool box of science: Tools for the building of models with a superconductivity example', in W. E. Herfel et al. (eds) *Theories and Models in Scientific Processes*, Amsterdam: Rodopi.

Fine, A. (1986a) *The Shaky Game: Einstein, Realism and the Quantum Theory*, Chicago: University of Chicago Press.

———. (1986b) 'Unnatural attitudes: realist and instrumentalist attachments to science', *Mind*, 95: 149–179.

———. (1993) 'Fictionalism', *Midwest Studies in Philosophy*, XVIII, 1–18.

Keefe, R. (2000) *Theories of Vagueness*, Cambridge: Cambridge University Press.

Lipton, P. (2002) 'The reach of law', *Philosophical Books*, XLII: 254–260.

Millgram, E. (forthcoming) *Hard Truths*.

Morrison, M. (this volume) 'Models as representational structures'.

Passmore, J. (1961) *Philosophical Reasoning*, London: Duckworth.

Putnam, H. (1975) 'Philosophy and our mental life', *Philosophical Papers*, 2: 291–303, Cambridge: Cambridge University Press.

Teller, P. (2004a) 'How we dapple the world', *Philosophy of Science*, 71: 425–447.

———. (2004b) 'The law idealization', *Philosophy of Science*, 71: 730–741

van Fraassen, B. (2002) *The Empirical Stance*, New Haven: Yale University Press.

Weinberg, S. (1987) 'Newtonianism, reductionism and the art of congressional testimony', *Nature*, 330: 433–437.

Williamson, T. (1994) *Vagueness*, London: Routledge.

Wilson, M. (1998) 'Mechanics, classical', in E. Craig (ed.) *The Routledge Encyclopedia of Philosophy*, 6: 251–259, Routledge: London.

Reply to Paul Teller

Half of Paul Teller's piece is a direct comment on my work, especially the first book, *How the Laws of Physics Lie*, and half develops his own view that the laws do not lie because they are not intended to make exact claims. I have three brief comments on the first half, then I turn to the second.

First, it is ontic fundamentalism, as opposed to epistemic, that I have wanted to develop and defend, indeed that I would bet on if I had to bet. But it is not the kind of bet that makes sense because these claims are so unlike any that we know how to gather proper evidence for.

Second, Teller suggests that my remarks imply that the "evidence is a draw" between ontic versus epistemic fundamentalism. I hope they do not imply this. Evidence, if the term is to do the kind of legitimizing work we put it to, is a serious matter. The kinds of facts we usually cite in discussing fundamentalism have far too loose a connection to either hypothesis to count as more than nods in some direction. Nor do I see what we could do to produce serious compelling evidence.

Third, Teller suggests that capacity claims must be "determinate" or "exact" because I argue that capacities are real. Why though does he not take the natural ontological attitude here as elsewhere? If capacity claims are central to science, as I argue, and Teller can successfully argue that most scientific claims are vague, why suppose capacity claims to be different from any of the rest?

Much of my work treats central claims of science that are radically false or are so abstract and removed from the empirical world that they have little empirical content. Teller treats different claims, those that are not true—on standard accounts of truth—but are nevertheless "approximately true". For these claims Teller offers a bold programme for science, turning on its head centuries of admonitions to make our scientific claims exact (not vague) and precise (narrow in range). With it he also turns on its head the usual problem of what I sometimes call "Aristotelian" abstraction. Aristotle worried about the relation of geometry to physics because it seems that ideal geometrical objects are not found in the world that geometry is supposed to treat: A real line always has width, a geometrical line never does. Following Aristotle's thought, Carl Menger argues that economics could—and should—be an exact science but that it would not then be true of "full empirical reality". Pierre Duhem argued that science itself is exact even though the facts we confront cannot dictate an exact scientific description. And one of the rallying crises of Positivism has been the call for exact science. My own hero, Otto Neurath, for example, urged a Positivist view that Teller seconds, that we can only compare scientific representations with other representations, not the world itself. But scientific representations for Neurath, as for the other Positivists, should be exact. This contrasts with the concepts that describe the evidence for science and in terms of which science will be put to use. These are "Ballungen": dense clusters with rough edges.

All these suppose that science can be exact though it is not likely that the world is. At the least, the world may not supply evidence that allows us to settle for one precise value of an exactly defined quantity as opposed to nearby ones. Teller however takes the terms themselves to be vague, just as "white", "headache", and "relief" are. Since they are vague, they can be judged to hold truly across a range of values, so what looks like getting it wrong really is not. A central project I would urge Teller then to take up is to catalogue the relative advantages of his view, which permits science to be inexact, over these others who cling to exactness in science though allowing that its claims may thus not be literally true of the "fuzzy" world.

In particular there were two important reasons stressed by both Popper and the Positivists for demanding that the claims of science be both exact and precise: To ensure that we know exactly what is being asserted and that what is being asserted is informative. This in turn contributes to the joint scientific goals of understanding and prediction. Genuine understanding requires, they maintained, that we know what our claims really assert, and vague predictions are of little use for constructing the kinds of precise technologies we expect from modern science.

The difference between the two converse ways of admitting fuzziness may not seem all that great when we look at just a single scientific claim. But science's real predictions depend on a network—usually a vast network—of interconnecting claims; consider for instance any of the hugely long complex evidences we use in mathematical physics to make a natural prediction. The semantics that allows us to link the claims together and make derivations supposes that each claim claims what it literally says. If we are instead to suppose that the mathematical claims themselves are vague, what inference rules will link them together? If we have fuzzy claims, will we not need fuzzy logic and, worse, fuzzy mathematics?[1] I hope not because fuzzy mathematics is a really tall order.

I should like to close by putting in a word for working as well on a different problem from the problem of inexactness that Teller tackles[2]—the problem of how to deal with all the other cases where the claims of a theory seem radically wrong and need ad hoc changes to describe the facts correctly. Perhaps though Teller does not believe that these cases—which look to me to be all too prevalent—really exist, since he dislikes *The Dappled*

World, which tries to show how the claims of science might be true despite often needing ad hoc additions to get the facts right.

NOTES

1. This does not deny that some claims in physics are about measured quantities, where reported values are expected to have measurement error. In precise practical applications, we often see fairly sophisticated ways to deal with this, such as keeping error charts, hypothesizing when errors should be independent, etc. But it seems this is always against a background where the bulk of the claims—especially laws and principles, like the equations of GTR or various conservation laws—are treated as exact.
2. As well as the idea that we judge statements false relative to specific improved claims, which I have not mentioned.

Part II

Causes and Capacities

6 Models, Metaphysics, and Methodology[1]

Ronald N. Giere

INTRODUCTION

This chapter constitutes my first attempt publicly to comment on Nancy Cartwright's philosophy of science. That I have not done this earlier is primarily due to the great similarities in our views on topics where our interests overlap.[2]

But Cartwright's work also covers topics I have never seriously considered, such as the use of linear models in economics and the measurement problem in quantum mechanics. Even the subject of probabilistic causation, to which I once contributed, is not one I now feel confident in examining in any detail. I will concentrate, therefore, on her views regarding the nature of scientific theories, laws, models, and causality in general—topics at the forefront of my own current thinking. More specifically still, I will focus on the picture of classical mechanics she presents in *The Dappled World* (Cartwright 1999).

CARTWRIGHT'S PICTURE OF CLASSICAL MECHANICS

Let us begin with the most general principles of classical mechanics, in particular, Newton's second law in it simplest form, $F = ma$. This looks like a statement. So apparently we can inquire about its semantic and epistemological features. What are the referents of the terms? Is the resulting statement true or false? Is its truth empirical or a priori? If empirical, what evidence is there for believing that the statement is true?

It helps first to consider answers that Cartwright rejects. She is quite clear that this statement should not be understood as an empirical generalization asserting the universal association of occurrent properties, forces of a given magnitude, and products of mass and acceleration of the same magnitude. Nor should it be understood as a modalized generalization asserting the necessary occurrences of the designated properties.

I think her main positive strategy begins by asking how one would *test* the statement that $F = ma$. The answer is that one can't test it in this form.

The fundamental principles of classical mechanics don't tell us what counts as a force. We don't know where to find the forces apparently referred to by the principles. To test the principle we must first introduce what she calls "bridge principles" or "interpretive models".[3] The historically most significant is the gravitational law for the force between two masses in free space, $F = Gm_1m_2/r^2$. Others include a constant force in a uniform gravitational field, $F = -mg$, and the linear restoring force resulting in harmonic motion, $F = -kx$, where x measures the displacement from an equilibrium position. So what then do we say about the statements resulting from substituting these expressions in the original statement of Newton's second law: $a_1 = Gm_2/r^2$, $a = -g$, and $a = (-k/m)x$? Are these to be understood as universal (necessary?) associations? Again, Cartwright's answer is no.

Take the simplest case, free fall in a uniform gravitational field. Here she invokes Otto Neurath's example of dropping a banknote from the steeple of Vienna's St. Stephen's Cathedral into the square below. Clearly the downward acceleration of the banknote will not follow the simple "law" $a = -g$. It is too light, of irregular shape, and there may be air currents or a wind. It will just float haphazardly to the ground somewhere in the square below. This case is to be contrasted with that in which one drops instead a heavy coin which pretty much will exhibit the indicated constant downward acceleration. So it is not the case that the "law" for free fall is never instantiated (at least to a fairly good approximation). It is just not universally valid. It works for some things and not others.

A contrary view is that Newton's principles apply equally well to the banknote in the sense that, at every instant, the acceleration of the banknote, in whatever direction, is proportional to the total force on the banknote at that instant, whatever the sources of the force might be (gravity, air friction, air currents, etc.). This is an expression of what Cartwright calls "fundamentalism", the view that there are true laws in force always and everywhere. She urges us to resist fundamentalism (Cartwright 1999: 34). But she does not reject altogether the idea that there are (approximately) true laws. It is just that it takes what she calls a "nomological machine" (Cartwright 1999: Ch. 3) to instantiate a law. The coin falling from St. Stephen's steeple is such a nomological machine. So is the natural system consisting of the planet Jupiter orbiting the Sun, which system instantiates Newton's gravitational law.

Now we can return to the original statements of Newton's laws. Although by themselves they make no empirically testable claims, they do, according to Cartwright, make claims about the world. They describe abstract capacities, something akin to Aristotelian natures. Newton's laws tell us something about the nature of mechanical motions. The Gravitational Law tells us that it is part of the nature of masses that they tend to attract other masses. But from natures alone we cannot predict how any particular system will behave. For that we need more specific interpretive models. We know how to construct such models for falling coins but not for freely falling banknotes.

As Cartwright notes, there are two ways of generalizing about a nomological machine, which I will call "internal" and "external" respectively. Internal generalization concerns repetitions of the same nomological machine. External generalization goes from one instance of a nomological machine to other, relevantly similar, nomological machines. Cartwright argues that both types of generalization require something like capacities because otherwise we do not have a reliable means for determining which features of a repetition or of another nomological machine are the ones we need to control for a successful prediction of the behavior of that new repetition or new nomological machine. The features to focus on are just those that permit the natures to express themselves in the desired manner (Cartwright 1999: 89–90).

This snapshot hardly does justice to the richness of the discussion in *The Dappled World*, even as it applies just to classical mechanics. Hopefully it is faithful enough for what follows.

PRINCIPLES AND MODELS

My own view of theorizing is very similar to Cartwright's, though perhaps my way of thinking about these things is a little more regimented.[4] I would first distinguish between fundamental principles and descriptions of models. Newton's three laws are primary examples of what I call principles, The Principles of Mechanics. Other principles include: Maxwell's Principles of Electrodynamics, The Principles of Thermodynamics, The Principles of Quantum Mechanics, The Principles of Relativity, The Principle of Natural Selection, The Principle of Nash Equilibrium.[5]

I agree with Cartwright that these principles are not to be understood as empirical generalizations. At a minimum, I would regard principles as defining highly abstract entities that I would include in the category of models. This has the result that the principles are indeed true. But they are true in the way definitions are true. They are true of the abstract models. This is not so different from the old Logical Empiricist idea that scientific laws provide implicit definitions of their terms. The difference is that now we focus on highly abstract models rather than axioms, abstract entities rather than linguistic entities, abstract structure rather than linguistic structure.

These models are abstract in two well-known ways. First, they are abstract objects such as numerical relationships or geometrical figures, square roots, perfect squares and circles, or never-constructed buildings described in architect's drawings. They are not physically realized. Second, they are abstract in that they are not fully specified. Newton's laws refer to forces, masses, accelerations, velocities, positions, and times but not to any specific such objects or quantities.

To illustrate these two senses of abstractness, let us stay with Cartwright's example of a body falling in a uniform gravitational field. Setting $F = -mg$ (with the value of g undetermined) in the Second Law yields the prediction that, when released from height h, a suitable body will hit the ground in t seconds according to the relationship $h = 1/2gt^2$. Such an object is abstract both in the sense that it is an idealized object and also not fully specified. Neither g nor either h or t have been assigned specific values, nor has the body to fall been identified. It is just an unspecified body. If we now specify values for, say, g and h, we still have an idealized abstract object, but now it is fully specified in the sense that all the mathematical variables have assigned specific values. It remains abstract in the first sense noted above but is no longer abstract in the second sense.

Once we have set F equal to $-mg$, all the models in the resulting little hierarchy of more and more specific models seem to me examples of what Cartwright calls "interpretive models". We would get another hierarchy if we set F equal to $-kx$, and so on. The various equations, $F = -mg$, $F = -kx$, etc. are what she calls "bridge principles". I think she would say they provide a bridge between the theory and models of the theory. I would say, rather, that these various forms of the force function make it possible to define more specific models than those defined by the fundamental principles, models that can be made fully specific in the sense just noted. Moreover, as I will discuss further a bit later, they are models that can be used to represent actual systems in experimental situations, unlike the highly abstract models defined by the fundamental principles.

Cartwright rejects both the Logical Empiricist account of theories as sets of axioms and the later "semantic" account of theories as sets of models, but she never says explicitly what she herself means by the term "theory". My best guess at a reconstruction of what she does say is that, for her, a theory is to be identified with a set of fundamental principles *plus* a set of bridge principles.[6] Together these statements may be used to define the various little hierarchies of models (her "interpretive models") that, on a "semantic" approach, could be identified as constituting "the theory".[7]

Late in *The Dappled World* Cartwright discusses another category of models she calls "representative models" (Cartwright 1999: 180–198), but I think are better labeled "representational models". They are used to represent things in the real world. What is the difference between "interpretive" and "representational" models? I think it is this: Cartwright is much impressed by and concerned with what most people think of as applied physics or engineering. The Stanford Gravity Probe and SQUIDS (Superconducting Quantum Interference Devices) are among her favorite examples. Our models of such devices draw on diverse sets of principles, including principles not associated with any recognized fundamental theory. Moreover, in such cases we construct the devices to fit our models as much as we construct models to fit the devices. The models serve, she says, as blueprints for constructing these nomological machines as well as serving to represent the devices. Her "interpretive" models, by contrast, are generated from a single set of fundamental principles associated with a single fundamental theory. They result from the various "interpretations" of the abstract terms in the principles provided by the bridge principles.

Now the distinction between what Cartwright sometimes calls "models of the theory" and models constructed using a variety of principles is a real distinction. And her emphasis on the more applied aspects of physics (and science in general) provides a healthy antidote to so-called "foundational studies" which focus almost exclusively on the principles of individual theories to the exclusion of actual scientific (including theoretical) practices. Nevertheless, I would insist that both her "interpretive" and "representative" models are *representational*, and in the same way.[8] But what way is that?

USING MODELS TO REPRESENT REALITY[9]

It is tempting to think that there is a binary representational relationship between a model and a system in the real world that it represents. I agree with Suárez (Suárez 2003), however, that, whether based, for example, on either similarity or isomorphism, no such binary relationship exists. We need, rather, to introduce agents who consciously use models to represent things. And once we have agents, we must consider the purposes for which they are doing the representing. I need not argue this point here since it seems completely within the spirit of Cartwright's approach (but see Giere 2004).

Consider, then, a situation more even regimented than dropping a coin from St. Stephen's. Imagine a steel ball suspended by an electromagnet in a laboratory so it can be released by turning off the current in such a way as simultaneously to start a clock. On the floor is a switch arrangement that stops the clock when the ball lands. If the steel ball is suspended 10 meters above the surface of the earth, we can then construct a fully specific gravitational model of this situation. All that remains to be done is to identify the particular real ball in the laboratory as the object to be represented by the body described in the model. So we represent the real ball by the body in the model, the real height by h = 10m in the model, etc. Within the model we can calculate that, when released, the body will hit the ground in t = 1.428571 . . . seconds, assuming a uniform acceleration, g, of exactly 9.8 m/sec^2. I want to say that this prediction still describes an idealized abstract object and is exactly true of this abstract object. But now that we have established a correspondence between the objects and quantities in the model and those in the laboratory, we can transfer the prediction to the real ball, concluding that it will hit the ground in about 1.43 seconds.

What makes it possible for a model to be used to represent something? One thing, and I would not claim the only thing, is similarity in relevant respects and degrees. So here I (like Cartwright, 1999: 193) do invoke the notion of similarity (or resemblance). One cannot, however, define an *objective* similarity relationship between an abstract model and any physical objects, one independent of human intent. In fact, I don't think anything is in this way objectively similar to anything else except in the vacuous sense that everything is similar to anything else in at least a countable infinity of ways. But the activity of representing the world as characterized above does not require such an absolutely objective notion of similarity. The way scientists use a model to represent some real system is by picking out some specific features of the model which are then claimed to be similar in specified ways to features of the real system to some, perhaps fairly loosely indicated, degree of fit. It is the relatively objective existence of the specified similarities that makes possible the use of the model to represent the real system.

I insist that throughout the theoretical process we continue to distinguish between the model (an abstract object) and the real objects the model is used to represent. In particular, we should *not* think of the real objects as themselves constituting models. This is a standard way of thinking about models in logic where real objects may constitute a model of a formal set of axioms by instantiating the formal relationships.[10] I think a sufficient reason for insisting on the separation of real objects and representational models is that real objects cannot be expected to satisfy exact formal relationships. The gravitational constant on earth, for example, varies between the poles and the equator, so is hardly anywhere equal to exactly 9.8 m/sec^2.

I have appropriated the term "hypothesis" for claims that there is a good fit between a fully specific model and a concrete object or system. Thus hypotheses, unlike models themselves, are statements that may be true or false depending on whether the indicated good fit is realized or not. I continue to think that to understand the role of hypotheses we do not need a substantive theory of truth. A minimalist, redundancy account will do. To say that a hypothesis of this form is true is to say no more or less than that there is indeed a good fit between the designated aspects of the model and the indicated real system. Given an understanding of how good a good fit should be in the circumstances, that claim is in principle subject to empirical investigation.

CAPACITIES AND CAUSES

Cartwright embraces a distinction between what is "abstract" versus what is "concrete" that seems different from either of the two senses of "abstract" I introduced earlier. She says that the force referred to in Newton's Second Law is abstract while gravitational and linear restoring forces are concrete. Perhaps her main reason for this claim is that there is nothing physical that, for example, a gravitational force and a restoring force of a spring, have in common except being forces that can be manipulated so that they satisfy Newton's principles of motion. They are concrete realizations of the abstract forces referred to in the principles of motion.

In addition, for Cartwright, both Newton's three fundamental principles of motion and various bridge principles, such as the principle of universal gravitation, do more than merely define abstract models. They describe real capacities in nature. The gravitational law tells us that bodies have the capacity to attract other bodies. The bridge principle, $F = -kx$ tells us that springs have the capacity to resist displacement from an equilibrium position. It is in the nature of bodies and springs to do these things. But from these natures we cannot directly infer anything about observable motions. Only when combined with the principles of motion do explicit references to forces drop out, and we are left only with references to measurable quantities such as mass, velocity, and time. We then have the possibility of constructing models of actual situations. But only the possibility: for only after we have eliminated interfering factors can we expect to have a nomological machine for which the remaining equations define a model.

Consider again the laboratory set-up with the suspended steel ball. Now imagine an electric coil surrounding the path of its fall, such that the steel ball passing through the coil creates a magnetic field which impedes, though does not stop, its fall. The net result is that the time it takes to reach the floor is measurably less than the simple model of free fall predicts. I choose this example because there are no models constructible on Newtonian principles alone that can account for the actual time to reach the ground. For that we need principles of electromagnetism. This supports Cartwright's contention that one cannot be confident of constructing reliable models if one is confined to a single set of theoretical principles.

In spite of my considerable sympathy with Cartwright's program, I still find her invocation of capacities and natures to be anachronistic, even quixotic. I still feel there was something profoundly correct about the rejection of such notions that was part of the scientific revolution of the seventeenth century.[11] Which is not to say that I support the empiricist account of laws she is at such pains to reject. But rejection of the arid metaphysics of both classical and contemporary empiricism does not necessarily drive one back to Aristotelian natures, as Cartwright herself realizes. She several times notes that what she thinks we need is either capacities 'or some related non-Humean notion' (Cartwright 1999: 89). She even highlights Max Weber's notion of "objective possibility" as 'very much worth pursuing in our contemporary attempts to understand scientific knowledge' (Cartwright 1999: 72). I take this as a hint that we can do as well with a robust notion of causality more in tune with the modern scientific tradition as we can with Aristotelian natures.

I earlier endorsed the minimal view of the principles of mechanics as defining abstract models. As an alternative to Cartwright's view of principles as describing capacities, I would now like to suggest that we can instead take them to be describing an abstract *causal structure*, or, perhaps better, as abstractly describing a causal structure. By making these models more specific in just the ways described above we can reach general models of causal systems such as Cartwright's nomological machines, arriving finally at models of specific causal systems.

It seems to me that the virtues Cartwright finds in capacities are equally well to be found in causal structures. Invoking causal structure, for example, provides as much (or as little) support for both internal and external generalizations. Indeed, the fact that one can talk about "causal powers" suggests there may not be much real difference between invoking causal structures and invoking capacities. In the end, one may be able to do all the work that needs doing with nothing more than a robust notion of causality and thus avoid introducing capacities in addition to causes.[12]

Cartwright argues that capacities are implicit in our scientific practice and thus not objectionably metaphysical. I would argue that the practice she cites only goes so far as implicating causal structures, not capacities. Consider again my laboratory version of the coin falling from St. Stephens. By varying the height from which the ball is released, deliberately or at random, we can effectively sample the causal possibilities in that system. The fact that the time of fall continues to agree with the predictions from the respective models is evidence for the causal structure reflecting the structure of the models. The introduction of human agency here is important. A stricter empiricist would claim that our evidence consists only in the set of observed ordered pairs of the form (h, t). The fact that we are free to choose whatever h we wish (within the limits of the apparatus) is regarded as irrelevant. I think this little bit of human agency makes all the difference, as I am confident Cartwright would agree. We both reject the spectator view of scientific observation in favor of a view incorporating a role for active human intervention.

METAPHYSICS AND METHODOLOGY

Let us return, finally, to what Cartwright calls (scientific) "fundamentalism", the view that scientific laws rule everywhere and always. I agree with Cartwright that this sort of fundamentalism is a metaphysical extrapolation from actual scientific practice. Our practice finds the world agreeing with the laws describing our models mainly in limited and highly contrived circumstances. For most of our everyday experience, no scientific models have ever been devised. She urges us to adopt the contrary metaphysics of a dappled world. I would insist that science has no need of any metaphysics whatsoever. Sound methodology is enough.

Staying with classical mechanics, we can agree that the gravitational inverse square law provides well-fitting models for systems both below and above the sphere of the moon. But what about the stars? Or, in the twentieth century, other galaxies? Fundamentalists would insist that this law must apply to these other regions as well. But a fundamentalist belief in the truth of this extrapolation is not necessary for scientific progress. It is sufficient that scientists adopt the methodological rule of first trying models that have worked in situations judged similar to the new one. That provides at least some evidence that they *might* work in the extended situation. So trying gravitational models is well justified. But thinking that these models *must* fit is not justified. Their fit had to be certified by more direct evidence from the motions of the stars, galaxies, or whatever. So a methodological shadow of fundamentalism seems in line with sound scientific practice.[13]

What about a methodological version of a dappled metaphysics? It seems to me that a dappled world with capacities gives sufficient reason to expect that familiar models might work in an apparently similar situation without sanctioning belief that they must work. But the same goes for a dappled world with a causal structure but without capacities. So, at the level of methodology, there is little to choose between a fundamentalist and a dappled methodology.[14] This is as it should be. Sound scientific methodology should be independent of metaphysics.

There may, however, be broader reasons for advocating a dappled over a fundamentalist methodology. Human psychology being what it is, scientists may be all too prone to inflate sound methodological guidelines into extravagant metaphysics. This leads to the kind of scientific hubris Cartwright opposes. Keeping in mind the mere *possibility* of a dappled metaphysics may promote a desirable degree of scientific modesty.

CONCLUSION

Among contemporary philosophers of science, there are none that I admire more, or whose views seem to me closer to my own, than Nancy Cartwright. I fear the above does not do justice to my admiration for her work and her courage in advancing views contrary to the ideology of many contemporary philosophers of science and scientists as well. Where I am critical, it is in the interest of making our overall shared project simpler and more accessible both to our colleagues in the philosophy of science and in science studies more generally and also to reflective scientists.

NOTES

1. The support of the Netherlands Institute for Advanced Study in the Humanities and Social Sciences for the academic year 2002–2003 is hereby gratefully acknowledged.
2. I first met Nancy at Fred Suppe's conference on The Structure of Scientific Theories in Urbana, Illinois, in 1968, when she was still a graduate student at the old Chicago Circle Campus of the University of Illinois and I a beginning Assistant Professor of History and Philosophy of Science at Indiana University. For the past thirty plus years our views have developed in parallel but with only occasional interactions. More important have been common influences on our work by people like the late Wes Salmon, Pat Suppes, Ian Hacking,

Arthur Fine, and Bas van Fraassen. I am grateful to the organizers of the Konstanz workshop for providing the motivation for me at last to consider systematically some of the similarities and differences in our views.

3. These two terms are echoes of terms developed within earlier accounts of the nature of scientific theories. "Bridge principles" or, more exactly, "bridge laws" were what the logical empiricists invoked to link terms in theories when one theory was to be reduced to another (Nagel, 1961). In the later set-theoretical understanding of scientific theories (Suppes 1960), "interpretive models" are sets of objects that instantiate purely formal statements in such a way as to make the now interpreted statements true. Cartwright's use of the terms seems somewhat different from either of these. I will use her term "interpretive model" until I introduce my own terminology later in this chapter.
4. I say "theorizing" rather than "theories" because, like Cartwright, I now like to emphasize the process of doing of science, a human activity, rather than just the more static products of scientific work. Even so, one cannot avoid thinking to some extent about more static structures (Giere 1988: 1999a).
5. I just give examples because I do not think one can find useful necessary and sufficient conditions for what counts as a fundamental principle.
6. This is similar to the Logical Empiricist idea that a theory is a set of axioms plus a set of correspondence rules linking theoretical and observational terms, though Cartwright does not employ a traditional theory/observation distinction.
7. I myself would not now identify a theory with a set of models. Indeed, I think the term "theory" is used so ambiguously in general scientific and philosophical discourse that a sound understanding of science is better achieved by not trying to adopt "theory" as a technical notion. What people want to say using this term in various contexts can better be said using other, more specialized, terms.
8. Margaret Morrison (this volume) also says that interpretive models are used representationally.
9. A number of students of the scientific enterprise, including Cartwright and Margaret Morrison (1999), have insisted that scientists use models for all sorts of purposes other than representing the world. In focusing on representing, I do not mean to deny this. I do think, however, that representing the world is a very important function of models and is often presupposed in discussions of other roles for models. So a focus on the representational role of models is well justified.
10. It was on this conception of models that Pat Suppes based his famous claim that the concept of models is the same in mathematics and the physical sciences (Suppes, 1960). It took me a long time to realize that my understanding of the use of models in science was not the same as Suppes'. My representational models are different than his "instantial" or "interpretational" models (Giere 1999b).
11. Cartwright was in part inspired to invoke these notions by Elizabeth Anscombe, to whom Chapter 5 of *The Dappled World* is dedicated. I fear that my basic sympathy with the Enlightenment makes me suspicious of Anscombe's Catholicism and its attendant metaphysics.
12. The historical transition from classical to general relativistic accounts of gravitation provides a nice example. In the context of classical mechanics, Cartwright says that masses have the capacity to attract other masses. In the context of general relativity, she would presumably say that masses have the capacity to distort the space-time around them so that other bodies naturally move on geodesics in the distorted space-time (and thus only appear to be attracted). So we have changed our account of the capacities masses have. But

here it seems to me that talk about capacities is clearly redundant. It seems sufficient to say that we have changed our understanding of the causal nature of interactions among masses from action at a distance to action mediated by the induced structure of space-time.

13. Indeed, recent observations suggest that the rate of expansion of the universe is *increasing*, thus indicating that there are forces other than gravitation at work.
14. This conclusion is in agreement with Paul Teller's view (this volume) that the proper attitude toward fundamentalism is agnosticism.

REFERENCES

Cartwright, N. (1999) *The Dappled World: A Study of the Boundaries of Science*, Cambridge: Cambridge University Press.

Giere, R. N. (2004) 'How models are used to represent reality', *Philosophy of Science*, 71: 5.

———. (1999a) *Science without Laws*, Chicago: University of Chicago Press.

———. (1999b) 'Using models to represent reality', in L. Magnani, et al. (eds) *Model-Based Reasoning in Scientific Discovery*, New York: Kluwer/Plenum.

———. (1988) *Explaining Science: A Cognitive Approach*, Chicago: University of Chicago Press.

Morrison, M. (1999) 'Models as autonomous agents', in M. S. Morgan and M. Morrison (eds) *Models as Mediators: Perspectives on Natural and Social Science*, Cambridge: Cambridge University Press.

Nagel, E. (1961) *The Structure of Science*, New York: Harcourt, Brace, & World.

Suárez, M. (2003) 'Scientific representation: Against similarity and isomorphism', *International Studies in the Philosophy of Science*, 17:225–244.

Suppes, P. (1969) 'A comparison of the meaning and uses of models in mathematics and the empirical sciences', in P. Suppes (ed.) *Studies in the Methodology and Foundations of Science*, Dordrecht: D. Reidel.

Reply to Ronald N. Giere

I thank Ronald Giere for his admiration, which is reciprocated, and I agree with him that his views and mine have a great deal in common. I suspect this comes from a lot of conversations in our youth and a shared kind of disaffection with all the attention to abstract theory and the system of physics. I should like here to discuss these points that Giere raises.

First is on *specification*. Within Giere's single category, I distinguish two different kinds of specification. Theories contain principles with variables in them representing putative features for the world, such as *f* for force, *a* for acceleration, Φ for the quantum state, or *H*, the quantum Hamiltonian, for the possible energy states of a system. To make predictions about a given situation, we need to fill in, or specify, values for these variables.

I maintain that for some quantities this must be a two-step process if the treatment is to be principled as opposed to ad hoc. That is because some concepts in physics are abstract in a very particular sense—they piggyback on other more concrete descriptions. Force is like this, as is the quantum Hamiltonian, but acceleration and the quantum state are not. There are rules for how to fill in the force variable with a more specific force function. These rules are given in the bridge principles of the theory, which link specific forms of the force function with what I call "interpretative models". The ascription to a given situation of a particular form of the force function will be ad hoc unless the situation also satisfies the description in the associated interpretative model. For instance, the specific form "–*kx*" can be used only when the system can be described as a harmonic oscillator. Of course after that, in order to make specific predictions, it is still necessary to specify a particular value for the force, such as "10 dynes".

I stress this two-tiered process because it shows clear bounds to the scope of theories that use abstract concepts of this kind. For instance, *f* = *ma* can only apply to situations that are appropriately described by some combination of the interpretive models supplied by the bridge principles of Newtonian mechanics. For most of Giere's purposes, however, the two varieties of specification can be collapsed into one.

Second is on *similarity*. Giere is often criticized for insisting that models must be similar to the systems modelled, though only in designated ways. I think these criticisms are not sufficiently generous in this interpretation of "similarity". Let us look at it as I do first, in terms of the *claims* one is allowed to derive from the model. As I describe in my comments on Daniela Bailer-Jones in this volume, I follow Mary Morgan in supposing that models need stories to tell us what to do with them and how to draw conclusions about the target from what we do.[1] I suppose, for example, that we have a hydraulic model of an economic process. We are to experiment with the model by pouring more water in vat one and looking to see if pressure in vat four goes down. We understand from the story that if it does we are to interpret this as telling us that if we lower taxes spending will go up. Where is the similarity? But of course the requisite similarity is just that: The model and the target are similar in the *relevant* respect, namely pressure in vat four goes down as water is poured into vat one just in case spending goes up as taxes go down. This does not make "similarity" an empty concept. It just shows the importance for use of making clear what the intended similarity is supposed to be; and Giere is explicit that that job must be done by the hypothesis, just as for Morgan it is done by the story.

Third is the question of why I endorse the notion of *capacities*, which is a stronger modal notion than *causal structure*. I take it that a causal structure is a specific arrangement of features of the world—causes—that act together to produce different effects. We are supposed to imagine experimenting on the various causes in the structure to see how a given variation in a particular cause affects the effect.

What do capacities do for us beyond that? They articulate what the given cause contributes across all possible causal structures,[2] where this will in general be different from the effect produced in any one causal structure by varying the cause. An electron, it seems, always *repels* another electron; it "tends" to cause the second electron to move away. This is true despite the fact that in some causal structures moving the first electron towards the second will cause the second to move even closer; in others it will cause a particular motion; in others no motion at all. The actual effect depends on the set-up. Yet we know how to calculate that effect from the "law" of electron-electron repulsion.

But what does this "law" say? It cannot be formulated as a claim about regularities among occurrent properties, nor about what electrons always cause, nor about what they cause in a given causal structure, nor in every causal structure. What then? I suggest that it says that electrons—because they are electrons—have the capacity of the given strength to repel other electrons, where for nice situations we have some rules for how to calculate the results that occur when this capacity operates jointly with others, and where in messier situations we are entitled at least to claim "the electron might cause a second to move away". What better alternatives are available?

NOTES

1. In general much of the story will be implicit, relying on often unarticulated convention in the community of users.
2. Or, across all causal structures in which the given cause retains that particular capacity.

7 Experimental Realism Reconsidered

How Inference to the Most Likely Cause Might Be Sound[1]

Mauricio Suárez

INTRODUCTION

In her first book, *How the Laws of Physics Lie*, Nancy Cartwright provided a wide-ranging critique of a realist attitude to any explanatory scientific theory. Cartwright argued, roughly, that the explanatory power of a theory is at odds with its descriptive accuracy. The greater the "covering power" of a theory the more idealised and further from the truth the theory will be. Cartwright promoted instead the position known as "entity realism" or "experimental realism", according to which it may be possible to have a justified belief in the existence of some unobservable entity postulated by science, independently of any justification for our current best theory about that entity. Experimental realism thus achieves a combination of common sense realism about some unobservable entities with a principled nonrealism about theories.

By contrast Cartwright's last book *The Dappled World*, is perfectly happy to accept a robust form of scientific realism about theories. As she puts it herself: 'Nowadays I think I was deluded about the enemy; it is not realism but fundamentalism that we need to combat' (Cartwright 1999: 23). Fundamentalism entails that there is only one true set of laws about the world; antifundamentalism on the other hand allows a large number of scientific theories, postulating alternative sets of laws, to be true at once, each of them in their particular domain. Cartwright's current view is "anomalous dappling", according to which different laws govern different patches of the world, but no law may govern some patches at all (Cartwright 2002).[2] "Anomalous dappling" allows us in principle to take a full-blown realist attitude to many more than just one empirically adequate theory, as long as they don't contradict each other, thus yielding the promiscuous or patchwork realism that is in accordance with the metaphysics of the disunified world.[3]

Cartwright's quote suggests that she might be happy to accept that if one remains unconvinced about realism about theories, one would not be particularly inclined to defend either antifundamentalism or fundamentalism, since both presuppose realism. The fundamentalist claims that all regions of the world are law-governed, and moreover by the same laws; the antifundamentalist, such as Cartwright's "anomalous dappler", claims that only some regions are law-governed, and not necessarily by the same laws. Both presuppose that science aims, through its laws, to represent the way the world really is. So their dispute about whether there is one law that subsumes all phenomena is also a substantive ontological dispute about what the world is really like.

By contrast, an antirealist, or nonrealist, will find that there is no substantive ontological issue at stake. He or she will find no offence in the search for the system that best organises and economises our thought or even in supposing that there is one system that does it best. For the existence of such system does not show anything about what the world is really like but only, at best, about how we conceptualise it. To put it in a nutshell, only from the perspective of realism about theories can fundamentalism be the "enemy"; and only from that perspective can anomalous dappling be defended. From the perspective of nonrealism, both fundamentalism and anomalous dappling are metaphysical views underdetermined and not required by the practice of science or by an abductive inference to the best explanation of that practice. From this perspective, it is just as mistaken to draw metaphysical lessons from scientists' failure to find a unified system of laws that fits all phenomena as it is to draw them from their success in finding it. Fundamentalism and anomalous dappling appear to be equally unwarranted and unnecessary.[4]

This suggests that the choice that we are presented with is not really exhaustive or, more precisely, that we need not share its presuppositions. I was surprised to find Cartwright essentially conceding this point at the Konstanz conference, where she presented anomalous dappling not so much as a metaphysics of its own but as an attempt to break away from the dominance of fundamentalist metaphysics. It then seems to me that the anomalous dappler and her fundamentalist opponent share the mistaken assumption that theoretical physics and philosophy of physics have so far been totally dominated by fundamentalist metaphysics (see for instance Hoefer 2003). At least the work of those physicists and philosophers of physics that I am most familiar with (e.g., Bohr, Schrödinger; Hempel, Reichenbach, Fine, Van Fraassen) presupposes neither fundamentalism nor anomalous dappling. I don't see, for instance, how one attacks, or defends, philosophy of physics as a discipline by attacking, or defending, fundamentalism.

Cartwright would probably reply that the antirealist is still "deluded about the enemy". She has advanced several arguments against fundamentalism and in favour of anomalous dappling. These are detailed and intricate arguments, which require careful and thorough philosophical analysis. I will not attempt to evaluate them here. My aim in this chapter is the more modest one of discounting one possible motivation for Cartwright's present belief that she was "deluded about the enemy". One way to motivate the dilemma between anomalous dappling and fundamentalism is of a negative sort: These are the only alternatives to understand science; no other alternatives will work.

More specifically, one reason that may lead (and which may have led Cartwright) to the theoretical realism that underpins the dilemma of *The Dappled World* is dissatisfaction with the experimental realism of *How the Laws of Physics Lie*. There have been many papers directly criticising experimental realism since 1983. The objections fall into three different kinds, depending on whether they charge experimental realism with (i) inadequacy, (ii) incoherence, or (iii) implausibility. Did Cartwright abandon experimental realism because these critics convinced her? Is that the reason why she changed the focus of her criticism? I would like to argue that there is no sound argument, even of such a negative kind, in favour of Dappled World metaphysics. There was no need to abandon the antirealism about theories of *How the Laws of Physics Lie* in the first place: The arguments against "experimental realism" are inconclusive, and a version of the position is defensible.[5]

The experimental "realism" that is rendered plausible by the arguments in this paper is distinct from what Cartwright intended in the first place, and seems to me to be incompatible with her recent *dappled world* metaphysics. In fact, I do not think of it as a realism at all. A more appropriate name for my views would be "the experimental attitude", as among all contemporary epistemological views it is closer to the "neither realism-nor-antirealism" of Arthur Fine's NOA. In the remainder of this paper I try to sketch out what this position would amount to if forced to express it as an epistemic thesis—as I believe Cartwright should have done in 1983.

EXPERIMENTAL REALISM: COMMON-SENSE EPISTEMOLOGY OR FANCY METAPHYSICS?

It is widely believed that, when first introduced by Hacking and Cartwright, experimental realism was conceived primarily as a metaphysical thesis about what kind of entities are, or can be, real. Hacking's manipulability criterion, in particular, is often taken to express a condition on what should count as real. For example, Margaret Morrison writes:

> Hacking contrasts the metaphysical questions concerning scientific realism with those that deal with rationality, the epistemological questions. The former raise issues such as, Are those entities postulated by physics theories real?, What is true of those entities?, What is truth?, and so on . . . In arguing for entity realism Hacking takes himself to be addressing only the metaphysical questions. (Morrison 1990: 1)

I shall argue that Morrison's description of Hacking's position as *only* metaphysics cannot be correct. Certainly Hacking's slogan 'if you can spray them then they are real' lends itself to this interpretation, and indeed Hacking has explicitly defended a metaphysical claim on behalf of experimental realism:

> Reality has to do with causation and our notions of reality are formed from our ability to change the world . . . We shall count as real what we can use to intervene in the world to affect something else, or what the world can use to affect us. (Hacking 1983: 146)

However, in addition to this metaphysical claim about what should be counted as real, Hacking also makes an explicit epistemic claim:

> The *best kinds of evidence* for the reality of a postulated or inferred entity is that we can begin to measure or otherwise understand its causal powers. The *best evidence*, in turn, that we can have this kind of understanding is that we can set out, from scratch, to build machines that will work fairly reliably, taking advantage of this causal nexus. (Hacking 1984: 170, my emphasis)

The metaphysical claim then aims to establish a conceptual link between reality and manipulation. The epistemic claim asserts that manipulation provides particularly robust warrant for our ontological commitments. Which one is primary? Suppose that the metaphysical claim was primary. This might be enough to substantiate Morrison's description of Hacking's experimental realism as a piece of metaphysics. Hacking sometimes writes as if the epistemic claim is required only to bring out the practical consequences of the metaphysical one. Or as he puts it, the metaphysics would be "idle" without the epistemology (Hacking 1983: 28); it is because of the tight conceptual link between reality and manipulation that our best evidence for an entity's existence is our manipulating it.

However, Hacking accepts that manipulation is not a necessary condition on reality. There could be real entities out there that we would never be able to manipulate: black holes and gravitational lenses are possible examples. Our failing to manipulate them does not necessarily mean that these entities are unreal; it simply precludes us from having grounds to justify their existence (Hacking 1989). Manipulation is then meant, if anything, as a sufficient condition on reality. Hacking is not defending the conceptual equivalence of what is real and what can be manipulated, but rather, it seems, that manipulation is one important hallmark of reality. We may enunciate Hacking's metaphysical claim as follows:

Metaphysical Experimental Realism (MER)

Manipulation is a sufficient condition on reality: *x* is real if *x* can be manipulated.

A fair amount of criticism has been devoted to this metaphysical claim. In particular the critics have argued that if manipulation is a hallmark of reality, it is hard to see how we could classify or describe some type of entity as "real" independently of the theory that describes its causal powers and our possible manipulations of them.[6] Would it not be incoherent to classify entities as "real" that we have no theoretical description of?

I think that there is something right about this incoherence objection as applied to (MER), and I will return to it in due course; but I'll argue that it can only be an objection to the metaphysical version of experimental realism. I noted that Hacking also makes an epistemic claim on behalf of experimental realism, and I want to suggest that experimental realism must be understood as primarily making this epistemic claim:

Epistemic Experimental Realism (EER)

Manipulation is a necessary and sufficient condition on causal warrant: Our belief that *x* exists acquires this special kind of warrant if and only if we believe that we manipulate *x*.

EER is consistent with many passages in both Hacking's and Cartwright's original papers. I already quoted a passage from Hacking's *Representing and Intervening* to this effect. Let me quote a couple from Cartwright's *How the Laws of Physics Lie*, where I have also emphasised the phrases with undeniable epistemic content:

> Causal reasoning provides good grounds for our beliefs in theoretical entities. (Cartwright 1983: 6)
>
> I agree with Hacking that when we can manipulate our theoretical entities in fine and detailed ways to intervene in other processes, then we have *the best evidence possible* for our claims about what they can and cannot do; and theoretical entities that have been *warranted* by well-tested causal claims like that are *seldom* discarded in the progress of science. (Cartwright 1983: 98)

The view that I want to defend manipulation is indeed taken as an indication, or symptom of reality, but not a certain one; for it is not part of the notion of warranted belief that warrant be infallible and the corresponding belief always true. Hence our taking ourselves to manipulate *x* cannot be, on this view, a sufficient condition on *x*'s reality. MER does not follow from EER.[7]

This chapter constitutes a first step in an argument to the effect that experimental realism needs to make no metaphysical commitments at all and is in particular not committed to MER.[8] In other words, I want to turn the presumed primacy of MER on its head, in order to defend experimental realism as *only* epistemology. Our belief in the existence of *x* acquires a special sort of warrant when we come to convince ourselves that we manipulate *x*; and it is precisely this fact about our epistemic practice that grounds the secondary claim that manipulation is a good indicator of reality; a good guide—not an infallible one.

Experimental realism then needs to establish that EER is true by (i) elucidating the notion of causal warrant and (ii) showing that manipulation affords it. We find some clues for (i) and a partial but essentially sound defence of (ii) in Cartwright's arguments in favour of inference to the most likely cause.

INFERENCE TO THE MOST LIKELY CAUSE (IMLC)

In Chapter 4 of *How the Laws of Physics Lie*, Cartwright argues that inference to the most likely cause (IMLC) is a success term: a putative causal explanation of a phenomenon is only a genuine explanation if the cause is real.[9] By contrast, inference to the best theoretical explanation, or explanation by subsumption under a theory is, according to Cartwright, not a success term: A theory may provide a good explanation of some phenomenon, regardless of its truth-value. So our acceptance of a theoretical explanation of a phenomenon *qua* explanation is not in itself a reason to believe in the explanation; but our acceptance of a causal explanation is a reason to believe in the existence of the cause or causes cited. As an illustration, Cartwright gives the following everyday example:

> My newly planted lemon tree is sick, the leaves yellowing and dropping off. I finally explain this by saying that water has accumulated in the base of the planter: the water is the cause of the disease. I drill a hole in the base of the oak barrel where the lemon tree lives, and foul water flows out. That was the cause. Before I had drilled the hole, I could still give the explanation and to give that explanation was to present the supposed cause, the water. There must be such water for the explanation to be correct. An explanation of an effect by a cause has an existential component, not just an optional extra ingredient.
>
> (Cartwright 1983: 91)

Cartwright takes the reality of the cause to be an intrinsic characteristic of the causal explanation: pointing to a nonexistent cause cannot explain anything. A cause can only constitute a genuine explanation if it actually exists. By contrast, Cartwright follows Duhem and Van Fraassen in arguing that theoretical explanation is not a success term. Providing a satisfactory explanation of a phenomenon by subsuming it under a theory gives no reason to believe that the theory is true. The theory can be explanatory without being true. This is because theoretical explanation does not in general meet the requirement of nonredundancy, which is met in the case of causal explanation:[10]

> We can infer the truth of an explanation only if there are no alternatives that account in an equally satisfactory way for the phenomena. In physics nowadays . . . there is redundancy of theoretical treatment, but not of causal account. (Cartwright 1983: 76)

In addition, Cartwright distinguishes between two different senses of "theoretical explanation". First, there is the explanation of a phenomenon by showing that the law that describes the phenomenon is a special case of a theoretical law. Hempel's Deductive Nomological account of explanation is a case in hand. Inference to the best D-N explanation, admits Cartwright, gives *some* reason to believe in the theoretical law. This reason is not conclusive, though, since the phenomenological law could have been derived from an alternative theoretical law. We are never in a position to rule out such alternative law, hence the requirement of nonredundancy is not generally met. For instance, suppose that we are able to deductively derive the exact future positions of most planets from the laws of Newtonian mechanics plus facts about the present positions and forces. This provides some explanation of the motions of the planets in the solar system; and the success of this explanation in turn provides some evidence in favour of Newton's theory; but it does not conclusively show that Newtonian mechanics is true, since the motions might also be derived from another set of laws (as indeed is the case with Einstein's theory).

Cartwright distinguishes D-N explanations carefully from the type of theoretical explanation that Duhem and Van Fraassen discuss. A theory according to Duhem is an abstract system whose aim is to summarize and classify logically a group of experimental laws. The requirement of nonredundancy fails maximally here: We can be certain that alternative summaries and classifications of experimental laws always exist; ours is just one that we find convenient for our own purposes. Hence there is no reason at all, conclusive or otherwise, to expect a theory, in this Duhemian sense, to be true.

What underlies this discussion is a difference between two types of inference to the best theoretical explanation (IBTE): those IBTEs that are completely unwarranted (inferences to the best Duhemian theoretical explanation) and those that are warranted but only mildly, on the supposition that no alternative superior explanations are available (inferences to the best Hempelian theoretical explanation). The former type of IBTE transmit no warrant to their conclusions; while the latter transmit what we may call theoretical warrant. By contrast, according to Cartwright, the requirement of nonredundancy is always met in the case of causal explanation. This is the explanation of a phenomenon by direct appeal to its cause. Since this type of explanation is only successful to the extent that the cause is real, inference to the most likely cause (IMLC) is strongly warranted. I will refer to this type of warrant, which accrues from successful causal explanation, as causal warrant.

We have thereby established a distinction between three types of inference and the corresponding type of warrant that they are able to transmit to their conclusions. An inference to the best Duhemian theoretical explanation of a phenomenon is fully unwarranted, one to the best Hempelian theoretical explanation is theoretically warranted, while an inference to the most likely cause of a phenomenon is causally warranted. What is left to establish is the sense in which causal warrant is stronger than theoretical warrant.

Cartwright claims that causal explanations obey the requirement of non-redundancy because experimental testing and manipulation of the cause under controlled laboratory conditions allows us to establish the most likely cause of a phenomenon. Only then can we say that we have provided a causal explanation for it. Hence success in causally explaining a phenomenon by citing some entity *x* and its causal properties gives us the most conclusive reason that we may ever have to believe in *x*'s existence. Cartwright writes:

> If God tells you that the rotting of the roots is the cause of the yellowing of the leaves, or that the ionisation produced by the negative charge explains the track in the cloud chamber, then you do have reason, conclusive reason, to believe that there is water in the tub and that there is an electron in the chamber. (Cartwright 1983: 93)

It is tempting to interpret "conclusive reason" in this passage as "infallible reason", in a sense that would defeat a Cartesian sceptic: The existence of causal explanations would entail the existence of real entities and their causal properties; and this in turn would entail the existence of an external world. IMLC would then become the best proof of metaphysical realism. This interpretation is surely mistaken, however, even if it is unfortunately suggested by Cartwright's above appeal to divine revelation and by the contrast she draws between inference to the best (theoretical) explanation and the Cartesian *cogito ergo sum* argument (Cartwright 1983: 89). But it would be a mistake to suggest that IMLC has the global or radical scepticism-defeating character of Descartes' argument in favour of the existence of the external world. For that would make experimental realism susceptible to the dazzling battery of arguments in favour of global scepticism and would thus turn it into a thoroughly untenable position. In addition, it would be attempting to provide much more than is needed to defend a qualified form of selective realism. Experimental realism should not be required to show that an external world exists, but rather—on the assumption that there is such a world—that our beliefs in the unobservable entities of science are no less warranted than our beliefs in the objects of our ordinary experience.

Is there a coherent interpretation of the term "conclusive" that could make Cartwright's statement true? I believe there is. It turns on the fact that the warrant afforded by the inference to the most likely cause of a phenomenon is more robust than the warrant that inference to the best theoretical explanation could ever provide. "Conclusive" is thus to be understood as a relative term: Causal warrant is conclusive in comparison with the warrant provided by an IBTE. No existential commitment derived from an IMLC can be defeated by any amount of theoretical warrant to the contrary. If we believe that we have manipulated an entity, in the way required for an IMLC, in order to causally explain a phenomenon, then no theoretical explanation of that phenomenon, no matter how empirically successful the theory, ought to lead us to withdraw the causal commitment.

Once I come to be convinced that I have manipulated the water in the basis of the lemon tree planter, in such a way as to establish it to my satisfaction as the cause of the yellowing, then I would not give up my belief that water causes the yellowing (and it would be epistemically irresponsible for me to do so) on account of any alternative, purely theoretical or speculative, explanation. The only defeater of causal warrant in favour of the existence of *x* is causal warrant of the same strength against *x*. I would only abandon the belief that there is water that causes the yellowing if I obtained causal warrant of the same strength in favour of a different substance causing the yellowing—forcing me to conclude that I was wrong to believe that I was manipulating water in the first instance.

To sum up my improved version of Cartwright's argument in three compact theses:

(i) Duhemian theoretical explanation is not a success term—in the sense that a false theory *T* may provide a satisfactory explanation of a phenomenon. But causal explanation is a success term—if the cause is not real there is no genuine explanation.

(ii) From the fact that a theory *T* explains a phenomenon *y*, we cannot infer that *T* is true. But from the fact that *x* (probably) causally explains *y* we may infer that *x* is (probably) the real cause of *y*.

(iii) We can accept a theoretical explanation, *qua* explanation, even if we do not believe that the theory is true. But we cannot accept a causal explanation, *qua* explanation, unless we believe that the cause is real.

Many of the objections to experimental realism, which I review in the next section, suppose that (i), (ii), and (iii) are the epistemological consequences of MER. Consequently, the inference to the cause in (ii) would have to be certain, or infallible. For in order to establish that *x* causally explains we need to have manipulated *x*, and if *x* has been manipulated then, by MER, it surely is real. Yet, (i), (ii), and (iii) can be argued for directly, without presupposing MER; and thus without presupposing that an inference to the most probable cause is certain, or infallible.

INADEQUACY, INCOHERENCE, IMPLAUSIBILITY

The objections to experimental realism fall roughly into three types. I will discuss them in order of what I take to be their increasing importance.

Inadequacy

The first objection to experimental realism is that it provides an inadequate picture of the actual aims and particular objectives of scientific research and inquiry. Such an artificially mixed combination of realism about entities and antirealism about theories does not drive any particular scientific inquiry, and it does not accurately describe science, including the most experimental, nontheoretical branches of applied science. Resnik, for instance, writes:

> Experimenters do not operate without genuine scientific theories and laws about the phenomena they investigate: The gulf between experiment and theory is not nearly as large as Hacking supposes. . . . A person running experiments with a particle accelerator may not be aware of the latest developments in theoretical physics, but he (or she) is likely to be familiar with most of the commonly accepted background theories in physics, including some theories about the particles he (or she) is studying.
>
> (Resnik 1994: 410)

Let us suppose Resnik is correct about the intertwining, in practice, of theoretical and practical knowledge. It seems plausible that scientists use both theoretical and experimental knowledge in their work; and of course many of their beliefs are infused by theory. The experimental realist need not deny any of this, though. He or she need only claim that theoretical and experimental knowledge have some distinct cognitive and epistemic functions but not that all their functions are distinct or fully separate. Resnik's objection will draw out slightly different replies depending on whether experimental realism is understood as primarily defending MER or EER, but the core of the reply will be the same in either case: The origin and content of scientists' beliefs and knowledge is irrelevant to experimental realism. MER claims that scientists' ontology is ultimately derived from their phenomenological knowledge and the results of their laboratory manipulations. EER claims that experimental laboratory practice provides the strongest form of warrant for scientist's existential commitments. Each of these views is fully compatible with scientists' possessing a mixture of theoretical and experimental knowledge.

The issue concerning experimental realism is a more fine-grained question about the specific role that these two types of knowledge play in actual scientific practice. For MER the question is what type of knowledge scientists ultimately base their ontologies on; for EER it is what warrants their existential commitments; but both can accept that theoretical and experimental knowledge is always deeply intertwined in practice.

Incoherence

The second objection is that experimental realism is incoherent. It is not possible to coherently separate an unobservable entity from the theory that describes it, because the only concept that we possess of an unobservable scientific entity is the one given to us by the theoretical description that we happen to accept. We may have both a theoretical and a perceptual concept of an observable entity, such as the Konstanz train station; but about an unobservable entity, we can have no perceptual concept, only theoretical.

Hacking is clear that in order to describe our causal interaction with an unobservable entity we need only appeal to a set of phenomenological "home truths" about the entity and its properties; we need not believe the full theoretical description of it. Suppose we infer the reality of an entity *x* on the basis of our interaction with it, as described by some such set of home truths. The thorny question for the experimental realist is this: what is it that we are inferring to, when we infer to the reality of an entity on the grounds that we can manipulate it? Whatever it is, it is not that entity that we take ourselves to be inferring, since the properties that we must ascribe to the entity in order to manipulate it are typically only a subset of the full set of properties that informs our theoretical concept.

For instance, the entities that we must suppose are real because we manipulate them in the electron microscope are not quite electrons as we understand them: They are particles (call them *flectrons*) that have some of the properties of electrons but not all of them. (MER) would then not allow us to claim that electrons are real but only that flectrons are. And in each distinct manipulation of "electrons", each circumscribing our causal interaction to a different subset of their properties, we would *really* be manipulating different entities: "flectrons" this time, "plectrons" the next, and so on. Yet, this is deeply counterintuitive, if not plainly wrong. Scientists do not take themselves to be confronting different particles when they carry out a scattering experiment on electrons as opposed to operate an electron microscope. In both cases they take themselves to be confronting electrons.

This is a powerful objection against the metaphysical version of experimental realism because, on its most natural reading, MER sanctions only those inferences to the existence of the particular properties that are actively being manipulated. Since it's only the properties of flectrons that scientists manipulate in an electron microscope, it should only be flectrons that scientists are entitled to presume are real, and so on. Hence the incoherence charge shows that MER comes into conflict with the actual ontological commitments of scientists.

In addition, the incoherence objection also undermines the following cognitive claim:

Belief Possession Experimental Realism (BPER)

Manipulation is a necessary and sufficient condition on possessing existential beliefs: A subject *S* can have the belief that *x* exists if and only if *S* manipulates *x*.

For suppose BPER is true; the incoherence charge then shows that scientists rarely ever have—warranted or unwarranted—beliefs in any theoretical entities, as they rarely ever manipulate all of an entity's properties at once. But this would conflict with scientific practice once again. For example, Margaret Morrison has provided the details of two case studies, the cloud chamber and charmed quarks, which show that manipulation can fail to induce the appropriate beliefs and even 'can occur in a context where there are no firmly held beliefs about the entities being manipulated' (Morrison 1990: 6). These two case studies show empirically that BPER is false in general.

However, I argue that experimental realism is not committed to BPER or MER but only to EER. On this epistemic version of experimental realism the "home truths" about an entity *x*, which we need to believe in for our inference that *x* is real to be causally warranted, need not in any way exhaust our concept of *x*. EER does not entail MER, so it does not entail that we can only infer those properties of *x* that we can manipulate or interact with. It entails instead that these are the properties that best ground our inference that *x* exists.[11]

This is just the commonsense view that we apply in our ordinary life. I infer certain properties of the city of Konstanz (for instance those pertaining to the relative positions of its station with respect to the Hotel Barbarossa and the University, and the average time intervals to walk or travel by bus between them, which Stephan Hartmann accurately described in his instructions sheet on how to get there) because I have manipulated and interacted with them in order to find my way around. But those "home truths" neither exhaust the city of Konstanz nor my concept of it. There are many other interesting properties of Konstanz, some of them observable (i.e. its Town Hall building), some of them not (i.e. its founding year, the number of its inhabitants, or its geographical borders) that I can suppose are real, and indeed believe to be real, but because I've had no opportunity to manipulate them, I have no causal warrant for them.

Neither does EER entail BPER: Manipulation is neither a sufficient nor a necessary condition on having, or acquiring, the belief that *x* exists. The source of our belief in some unobservable entity may in no way be related to the grounds that warrant that belief. Theory is undoubtedly, for most of us, the source of our belief in most unobservable scientific entities, including electrons. Most of us learn about electrons from our high-school classes on electromagnetism and particularly Maxwell's theory. This is a fact about our psychology. It is also possible that this is indeed the most educationally efficient and appropriate way to acquire such a belief. That would be a fact about pedagogy. Experimental realism, in the epistemic version that I wish to defend, makes the distinct and additional claim that our belief in electrons possesses a special sort of warrant, causal warrant; we can be particularly confident in our belief in electrons because we are confident that we routinely manipulate electrons, or their causal properties, in experimental conditions. The fact that this belief has this type of warrant pertains to epistemology and is *prima facie* independent of the psychological and pedagogical facts.

Most of us will only know that electrons have been manipulated by description; only a few of us have direct acquaintance with the operations required to correctly use an electron microscope. No matter. According to EER it is only because we believe that we can manipulate electrons that we have a causally warranted belief in their existence: The methods employed in acquiring the beliefs that there are electrons and that they can be manipulated are irrelevant to the epistemic, warranting-transferring relation that holds between them. Again this is the common sense view: I may have learnt about the facilities at Hotel Barbarossa from a theoretical description; but my corresponding beliefs acquire causal warrant only to the extent that I have manipulated those facilities. The acquisition of the belief need not have the same source, nor follow the same route, as the acquisition of the causal warrant.

To sum up, experimental realism is not a metaphysical thesis about what is real, nor is it a psychological thesis about the source or origin of our beliefs in the entities postulated by science; it is an epistemological thesis about the grounds that warrant those beliefs. The incoherence charge does not undermine *this* thesis.

Implausibility

The third, and in my view the sharpest, objection to experimental realism accepts that the position is coherent and that it does not conflict with scientific practice, but argues that it provides an implausible epistemology for science. Although one could come to have a warranted belief in the existence of an entity on the basis of laboratory manipulations, without believing in the truth of any particular theory about it, and although scientists often do so, it cannot be an epistemological principle that they *ought* to do so.

The objection has been pursued in an interesting paper by Christopher Hitchcock, which addresses directly Nancy Cartwright's argument for IMLC (Hitchcock 1992). In this paper Hitchcock presents two challenges for the experimental realist, which pertinently track what in my view are the two stages in the overall argument in favour of experimental realism. Hitchcock's first challenge is that antirealists about theoretical entities may also be able to accept causal explanations without thereby committing themselves to the reality of the cause.[12] In other words, Hitchcock questions that causal explanation is really a success term. 'What is special', he asks, 'about the role of explanation such that causal stories filling that role, and not some other, must be believed if accepted?' (Hitchcock 1992: 174).

The core of Hitchcock's first challenge is a couple of examples of putative causal explanations where the cause is most definitely not real. I will discuss only one of the examples, as they are argumentatively identical. He considers an explanation of the two-slit experiment that is sometimes offered in quantum mechanics textbooks. Electrons are fired through a screen with two slits *A* and *B* on it and are then detected in a further screen (see Figure 7.3). The pattern of detections of particles in the faraway screen does not correspond to the sum of the patterns registered when the experiment is repeated with slit *A* closed and *B* open (Figure 7.1), and slit *A* open and *B* closed (Figure 7.2). This is true even when only one electron at a time is passed through the slit.

Interestingly the pattern of Figure 7.3 is destroyed as soon as a measurement is made to detect which slit the electron actually goes through. Hitchcock considers the following possible explanation of this fact. A detection process of the electron in either slit will ultimately consist in bouncing a photon off the electron as it passes through the slit. This will impart momentum on the electron, which will affect its trajectory, thereby destroying the interference pattern. This story, argues Hitchcock, seems perfectly explanatory and is a description of causal processes. We must thereby, by IMLC, infer that there is such photon-electron interaction taking place.

However, as Hitchcock quickly points out, this explanation is unacceptable because it 'contradicts almost every interpretation of quantum mechanics, and as such would not be believed to be true by any but the most

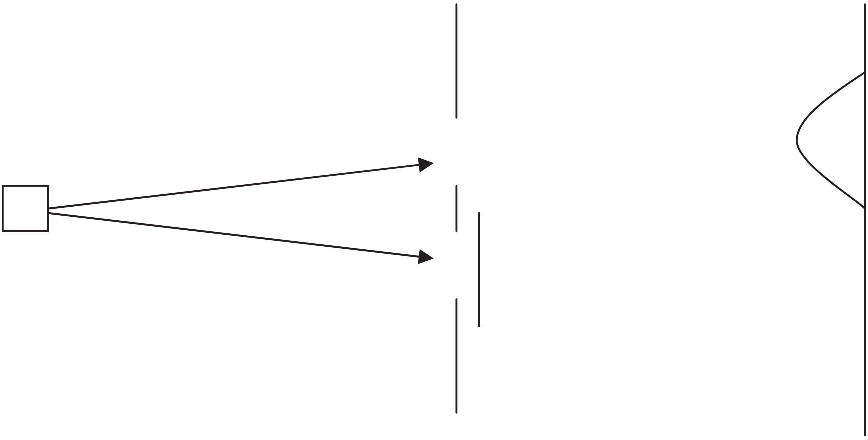

Figure 7.1

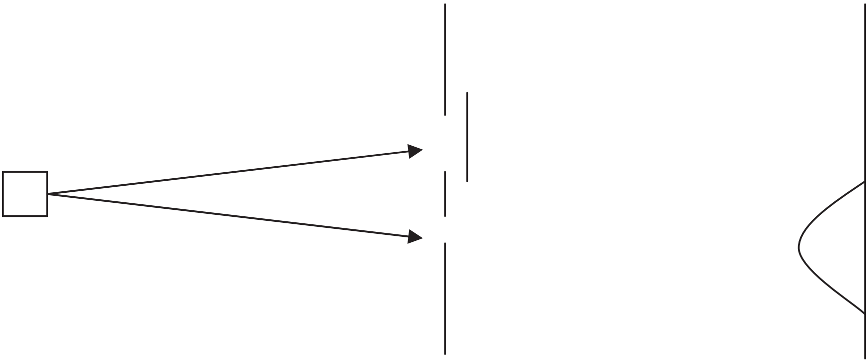

Figure 7.2

stubborn believer in hidden variables' (Hitchcock 1992: 171). According to quantum theory electrons do not have classical, continuous trajectories; but according to the causal story above, the entity putatively responsible for the observed interference pattern is precisely taken to be the electron's trajectory. Hitchcock concludes that IMLC cannot provide the sort of warrant that Cartwright takes it to: A causal story may be acceptable, as an explanation, even if the cause is most certainly not real.

However the defender of EER will reply that Hitchcock's argument is flawed, and cannot refute EER, because there is no causal explanation in the first place. In a causal explanation it is not the causal *story* that does the explaining but the causes themselves. The problem with the causal story above is that it presupposes an account of the interference pattern that we lack causal warrant for—and arguably have some causal warrant against. We are invoking the photon-electron interaction in order to explain not the interference pattern but rather its disappearance when we detect the electron

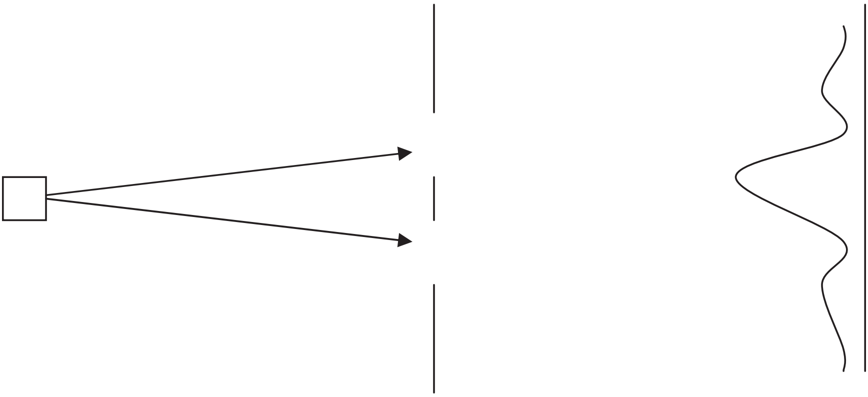

Figure 7.3

in the first screen. So the causal story presupposes that were we *not* to detect the electron's passage in the first screen, the electron's trajectory would have been causally responsible for the interference pattern. In other words we are presupposing that the interference pattern in a two-slit experiment is causally explained by the electrons' trajectories. And this "explanation" of the interference pattern is not causal—if it is an explanation at all—since it "would not be believed to be true by any but the most stubborn believer in hidden variables". And this is not so on simply interpretational or theoretical grounds but on the grounds of the experimental evidence against the existence of classical trajectories in quantum mechanics, in the form of all kinds of interferometry experiments.[13]

The defender of causal explanation takes causal explanation to be a success term. So if we don't believe in the "causes" appealed to in the story then the explanation the story offers—if any—cannot be said to be causal. The causal story that we are told about photon-electron interactions can only be accepted as an explanation in the sense that theoretical explanations can be; that is, we are given a theoretical account, and we are invited to deduce the phenomenon from it. But it has already been established that there is no truth-requirement on theoretical explanation. The fact that the theoretical account employs causal vocabulary is irrelevant. For a causal explanation of a phenomenon does not merely subsume a phenomenon under a theory that uses causal vocabulary: A causal explanation of a phenomenon such as the breakdown of the two-slit pattern must cite the actual causes of that phenomenon. And this is patently not the case in Hitchcock's example, as he himself acknowledges.

WHAT IT TAKES TO MAKE IMLC SOUND

In the epistemological version of experimental realism that I am defending, successful causal explanation provides warrant for the existence of the cause cited. But similarly, following Cartwright, some theoretical explanations provide warrant for the truth of the theories involved. The key to experimental realism, I am suggesting, is to distinguish carefully two types of warrant, which we may refer to as causal and theoretical.[14] The claim is then that causal warrant, i.e. warranted inference to the most likely cause, is conclusive (not "infallible") in the sense that only causal warrant to the contrary would force us to withdraw the existential commitment. In other words the commitment won't be defeated by an alternative theoretical explanation that dispenses with the entities so warranted.

By contrast, even the most warranted inference to the truth of a theoretical explanation would *ipso facto* be defeated by causal warrant to a cause whose existence contradicts the theory. Hence I am suggesting that inference to the most probable cause provides a type of warrant that is uniquely strong in that it can only be defeated by warrant of the same type.

Thus the full argument in favour of experimental realism has not one but two stages. The first stage has been provided by Cartwright's claim, already reviewed, that causal explanation, unlike theoretical explanation in general, is a success term. When "causal" and "theoretical" are understood properly this argument is, in my view, sound. Causal explanation has a truth-requirement built in: A cause can genuinely explain only to the extent that it is real, and we can only take it to explain if we believe it to be real. I have argued that Hitchcock's putative counterexamples are not actually such; and I know no other convincing counterexamples to Cartwright's claim so far.

The second stage of the argument is equally important but was, if anything, left implicit in Cartwright's original writings. It is the claim that an inference to the existence of an explanatory cause is more robustly warranted than an inference to the truth of an explanatory theory. Note that this claim, although not unrelated, is distinct from the truth-requirement one. Hitchcock's second challenge directly addresses this fact. (In what follows I change Hitchcock's terminology in order to distinguish a causal explanation from a theoretical explanation that employs causal vocabulary. I assume throughout that capital letters *P*, *Q* refer to sentences in some language, which may include causal terms—so "*P* causally explains *Q*" is a sentence in some theory that uses causal language. I reserve noncapital letters *p*, *q* to refer to entities, their properties, or phenomenological facts directly—so "*p* causally explains *q*" is not a theory but a statement of fact).[15]

Suppose that causal explanation *is* a success term; and suppose that I know that some phenomenon *q* occurs, and I find out, through careful manipulation in laboratory conditions, that *p* is the most likely cause of *q*. I am then invited by IMLC to infer that *p* is real. Hence "*p* is real" is the conclusion of an argument that has, as premises (i) *q* occurs and (ii) "*p* causally explains *q*". But, asks Hitchcock, since causal explanation is a success term, do I not need to believe in the existence of *p* in order to accept that '*p* causally explains *q*' in the first place? And if so, in what way is the inference to *p* in this argument providing me with any warrant in *p*'s existence that I did not already have?

The challenge is interesting and to the point, but it can be answered—and in an illuminating way. Let me first discount a trivially off-the-mark interpretation of Hitchcock's second challenge. Hitchcock is not criticising experimental realism for supposing that we must simultaneously believe that *p* is real and that *p* causally explains *q*; that is, he is not merely pointing out that in order to believe that *p* causally explains *q*, we need to already have the belief that *p* is real. This at best would argue against BPER, which I have already discounted as a misinterpretation of experimental realism. The fact that my belief "*p* causally explains *q*" necessarily presupposes my belief "*p* is real" is part and parcel of what it means for causal explanation to be a success term, so EER cannot be in the business of denying it.

The potential problem that Hitchcock is pointing to here is deeper, and quite general. The question is: How can we take the argument above—which appeals to a specific explanation of a particular phenomenon *q* by means of some putative cause *p*—to provide warrant for my (antecedent) belief that p is real? What are the conditions for a deductive argument to transmit warrant from its premises to its conclusion? Is it not clear that any such argument (where the truth of the conclusion must be presupposed in order to believe in the truth of the premises) would fail to transmit warrant and fail to provide me with a new reason to back up my belief in the conclusion?

Crispin Wright and Martin Davies have for some time been studying the mechanisms that underlie loss of warrant transmission in a deductively valid argument (Davies 1998; Wright 2000, 2003). Their particular target is McKinsey's argument for the incompatibility of externalism about mental content and privileged access to one's own mental contents. Roughly, McKinsey tried to show that these two premises together entail a priori knowledge of natural kinds, which he took to provide a reductio refutation of externalism (McKinsey 1991). More specifically, the argument is as follows: (i) I believe that water is wet; (ii). If I believe that water is wet then I belong to a community of speakers that has had contact with water; hence (iii) my community has had contact with water. Since (supposing privileged first person access) I know (i) a priori, and (supposing strong content externalism) I know (ii) a priori; it follows that I can know (iii), that there is water in my environment, a priori.

Wright and Davies defend the compatibility thesis against McKinsey's argument. They suggest that McKinsey's argument may be valid, but not cogent, in the sense that it fails to transmit warrant to its conclusion. So it offers us no reason to abandon externalism—or first-person access: We do not, in following the argument, acquire any warrant or justification for the conclusion. Wright and Davies claim, roughly, that the only evidence that I may ever possess in favour of the first premise in McKinsey's argument ("I believe that water is wet") is an introspective experience of the content of my mental state, which entails, on the strong conception of externalism that grounds the second premise, that I can know a priori that there is water in my environment. So the evidence for the first two premises jointly entail the conclusion; and McKinsey's argument, although valid, and possibly sound, does not transmit warrant.

The method is rather general, and can be—now quite precisely—summed up as follows: If the conclusion of an argument is a necessary presupposition for the evidence that we actually have to hand for its premises then the argument is not capable (for us) of transmitting warrant from the premises to the conclusion (even if the conclusion as well as the premises is true—and even if we correctly believe them all to be true!) A deductively valid argument with true premises will not warrant its conclusion if the only evidence that we possess in favour of the premises would not be evidence had the conclusion of the argument been false.[16]

With this in mind, let us now turn to Hitchcock's second challenge. Is the argument in question—(i) *q* is real, (ii) *p* causally explains *q*; therefore (iii) *p* is real—warrant transmitting? Is the conclusion (iii) a necessary presupposition of the evidence that we possess in favour of (i) and (ii)?

The reality of *p* is not a presupposition of any evidence we may possess for *q*—otherwise it would be impossible to empirically establish a phenomenon without thereby also establishing the reality of distinct, apparently unrelated, unobservable entities: Any empirical evidence in favour of, say, electrical conductivity would ipso facto establish the reality of electrons, without any need for further experimenting or reasoning. So the question is whether the reality of *p* is required for the evidence that we need to have to hand in order to accept premise (ii) to be counted as evidence. I do not believe this is generally the case, and hence I do not believe that IMLC fails, in general, to be warrant-transmitting. Let me explain why. There are two features that typically distinguish causal explanation and provide evidence that some explanatory claim is causal: lack of redundancy and a material mode formulation. I argue that these features provide evidence for "*p* causally explains *q*", regardless of whether *p* is real.

Nonredundancy

The nonredundancy requirement is met to a much larger degree by IMLC than by IBTE. Scientists establish which putative cause is nonredundant through controlled intervention and manipulation in laboratory conditions—and only then have they got reason to believe that the putative cause is genuinely responsible for the phenomenon.

By contrast, we have much poorer and controversial methods to establish which among all possible empirically adequate theoretical explanations of a phenomenon is the most probable one. (In the case of Duhemian theoretical explanation, we have no methods at all.) As is well-known, opinions on this matter differ enormously, both among philosophers and among practitioners: Is the most probable theoretical explanation the simplest one, the most ontologically parsimonious, the most familiar, the one that preserves the greatest amount of structure from previous theories, the one that explains a greater number of independent phenomena, etc, etc, etc. As a consequence there is much more redundancy, in the form of underdetermination, in the case of theoretical explanation. Lack of redundancy in an explanation is typical evidence in favour of it being a *causal* explanation.

In other words, we intervene and control variables in situations where we expect *p* to be operating in order to rule out redundancy in the explanation of *q*, which is in turn evidence that *p* causally explains *q*; and this provides warrant—by IMLC means—that *p* is real. The question is whether lack of redundancy on its own entails that *p* is real. If it does then Hitchcock is right, and IMLC cannot transmit warrant.

EER entails that we would only revise our causally warranted existential commitments through further experimentation that shows some other cause is more probable. This is an extremely rare occurrence, but it is not impossible. We arguably once had causal warrant for phlogiston but no longer do. The explanation of combustion is nowadays to be found in the interaction of oxygen with flammable materials—and we have acquired plenty of causal warrant in favour of oxygen and its role in combustion. So what we have here, arguably, is a case of causal warrant for the existence of an entity (phlogiston), and its role in combustion, that has been overturned by causal warrant for another entity (oxygen). We were convinced that we were able to manipulate phlogiston, and on that basis discarded any competing explanation of combustion; but we have since learnt that what we are actually able to manipulate is oxygen, which we have shown by experimental means to be involved in our present-day nonredundant explanation of combustion.[17]

This is in illuminating contrast with the case of the electromagnetic ether. It was generally accepted that it was not possible to manipulate or causally interact with the ether—even those few who thought that the ether might be manipulated were unable to convince themselves or others to have manipulated it.[18] So it would be wrong to say that we once had causal warrant for the existence of the ether. At best we had some theoretical warrant; and this was lost when we abandoned ether theories in favour of Einstein's relativity theory.

I am thus suggesting that manipulation provides only a particularly robust kind of warrant—causal warrant. We can never be certain that we are in fact manipulating *p*; at best we can be certain that we believe that we are manipulating *p*. This belief allows us to establish, by means of intervention in laboratory conditions, that *p* is nonredundant as a causal explanation of *q*, and this nonredundancy is one type of evidence that we need to possess in order to accept (ii) that "*p* causally explains *q*". But it should be obvious that both nonredundancy and the required belief in the manipulability of *p* are at best fallible evidence for (iii) "*p* is real" and do not, separately or jointly, logically entail it. (The fact that someone believes that he or she has interacted with aliens does not entail that there are aliens!)

Now the question is: does nonredundancy cease to be evidence in favour of the causal character of the explanation of *q* by means of *p* were *p* not real? Note that the question is not: "Can *p* be the causal explanation of *q* if *p* is not real?", to which we already know we must give a negative answer. Rather the question is: "Would the nonredundancy of explanation—established by what we take to be manipulation of *p* under experimental conditions—cease to constitute (defeasible) evidence in favour of the causal character of the explanation of *q* by *p* were *p* not real?"

Suppose that Priestley did establish to his own satisfaction, by the experimental means of intervention and (what he took to be) manipulation of phlogiston under laboratory conditions, that the only explanation of combustion involves the presence of phlogiston. Does all this painstaking and careful experimental and laboratory work not amount to (defeasible) evidence for Priestley in favour of the claim "phlogiston-release is the causal explanation of combustion?" It is hard to see what else could count as evidence in favour of such a claim. The reality or otherwise of phlogiston makes no difference whatever to the "evidential" character of the evidence in favour of the claim.[19] Priestley carried out every single experiment, intervention, and manipulation that he could have been expected to carry out in order to establish such a fact experimentally; and (we may suppose) he reported his experimental activity and results in full honesty. He was led by his prior belief in phlogiston to interpret all his experimental manipulations as providing grounds for the nonredundant role of phlogiston in the explanation of combustion.

Of course such evidence was defeasible and in fact turned out to be defeated; and we no longer interpret his work in those terms. But if (iii) is not presupposed by the character of the evidence for (ii) and (i), then IMLC, unlike McKinsey's argument, is in general capable of transmitting warrant. We do learn something after all when we infer from "*p* causally explains *q*" to "*p* is real"; what we learn is not that *p* is real—we already believed this—but that we have as good a reason as we could have to believe that *p* is real.

Hitchcock's second challenge then fails to refute EER. What it does undermine in fact is a competitor view, defended by neither Cartwright (1983) nor Hacking (1983), which we may refer to as the internalist version of "metaphysical entity realism":

> ***Internalist MER:*** *x* is real if we believe that we manipulate *x*.

This principle must strike everyone as obviously too strong, for reasons already pointed out: the mere belief that we have interacted with aliens does not make them real. But suppose it was true, then (iii) "*p* is real" would follow from the evidence that we need to possess in order to accept (ii) "*p* causally explains *q*", namely that we believe that we manipulate *p*; and IMLC could not transmit warrant, for the putative evidence in favour of (ii) would not be genuine evidence were the conclusion (iii) false. Hitchcock's second challenge thus refutes MER, and it provides yet one more argument in favour of understanding experimental realism as EER *only*.

Material Inference

One aspect of experimental realism that is rarely mentioned is Hacking's and Cartwright's insistence on the importance of semantic descent from the formal mode to the material mode. They themselves have not made clear what precise role this distinction plays in the argument in favour of experimental realism. The following conjecture is, I think, plausible: Causal warrant can accrue to the conclusion of an inference entirely carried out in the material mode; theoretical warrant on the other hand is always the result of an inference in the formal mode. Although the conclusions of such inferences are, as Carnap argued, intertranslatable, the vehicle of the inference, and the corresponding strength of the warrant transmitted to the conclusion, differ (Carnap 1935).

The difference might be best explained by means of the following example. Suppose that we would like to explain the phenomenon that metals dilate in the presence of heat. We could:

1. Formulate the corresponding phenomenological law and state it in a "protocol sentence".
2. State formally the solid-state physics theoretical treatment of metals, including the formal hypothesis that heat makes molecules vibrate with higher energy and thus forces them move further apart from each other.
3. Deduce from this theory together with the required boundary and initial conditions, the "protocol sentence" in 1.
4. Infer, by IBTE, the truth of the theory including the molecular assumption.
5. Infer by semantic descent the reality of highly energetic molecules in a solid.

Or, alternatively, we could:

1. Formulate the phenomenological law to be explained.
2. Assume that molecules vibrate with higher energy in heat, thus move further apart from each other.
3. Causally explain the law by appeal to the assumption (i.e. describe the experiments that show that no other cause of the expansion is as likely by manipulating the molecules of different samples in order to vary their energy and then checking whether the heat of the solid covaries accordingly).
4. Infer directly, by IMLC, the reality of highly energetic molecules in a solid.

The former inference is a formal inference to the best theoretical explanation, whereas the latter is a material inference to a most likely cause. The latter type of inference is more robust—it contains fewer steps where it may go wrong. In particular we need not worry about how appropriate or fair the translation is into the formal mode description in the first place.

But can the formal mode description not be applied to the causally explanatory argument directly? Yes, indeed: material-mode speech is not always required for causes; we can refer directly to the causes or refer to our statements and theories about those causes. The statement in material mode that 'magnetic fields can cause electrons to deflect' is translatable into (although not synonymous with) the statement: 'according to Maxwell's theory, "magnetic fields" are correlated with deflections in the trajectories of "electrons"'. What is relevant here is not how to couch the explanation but what the actual relata of the explanatory relation are: the distinction between genuine causal explanations that refer directly to the causes, and causal "sounding" theoretical explanations, or causal *stories*, is crucial in countering Hitchcock's first challenge to experimental realism.

However, no similar options are available for theoretical explanations. We cannot describe theories in the material mode, by definition, on pain of incurring a use/mention distinction fallacy. I cannot describe Newton's principles or Schroedinger's equation in a mode of speech that does not allow me to refer to sentences, theories, and language but only to real entities and their properties—unless I turn these principles and equations into the reality that they are aiming to describe. In other words, material mode speech is another hallmark of causal explanation, as only causal explanations can be cast entirely in this mode. The fact that an attempt at an explanation is given entirely in the material mode—that the relevant manipulations of the cause are presented or pointed to and not merely theoretically described—is fallible evidence that the attempted explanation is causal.

To conclude, the belief that we manipulate *p* turns out to be essentially involved in both types of evidence that we may possess for "*p* causally explains *q*". The question then arises as to whether our having this belief, on its own, entails "*p* is real". For if so, IMLC transmits no warrant and cannot provide us with any reason to believe that "*p* is real". But as a matter of fact, according to EER, the belief that we manipulate *p* does not entail that *p* is real. So there is no real reason to expect failures of warrant transmission in an IMLC: This type of inference transmits causal warrant, thus providing us with a new and particularly strong reason to back up our belief in the existence of the cause.

AN EXPERIMENTAL REALISM?

Material inference to the most likely cause is the norm in ordinary abductive reasoning; for examples one need go no further than one's own kitchen appliances or car mechanics. Whenever solving a problem with the normal functioning of our most familiar tools and appliances, we manipulate possible causes, provide evidence for causal explanations for the machines' malfunctions, and thereby infer most likely causes. Experimental realism, in the epistemic form that I defend it here (EER), brings the epistemology of science in line with our everyday epistemology. What I have called causal warrant is not special to science; it is precisely the kind of warrant typical of successful inquiries in everyday life. If scientific realism is characterised as the view that our beliefs in the unobservable entities postulated by science are in principle as warranted (or unwarranted) as the beliefs in the objects of ordinary life, then EER is just enough on its own to furnish a kind of realist epistemology.

However, this is a very limited and modest realism, and it is questionable whether it deserves the honorific, much sought-after title.[20] EER says that we have a stronger type of reason to believe in car engine pistons, washing machine filters, and electrons in electron microscopes than we do to believe in quarks and quasars. But it does not say we have no reason at all to believe in quarks and quasars, nor does it say that the existence of pistons, filters, and electrons is beyond any possible doubt.

Hence a commitment to EER is rather minimal, and it is hard to see how anyone, regardless of his or her additional epistemological commitments, would disagree with it. Even Van Fraassen's constructive empiricist could accept EER, although he would naturally add an independent principle to further privilege warrant in observable entities—a principle that the experimental realist will not accept. This suggests that EER is a good candidate for part of that elusive "core" that Arthur Fine argued realism and antirealism share (Fine 1997). Fine's Natural Ontological Attitude (NOA) was explicitly designed to capture this "common core", so I offer EER as a good candidate for a (part of) NOA.[21]

NOTES

1. This paper was written in 2002. Since then my views on this topic have sharpened considerably, in ways that do not line up with Cartwright's advice in her response. For instance, I would now be explicit that inference to the most likely cause (IMLC) is not defensible as a method of causal *discovery*. It can only be used as a method to warrant the ordinary presuppositions of laboratory work, whether or not they play any causal role. So I would no longer use the phrase "causal warrant", but "experimental warrant" instead. The ensuing shifts in terminology and argument are unfortunately too subtle to record in detail in the proofs (January 2008), and must await a paper of their own. In the meantime I emphasize these points in two brief footnotes in the final part of the paper.
2. The term "anomalous dappling" is due to Peter Lipton (Lipton 2002).
3. *The Dappled World* claims that theories postulating radically different laws and ontologies for different domains or at different levels of complexity may be simultaneously true; it does not follow from this that inconsistent theories may simultaneously be true in a way that would require a revision of classical logic.
4. Lipton's 2002 article was brought to my attention in the last stages of writing this chapter; it also argues for agnosticism, rather than atheism, about fundamentalism. Teller (this volume) essentially concurs with this agnosticism—but his arguments are orthogonal to those I present here.
5. I have occasionally tried to pull Nancy back to some of her earlier views. We even coauthored a brief paper (Cartwright et al. 1994) sketching a form of methodological instrumentalism about theories. This is an instrumentalism that, as I understand it, does not require any metaphysical backup, whether antifundamentalist or otherwise (see also Suárez, 1999).
6. This is a line of criticism adopted by Dorato (1988), Morrison (1990), and Resnik (1994). See also Elsamahi (1998), Gross (1990), and Reiner and Pierson (1995).

7. It is an interesting question (which I can not pursue here) whether the implication would be restored if EER were given an externalist twist, as follows. Externalist EER: our belief that *x* exists acquires causal warrant if we actually manipulate *x*. Manipulation is arguably a success term, so causal warrant would then become nondefeasible, and MER would seem to be implied by Externalist EER. There are however two important caveats. First, Externalist EEP stipulates that manipulation is a sufficient not a necessary condition on causal warrant—so there might be other sources of causal warrant, which would explain how it is sometimes defeasible. And second, even if there were no other sources of causal warrant (i.e. even if actual manipulation was a necessary as well as sufficient condition), one could still accept (Externalist EER) and insist that causal warrant is defeasible while denying that manipulation is in actual fact a success term. In either case (MER) would not follow from (Externalist EER).
8. Although of course, (MER) and (EER) may both happen to be true. My point is that experimental realism, in the epistemic version I wish to defend, neither requires nor provides grounds for (MER).
9. Cartwright's original term is "most *probable* cause". I prefer "most *likely*", because, strictly speaking, probabilities are only defined over statements, theories, hypotheses, or events but not entities or their properties.
10. It will become clear that I do not think the requirement is met exactly as worded here—but I do agree that there is a significant difference in the *degree* of redundancy in each case.
11. In correspondence, Peter Lipton points out that an analogy with descriptivist semantics illustrates vividly why the incoherence objection is off the mark: 'LSE philosopher whose name ends in "right" refers to all of Nancy Cartwright and her properties, even though it only mentions one most trivial property'. I agree: The flectron objection trades in a sense on ignoring the distinction between reference and mention.
12. This challenge was first raised by Fine (1991).
13. Neutron interferometry experiments are good instances. A noteworthy attempt to resist the weight of evidence is of course Bohmian mechanics, which notoriously restores well-defined trajectories by reinterpreting the evidence as a consequence of the radical nonlocal character of the quantum potential or guiding wave in configuration space. Brown et al. (1995) provide a critical analysis.
14. Nowadays I would distinguish "theoretical" and "experimental" warrant. (Note added to proofs, January 2008).
15. Hitchcock assumes that all inferences are carried out in the formal mode, since he assumes that the only way to infer that *p* is real from the fact that *p* causally explains *q* is to infer the sentence *P*: "*p* is real" from the sentences *Q*: "*q* is real" and the theory *T*: "*P* causally explains *Q*". That is, he assumes incorrectly that the relata of causal explanations are, as is the case for theoretical explanation, sentences or theories. Steve Clarke, in his otherwise cogent defence of Cartwright's argument, also turns IMPC into a subspecies of IBTE, i.e. precisely those IBTE that do not suffer from redundancy (Clarke 2001). The above considerations suggest that this move upwards in the semantic ladder is not without consequences and in fact already gives half the game away against experimental realism.
16. In other words, an argument does not transmit warrant if the evidence for the premises would lose its character as evidence were the conclusion false. And indeed, the introspective experience of my own mental states would remain the same even if there were no external world, but it could no longer be taken as evidence that the water in my environment is wet.

17. Musgrave (1976) is a good summary of the complicated history of the overturn of phlogiston theories by Lavoisier's theory.
18. Warwick describes the wonderful case of Joseph Trouton's failed efforts to build a perpertuum mobile machine out of the earth's interaction with the ether (Warwick 1995).
19. It of course makes a crucial difference to the actual truth-value of the claim. My point is that it makes no difference to the fact that Priestley's manipulations were rightly taken by him as evidence that the phlogiston explanation was nonredundant. We now have collected much stronger evidence in favour of oxygen, which makes Priestley's own explanation redundant—but he could not have anticipated this.
20. I would nowadays emphasize this point even more strongly. The view defended in this paper is not intended to provide an inductive method of discovery for unobservable entities. What I call *experimental warrant* can only provide support for our antecedently held existential commitments. The aim of this paper's project is to start deflating Hacking's and Cartwright's views, by extracting the "realism" out of the "experimental realism" (Note added to proofs, January 2008).
21. I would like to thank the participants at the Konstanz conference (December 2002) and the students at my doctoral course at Complutense (2002–2003) for their reactions; and Hasok Chang, Christopher Hitchcock, Carl Hoefer, and Peter Lipton for their detailed comments and helpful suggestions. This chapter has been supported by projects PR27/05-13879 and HUM2005-07187-C03-01 of the Spanish Ministry of Education and Science.

REFERENCES

Brown, H. et al. (1995) 'Bohm particles and their detection in the light of neutron interferometry', *Foundations of Physics*, 25: 329–347.

Carnap, R. (1935) *Philosophy and Logical Syntax*, Bristol: Thoemmes Press.

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Oxford University Press.

———. Cartwright, N. (1999) *The Dappled World*, Cambridge: Cambridge University Press.

———. (2002) 'Summary' and 'Reply', *Philosophical Books*, XLIII: 241–243, 271–278.

Cartwright, N. et al. (1995) 'The tool box of science', in *Theories and Models in Scientific Processes*, W. Herfel et al. (eds) Poznan Studies in the Philosophy of Science and the Humanities, Rodopi, 44: 137–149.

Clarke, S. (2001) 'Defensible territory for entity realism', *British Journal for the Philosophy of Science*, 52: 701–722.

Davies, M. (1998) 'Externalism, architecturalism, and epistemic warrant', in C. Wright et al. (eds) *Knowing Our Minds*, Oxford: Oxford University Press.

Dorato, M. (1988) 'The world of worms and the quest for reality', *Dialectica*, 42: 3.

Elsamahi, M. (1996) 'Could theoretical entities save realism?', *PSA 1994*, pp. 173–180.

Fine, A. (1997) *The Shaky Game: Einstein, Realism and the Quantum Theory*, 2nd edn, Chicago: University of Chicago Press.

———. (1991) 'Piecemeal realism', *Philosophical Studies*, 61: 79–96.

Gross, A. (1990) 'Reinventing certainty: The significance of Ian Hacking's realism', *PSA 1990*, 1: 421–431.

Hacking, I. (1983), *Representing and Intervening*, Cambridge: Cambridge University Press.

———. (1984) 'Experimentation and scientific realism', in J. Leplin (ed.) *Scientific Realism*, Berkeley: University of California Press.

———. (1989) 'Extragalactic reality: The case of gravitational lensing', *Philosophy of Science* 56: 578.

Hitchcock, C. (1992) 'Causal explanation and scientific realism', *Erkenntnis*, 37: 11–178.

Hoefer, C. (2003) 'For fundamentalism', *Philosophy of Science, PSA 2002*, 70.5, pp. 1401–1412.

Lipton, P. (2002) 'The reach of the law', *Philosophical Books*, XLIII: 254–260.

McKinsey, M. (1991) 'Anti-individualism and privileged access', *Analysis*, 51: 9–16.

Morrison, M. (1990) 'Theory, intervention and realism', *Synthese*, 82: 1–22.

Musgrave, A. (1976) 'Why did oxygen supplant phlogiston: Research programmes in the chemical revolution', in C. Howson (ed.) *Method and Appraisal in the Physical Sciences*, Cambridge: Cambridge University Press.

Resnik, D. (1994) 'Hacking's experimental realism', *Canadian Journal of Philosophy*, 24: 395–412.

Reiner, R., and R. Pierson. (1995) 'Hacking's experimental realism: An untenable middle ground', *Philosophy of Science*, 62: 60–69.

Warwick, A. (1995) 'The sturdy protestants of science: Larmor, Trouton and the motion of the earth through the ether", in J. Z. Buchwald (ed.) *Scientific Practice: Theories and Stories of Doing Physics*, Chicago: University of Chicago Press.

Wright, C. (2000) 'Cogency and question-begging: Some reflections on McKinsey's paradox and Putnam's proof", in *Philosophical Issues*, 10: 140–163.

———. (2003) 'Some reflections on the acquisition of warrant by inference', in S. Nuccetelli (ed.), New Essays on Semantic Externalism and Self-Knowledge, MIT Press, pp. 57–78.

Reply to Mauricio Suárez

I am happy to support Mauricio Suárez in his programme to develop and defend experimental realism. In particular I am happy to endorse as a sound rule of thumb his doctrine that good causal evidence can be expected to trump derivation from even very well-confirmed theory. Suárez here provides serious, careful answers to objections to my closely related claims in support of entity realism over theoretical realism, but he does not explain in any detail why causal evidence, or directly relevant evidence from a good experiment, will generally weigh more than deductions from a good theory. Here I shall review some of the reasons for expecting this to be true. In any particular case, though, I would suppose that the issue will hinge on how good the experiment is versus how good the theory and how secure the deduction. Importantly, it matters how well-confirmed the theory is for cases very like the one in question, where I mean "like" in well-understood ways; what matters about the confirmation of the theory is not variety of evidence but rather the weight of evidence for the very particular specification of the theory used in the derivation.

I begin with entity realism because the arguments I would deploy in favour of Suárez's experimental realism are, not surprisingly, similar to those that moved me to entity realism in *How the Laws of Physics Lie*. The context is theory versus less regimented claims to knowledge. Many take the claims of science to be far more secure than everyday claims because well tested; others take them to be groundless because they are about postulated, unobservable entities. I took a position in the middle and still do. The very abstract principles of high-level theory do not have much claim to knowledge precisely because they have not been properly tested. By contrast we do have good evidence for many of our claims about the existence of theoretical entities, their characteristics, and their behaviours in very specific circumstances. These latter will be complicated, detailed, highly concrete claims that generally use a mix of languages from different fields and of different types—mathematical, material, theoretical, technical, and everyday. These are the "phenomenological laws" of *How the Laws of Physics Lie*. Reasons for thinking that a low-level phenomenological law with good experimental support is likely to be far more warranted than claims that are derived from our best high theory fall under three headings.

EXCESS MATHEMATICAL STRUCTURE

Our mathematical representations are far richer than the phenomena they represent. So how much of the results we derive depend on aspects of the mathematics that represent genuine physically relevant features that we have good reason to believe in and how much depends on the excess structure? We need representation theorems to sort out this question and these are rare in physics. The problem is exacerbated as our mathematical representations build upon one another; as this happens there is seldom any attempt to provide independent characterizations of the physical features the mathematics is supposed to represent, let alone to prove representation theorems to justify the mathematical forms. For most of contemporary fundamental physics, it probably does not even make sense to talk of extracting the representative part of the mathematics from the excess structure. The mathematical representation has a life of its own. In such a situation we have little idea how much we can trust conclusions derived from theory unless they have independent warrant of the kind Suárez would call *experimental* or *causal*.

THE ROUTE FROM EVIDENCE TO APPLICATION

Often neither evidence nor application follow from the theory without correction, in which case it is not the hypotheses that have been confirmed that imply the conclusions derived. Warrant cannot then flow via the hypotheses from the evidence for the theory to its deductive consequences.

WARRANT AND THE SCOPE OF INDUCTIONS

We may infer a vary particular specification of a theoretical law from a number of its instances, similarly for a second and a third particular specification and perhaps many more, each from instances that clearly fall under it. Each induction is shaky in well-known ways. But if an induction to this or that particular specification of a law is shaky, the meta-induction to the general form of the law is far shakier still. Claims that follow from the law in this general form not backed up by independent evidence from the particular specification implicated in the derivation are bound to be less warranted than the ones backed up by particular specific forms of the general laws that have direct evidence in their favour.

The culprit, in all these cases, that provides the appearance of warrant for the high-level claims is the hypothetico-deductive method. Despite universal recognition that this method commits the fallacy of affirming the consequent, it seems to be generally presupposed by realists and antirealists alike; hence the scramble to find some possible truth-making characteristics for the "best" explanation; simplicity, unification, invariance. . . . I am convinced that the fallacy is a fallacy and also that the case that any of these "nice" features are truth-makers has not been made. Suárez aims to defend causal experimental warrants over theoretical warrant. I take it then that one of the principal jobs confronting him is to refocus our attention from deductive to inductive methods and to study how best to formulate these for contemporary science.

8 Cartwright's Realist Toil

From Entities to Capacities[1]

Stathis Psillos

INTRODUCTION

Nancy Cartwright has been both an empiricist and a realist. Where many philosophers have thought that these two positions are incompatible (or, at any rate, very strange bedfellows), right from her first book, the much-discussed and controversial *How the Laws of Physics Lie*, Cartwright tried to make a case for the following view: if empiricism allows a certain type of method in its methodological arsenal (inference to the most likely cause), then an empiricist cannot but be a scientific realist—in the metaphysically interesting sense of being ontically committed to the existence of unobservable entities. Many empiricists thought that because empiricism has been traditionally antimetaphysics, it has to be antirealist. One of the major contributions that Cartwright has made to philosophy of science is, I think, precisely this: there is a sense in which metaphysics can be respectable to empiricists. Hence, scientific realism cannot be dismissed on the grounds that it ventures into metaphysics. To be sure, the metaphysics that Cartwright is fond of is not of the standard a priori (or armchair) sort. It is tied to scientific practice and aims to recover basic elements of this practice (e.g., causal inference). But it is metaphysics, nonetheless.

Cartwright's realism has been described as "entity realism". This is not accidental. She has repeatedly made claims such as 'I believe in theoretical entities' (Cartwright 1983: 89, see also 92). Typically, she contrasts her commitment to entities to her denial of "theoretical laws". In the sections 'Causal explanation' and 'Causal inference', I examine in some detail the grounds on which Cartwright tried to draw a line between being committed to entities and being committed to theoretical laws, and I find them wanting. In 'Causal inference' I also claim that the method Cartwright articulated, Inference to the Most Likely Cause, is important but incomplete. Specifically, I claim that there is a more exciting method that Cartwright herself describes as Inference to the Best Cause, which, however, is an instance, or a species of Inference to the Best Explanation. But Cartwright has been against Inference to the Best Explanation (IBE). So, in the section 'Why deny inference to the best explanation?' I consider and try to challenge Cartwright's central argument against IBE.

At least part of the motivation for her early, restricted, realism was a certain understanding of what scientific realism is. She took scientific realism to entail the view that the world has a certain hierarchical structure, where the more fundamental laws explain the less fundamental ones as well as the particular matters of fact. In *The Dappled World*, she rightly disentangled these issues. 'Nowadays', she says, 'I think I was deluded by the enemy: it is not *realism* but *fundamentalism* that we need to combat' (Cartwright 1999: 23). What, I think, emerges quite clearly from her later writings is that Cartwright does not object to realism. Rather, she objects to Humeanism about laws, causation, and explanation. Insofar as Humeanism is a metaphysics independent of scientific realism, Cartwright is a more full-blown realist, without being Humean. And this is what she is. In the penultimate section, 'Capacities', I discuss in some detail Cartwright's central non-Humean concept, viz., capacities. Cartwright is a strong realist about capacities. They are the fundamental building blocks of her metaphysics. But there seem to be a number of problems with capacities. Though we can easily see how attractive it is to be a realist about capacities, I think it's really hard to be one. So, though Humeanism is certainly independent of scientific realism, I argue that we have not been given compelling reasons for a non-Humean metaphysics of capacities.[2]

It is helpful to state clearly five worries about Cartwright's views that I develop in this paper. The first is that though she was right to insist on the ontic commitment that flows from causal explanation, she was wrong to tie this commitment solely to the entities that do the causal explaining. This move obscured the nature of causal explanation and its connection to laws. The second worry is that when she turned her attention to causal inference, by insisting on the motto of "the most likely cause", she underplayed her powerful argument for realism. For she focused her attention on an extrinsic feature of causal inference (or, indeed, of any ampliative inference), viz., the demand of high probability, leaving behind the intrinsic qualities that causal explanation should have in order to provide the required understanding. The third is that her objections to Inference to the Best Explanation were unnecessarily tied to her objections about the falsity of fundamental laws. Fourth is that though her argument for positing capacities and being realist about them was supposed to take strength from its parallel with Sellars's powerful argument for the indispensable explanatory role of positing unobservable entities, there are important disanalogies between the two arguments that cast doubt on the indispensability of capacities. The final (fifth) worry is that laws—perhaps brute regularities—might well have to come back from the front door, as they are still the most plausible candidates for explaining why objects have the capacities to do what they can do.

CAUSAL EXPLANATION

One of Cartwright's central claims is that causal explanation is ontically committing to the entities that do the explaining (Cartwright 1983). Here are some typical statements of it:

> That kind of explanation succeeds only if the process described actually occurs. To the extent that we find the causal explanation acceptable, we must believe in the causes described (Cartwright 1983: 5).
>
> In causal explanations truth is essential to explanatory success (Cartwright 1983: 10).
>
> But causal explanations have truth built into them (Cartwright 1983: 91).
>
> (. . .) existence is an internal characteristic of causal explanation (Cartwright 1983: 93).

These assertions are not all equivalent to one another, but I do not dwell on that. For, there is indeed something special with causal explanation. So, let's try to find out what it is. As a start, note that it is one thing to say that causal explanation is ontically committing but quite another thing to say what a causal explanation *is*. Let's take them in turn.

Ontic Commitment

If *c* caused *e*, then, clearly there must be events *c* and *e* which are thus causally connected. This follows almost directly from the standard Davidsonian account of singular causal statements. Causation is not quite the same as causal explanation, but causes do explain their effects, and there is, to say the least, no harm in saying that if *c* causes *e* then *c* causally explains *e*. This feature of causal explanation by virtue of which it is ontically committing to whatever does the causing is not peculiar to it. Compare the relation *c* preceded *e*: *c* must exist in order to precede *e*. So, Cartwright's claim is an instance of the point that the relata of an actual relation *R* must exist in order for them to be related to each other by *R*. I think this is what Cartwright should mean when she says that '(. . .) existence is an internal characteristic of causal explanation' (Cartwright 1983: 93).

An equivalent way to show that causal explanation is ontically committing is this. To say that the statement "*c* causally explains *e*" is ontically committing to *c* and *e* is to say that "*c* causally explains *e*" is true. This way of putting things might raise the spectre of van Fraassen, as Hitchcock reminds us (Hitchcock 1992). Couldn't one just accept that "*c* causally explains *e*" without believing that it is true? And if so, couldn't one simply avoid the relevant ontic commitments to whatever entities are necessary to make this statement true? Indeed, insofar as we can make sense of an attitude towards a statement with a truth-condition which involves acceptance but not belief, van Fraassen is on safe ground here. He is not forced to believe in the truth of statements of the form "*c* causally explains *e*". Cartwright's point, however, is not meant to be epistemic. Her point is, I think, twofold. On the one hand, she stresses that we cannot avoid commitment to the things that are required to make our assertions true. On the other hand (and more importantly), insofar as we do make some assertions of the form "*c* causally explains *e*" (e.g., about observable events such as shortcircuits and fires or aspirins and headaches), there is no reason not to make others (e.g., about unobservable entities and their properties).

So, causal explanation is egalitarian: It sees through the observable–unobservable distinction. It is equally ontically committing to both types of entity, precisely because the relation of causal explanation is insensitive to the observability of its relata. In other words, what matters for ontic commitment is the causal bonding of the relata of a causal explanation. So, Cartwright's point is that there is just one way to be committed to entities (either observable or unobservable) and it is effected through causal explanation.

What Exactly Is a Causal Explanation?

This remains an unsettled question, even after it is accepted that causal explanation is ontically committing. The question, in a different form, is this: What exactly is the relation between *c* and *e* if *c* causally explains *e*? In the literature, there have been a number of attempts to explain this relation. I do not discuss them here.[3] Cartwright has offered a gloss of the relation *c causally explains e*. She put forward an early version of the contextual unanimity principle, viz., the idea that *c* causes *e* iff *c* increases the probability of *e* in all situations (contexts) which are causally homogeneous with respect to the effect *e* (Cartwright 1983: 25–26). I do not dwell on this principle here. But one thing is relevant. Although principles such as the above do cast some light on the notion of causal explanation, they do not offer an analysis of it, as they presuppose some notion of causal law or some notion of causally homogeneous situation. Cartwright is very clear on this when she says, for instance, that what makes the decay of uranium 'count as a good explanation for the clicks in the Geiger counter' is not the probabilistic relations that obtain between the two events 'but rather the causal law—"Uranium causes radioactivity"' (Cartwright 1983: 27). Still, it might be said that though Cartwright does not offer 'a model of causal explanation' (Cartwright 1983: 29), she does constrain this notion by objecting to certain features that causal explanation is taken to have. Most centrally, she objects to the deductive-nomological model of causal explanation. But it is not clear, for instance, that she takes a singularist account of causal explanation. In fact, it seems that she doesn't. For she allows that certain 'detailed causal principles and concrete phenomenological laws' are involved in causal explanation (Cartwright 1983: 8). Her objection is about laws captured by 'the abstract equations of a fundamental theory' (Cartwright 1983: 8). So, even if she objects to the thesis that all causal explanation should be nomological, she doesn't seem to object to the weaker thesis that at least some causal explanation should be nomological. In any case, it's one thing to deny that the laws involved in causal explanation are the abstract high-level laws of a theory and it is quite another to deny that laws, albeit low-level ones, are involved in, or ground, causal explanation. For all I know, Cartwright does not deny the latter (Cartwright 1983).

Here is the rub, then. If laws are presupposed for causal explanation, then it's no longer obvious that in offering causal explanations we are committed just to the relata of the causal explanation. To say the least, we should also be committed to a Davidson-style compromise that there are laws that govern the causal linkage between cause and effect. Though these laws might not be stateable or known, they cannot be eliminated. But this is not the end of it. Considering Davidson's idea, Hempel noted that when the existence of the law is asserted but the law is not explicitly stated, the causal explanation is comparable to having 'a note saying that there is a treasure hidden somewhere' (Hempel 1965: 349). Such a note would be worthless unless 'the location of the treasure is more narrowly circumscribed'. Think of it as advice: where there is causal explanation, search for the law that makes it possible. It's a side issue whether this law is a fundamental one or a phenomenological one or what have you. This is a worry about the kinds of law there are and not about the role of laws in causal explanation.

So here is my first conclusion. Cartwright's advertised entity-realism underplays her important argument for ontic commitment. In offering causal explanations, we are committed to much more than entities. We are also committed to laws, unless of course there is a cogent and general story to be told about causal explanation that does *not* involve laws. Note that it is not a reply to my charge that there might be a singular causal explanation. This is accepted by almost everybody—given the right gloss on what it consists in. Nor would it be a reply to my charge that, occasionally, we do not rely on laws to offer a causal explanation. A suitable reply would have to show that causal explanation is totally disconnected from laws. This kind of reply might be seen as being offered by Cartwright when she introduces capacities. But, as we shall see in the section 'Capacities', it is at least questionable that we can make sense of capacities without reference to laws.

CAUSAL INFERENCE

Given the centrality of causal explanation in Cartwright's argument for realism, one would have expected her to stay firmly in the business of explaining its nature. But Cartwright does something *prima facie* puzzling. She spends most of *How the Laws of Physics Lie* (1983) on an attempt to cast light on the nature of the inference that takes place when a causal explanation is offered and on the conditions under which this inference is legitimate. (Doesn't that remind you of what Hume did?) One way to read what Cartwright does is this: she is concerned with showing when a potential causal explanation can be accepted as the actual one. More specifically, she is concerned with showing that there is something special in causal explanatory inference that makes it sound (or, at any rate, makes it easier to check whether it is sound or not). She says:

> Causal reasoning provides good grounds for our beliefs in theoretical entities. Given our general knowledge about what kinds of conditions and happenings are possible in the circumstances, we reason backwards from the detailed structure of the effects to exactly what characteristics the causes must have to bring them about. (Cartwright 1983: 6)

Thus put, causal reasoning is just a species of ampliative reasoning. From an epistemic point of view, that the explanation offered in this reasoning is causal (that is, that it talks about the putative causes of the effects) is of no special importance. What matters is what reason we have to accept the conclusion about the putative cause.

This seems to me a crucial observation. Cartwright explicitly draws a contrast between "theoretical explanation" and "causal explanation" (Cartwright 1983: 12). But this is, at least partly, unfortunate. For it obscures the basic issue at stake. *Qua* inferential procedures, causal explanation and theoretical explanation are on a par. They are each species of ampliative reasoning, and the very same justificatory problems apply to both of them (perhaps to a different degree).

Cartwright does think that there is something special in the claim that the inference she has in mind relies on a *causal* explanation. She calls this inferential process 'inference to the most likely cause' (Cartwright 1983: 6)—henceforth, IMLC. But there is a sense in which the weight is on the "most likely" and not on the "cause". It's just that Cartwright thinks that it's most likely to get things right if you are looking for causes than if you are looking for something else (e.g., general theoretical explanations). Before we see whether this is really so, let us press the present point a bit more.

Inference to the Most Likely Cause

What kind of inference is IMLC? An obvious thought is that we infer the conclusion (viz., that the cause is *c*) if and only if the probability of this conclusion is high. But this is a general constraint on any kind of ampliative inference with rules of detachment, and hence there is nothing special in IMLC in this respect. A further thought then might be that in the case of IMLC there is a rather safe way to get the required high probability. The safety comes from relevant background knowledge of all sorts: that the effect has a cause, because in general effects do have causes; that we are offered a rather detailed story as to what the causal mechanism is and how it operates to bring about the effect; that we have controlled for all(?) other potential causes, etc. (Cartwright 1983: 6). All this is very instructive. However, thus described, IMLC gets its authority not as a special mode of inference where the weight is carried by the claim that *c* causally explains *e* but from whatever considerations help increase our confidence that the chosen hypothesis (viz., that it was *c* that caused *e*) is likely to be true. If these considerations are found wanting (if, for instance, our relevant background knowledge is not secure enough, or if we do not eliminate all potential alternative causes, or if the situation is very complex), then the claim that *c* causally explains *e* is inferentially insecure. It simply cannot be drawn, because it is not licensed as likely.

Indeed, my present complaint can be strengthened. Consider what Cartwright says: '(. . .) causal accounts have an independent test of their truth: we can perform controlled experiments to find out if our causal stories are right or wrong' (Cartwright 1983: 82). If we take this seriously, then all the excitement of IMLC is either lost or becomes parasitic on the excitement of a controlled experiment. It is lost if for every instance of an IMLC it is required that a controlled experiment is performed to check the conclusion of the inference independently. So, what if the excitement of IMLC becomes parasitic on the excitement of a controlled experiment? Controlled experiments are indeed exciting. But their excitement comes mostly from the fact that they are designed to draw safe causal conclusions, irrespective of whether there is on offer a causal explanation of the effect. When it is established by a clinical trial that drug *D* causes relief from symptom *S*, we may still be in the dark as to *how* and *why* this is effected, what the mechanisms are, what the detailed causal story is, etc. I think that causal explanation—*qua* inference—is exciting not just because we can get conclusions that are likely to be correct, but also because we get an understanding of *how* and *why* the effect is produced. But so far, we have got only (or mostly) the former. The hard question, I think, remains unaddressed: What is this (if anything) in virtue of which a causal explanation—*qua* an explanatory story—licenses the conclusion that it is likely to be correct? Put in more general terms, the hard problem is to find an intrinsic feature of causal explanation in virtue of which it has a claim to correctness and not just an extrinsic feature, viz., that there are independent reasons to think it is likely.

Inference to the Best Cause

Cartwright seems aware of the need for such an intrinsic feature. Occasionally, she describes IMLC as 'inference to the best cause' (Cartwright 1983: 85). I think this is not just a slip. Reference to "best cause" is not just meant to *contrast* IBC to Inference to the Best Explanation (IBE), by replacing "explanation" with "cause". It is also meant, rightly I think, to *connect* IBC to IBE. It is meant to base the inference (the detachment of the conclusion) on certain features of the connection between the premises and the conclusion, viz., that there is a genuinely explanatory relation between the explanation offered and the explanandum. The "best cause" is not *just* a likely cause; it is a putative cause that causally explains the effect in the sense that it offers genuine understanding of how and why the effect was brought about. Cartwright says of Perrin's "best cause": 'we are entitled to [infer the existence of atoms] because we assume that causes make effects occur in just the way they do, via specific, concrete causal process' (Cartwright 1983: 85). If all we were interested in was high probability, then we wouldn't go for specific, concrete causal processes—for the more detail we put in, the more unlikely they become. The specific, concrete causal processes matter for understanding, not for probability.

The upshot is that if we conceive causal inference as Inference to the Best Cause (IBC), then it is no longer obvious that it is radically different from what has come to be known as Inference to the Best Explanation (IBE). The leading idea behind IBE—no matter how it is formulated in detail—is that explanatory considerations are a guide to inference. The inference we are concerned with is ampliative—and hence deductively invalid. But this is no real charge. Inferential legitimacy is not solely the privilege of deductive inference. IBC can then be seen as a species of IBE. It's a species of a genus, whose *differentia* is that in IBC the explanations are causal (see Psillos 2002b for details).

What sort of inference is IBE? There are two broad answers to this. (1) We infer to the probable truth of the likeliest potential explanation insofar as and because it is the likeliest explanation. On this answer, what matters is how likely the explanatory hypothesis is. (2) The best explanation, *qua* explanation, is likely to be true (or, at least more likely to be true than worse explanations). That is, the fact that a hypothesis H is the *best* explanation of the evidence issues a warrant that H is likely. The late Peter Lipton noted that the first answer views IBE as an inference to the Likeliest Potential Explanation, whereas the second views it as an inference to the Loveliest Potential Explanation (Lipton 1991: 61–65). The loveliest potential explanation is 'the one which would, if correct, be the most explanatory or provide the most understanding' (Lipton 1991: 61).

Exactly the same distinction applies to causal inference. If we think of it as an Inference to the Most Likely Cause (IMLC), then, as we have seen, the inferential weight is carried by the likeliness of the proposed causal explanation. So, it's not that a causal explanation is offered that licenses the inference. Rather, it is that this proposed explanation has been rendered likely. This rendering is extrinsic to the explanatory quality of the proposed explanation and relates to what we have done to exclude other potential explanations as likely. On the other hand, if we think of causal inference as Inference to the Best Cause, we are committed to the view that the inferential weight is carried by the explanatory quality of the causal explanation offered, on its own *and* in relation to competing alternatives. Roughly put, the weight is carried by the understanding offered by the causal story and by the explanatory qualities that this story possesses.

Indeed, Cartwright speaks freely of "causal accounts" or "causal stories" offered by causal explanations. The issue then is not just to accept that there must be entities that make these causal accounts true. It is also to assess these accounts *qua* explanatory stories. If we take IBC seriously, there must be ways to assess these accounts, and these ways must be guides to whether we should accept them as true. It seems then that we need to take account of explanatory virtues (a) if we want to make IBC have a claim to truth; and (b) if we want to tie this claim to truth not just to extrinsic features of causal explanation (e.g., that it is more likely than other potential explanations) but also to intrinsic features of the specific causal explanatory story.

So, let me draw the conclusion of this section. Thinking of causal explanation as an inference to the best cause will require assessing the causal story offered, and this is bound to be based on explanatory considerations which align IBC to IBE.[4]

WHY DENY INFERENCE TO THE BEST EXPLANATION?

It is well known, however, that Cartwright resists IBE (Cartwright 1983). And it is equally well known that she thinks she is not committed to IBE, when she vouches for IBC. So the issue is by no means over. Cartwright explicitly denies that 'explanation is a guide to truth' (Cartwright 1983: 4) and discusses this issue quite extensively (Cartwright 1983). Due to lack of space, I focus on one of her arguments, which seems to me to be the most central one. This is the argument from the falsity of laws. But before I go into this, allow me to note an interesting tension in her current views on the matter.

The Transcendental Argument

Cartwright has always tried to resist global applications of IBE. In particular, she tried to resist versions of the "no miracles argument" for realism.[5] Consider her claim:

> I think we should instead focus on the causal roles which the theory gives to these strange objects: exactly how are they supposed to bring about the effects which are attributed to them, and exactly how good is our evidence that they do so? The general success of the theory at producing accurate predictions, or at unifying what before had been disparate, is of no help here. (Cartwright 1983: 8)

The last sentence of this quotation is, to say the least, overstated. But let's not worry about this now. For, in her current views, the general antitheory tone (Cartwright 1983) has been superseded by a more considered judgement about theories and truth. She concedes that 'the impressive empirical successes of our best physics theories may argue for the truth of these theories', but, as we have already seen, she denies that it argues 'for their universality' (Cartwright 1999: 4). In fact, her talk about 'different kinds of knowledge in a variety of different domains across a range of highly differentiated situations' implies that truth is in the vicinity. For knowledge without truth is an oxymoron. So, her objections to Inference to the Best Explanation do not try to challenge the very possibility of a link between explanation and truth. Rather, they aim to block gross and global applications of IBE.

Let us look at Cartwright's argument for "local realism", which, as she says, is supposed to be a Kantian transcendental argument (Cartwright 1999: 23). The way she sets it up is this: We have *X*—'the possibility of planning, prediction, manipulation, control and policy setting'. But without Φ—'the objectivity of local knowledge'—*X* would be impossible or inconceivable. Hence Φ. It's fully understandable why Cartwright attempts to offer a transcendental argument. These arguments are dressed up as deductive. Hence, they are taken not to have a problematic logical form. They compare favourably with IBE. But apart from general worries about the nature and power of transcendental arguments[6], there is a more specific worry: Is the above argument really deductive?

A cursory look at it suggests that it is: "Φ is necessary for *X*; *X*; Therefore, Φ". But it is misleading to cast it as above, simply because it is misleading to say that Cartwright's Φ is necessary for *X*. Kant thought that Euclidean geometry was necessary for experience. Of course, it isn't. He could instead have argued that *some* form of spatial framework is necessary for experience. This might well be true. But now it no longer deductively follows that Euclidean geometry must be true. In a similar fashion, all that Cartwright's argument could show is that something—call it Φ—is necessary for 'the possibility of planning, prediction, manipulation, control and policy setting'. But now, it no longer follows deductively that this Φ *must* be the realist's "objective local knowledge", no matter how locally or thinly we interpret this. To say the least, this Φ could be just empirically adequate beliefs, or unrefuted beliefs, or beliefs that the world cooperates only when we actually try to set plans, make observations, manipulate causes, etc. Put in a different way, all that follows from Cartwright's transcendental argument is a disjunction: Either objective local knowledge, or empirically adequate beliefs, or . . . is necessary for the possibility of planning, prediction, manipulation, control, and policy setting. But which disjunct is the true one? Further argument is surely necessary. There cannot be a transcendental deduction of objective local knowledge.

My suggestion is that the move from the "the possibility of planning, prediction, manipulation, control and policy setting" to a realist understanding of what needs to be the case for all of them to be possible (or, why not, actual) can only be based on an inference to the best explanation: "The objectivity of local knowledge" (as opposed to any other candidate) should be accepted on the grounds that it best explains "the possibility of planning, prediction, manipulation, control, and policy setting". The moral then is that Cartwright's recent, more robust, realism can only be based on the very method that she has taken pains to disarm. We can now move on to look at the credentials of one her stronger early arguments against IBE, viz., the alleged falsity of laws.

False Laws?

One of Cartwright's main theses is that explanation and truth pull apart (Cartwright 1983). When laws come into the picture, this thesis seems to be the outcome of a certain failure of laws. She puts it like this:

> For the fundamental laws of physics do not describe true facts about reality. Rendered as descriptions of facts, they are false; amended to be true, they lose their fundamental, explanatory power. (Cartwright 1983: 54)

So, we are invited to see that if laws explain, they are not true, and if they are true, they do not explain. What Cartwright has in mind, of course, is what she calls fundamental or abstract laws as well as the covering-law model of explanation. If laws explain by "covering" the facts to be explained, then, Cartwright says, the explanation offered will be false. If, she would go on, the laws are amended by using several *ceteris paribus* clauses, they become truer but do not "cover" the facts anymore; hence, in either case, they do *not* explain the facts. The reason why covering laws do not explain has mostly to do with the fact that the actual phenomena are too complex to be covered by simple laws. Recall her example of a charged particle that moves under the influence of two forces: the force of gravity and Coulomb's force. Taken in isolation, neither of the two laws (i.e. Newton's inverse-square law and Coulomb's law) can describe the actual motion of the charged particle. From this, Cartwright concludes that each loses either its truth or its explanatory power. Here is her argument:

> The effect that occurs is not an effect dictated by any one of the two laws separately. In order to be true in the composite case, the law must describe one effect (the effect that actually happens); but to be explanatory, it must describe another. There is a trade-off here between truth and explanatory power. (Cartwright 1983: 59)

I fail to see how all this follows. For one, it does *not* follow that there is not (worse, there cannot be) a complex law that governs the motion of massive *and* charged particles. If we keep our eyes not on epistemology (can this law be known or stated?) but on metaphysics (can there be such a law?), the above argument is, to say the least, inconclusive. For another, in the composite case, there is no formal tension between truth and explanation. In the composite case, none of the two laws (Newton's and Coulomb's) is strictly true of, in the sense of "covering", the effect that actually happens. Why should we expect each of them on its own to "cover" the complex effect? After all, the complex phenomenon is governed by both of them jointly, and hence it cannot be covered by each of them separately. This does not imply that laws lose their explanatory power. They still explain how the particle would behave if it was just massive and not charged or if it was charged but not massive. And they still contribute to the full explanation of the complex effect (that is, of the motion of the charged and massive particle). To demand of each of them to be explanatory in the sense that each of them should somehow cover the actual complex effect is to demand of them something they cannot do. The laws do not thereby cease to be true, nor explanatory. Nor does it follow that they don't jointly govern the complex effect. Governing should not be conflated with covering.[7]

My argument so far might be inconclusive. So I want to suggest that there is an important independent reason why we should take laws seriously. Laws individuate properties: Properties are what they are because of the laws they participate in. Cartwright says:

> What I invoke in completing such an explanation are not fundamental laws of nature, but rather properties of electrons and positrons, and highly complex, highly specific claims about just what behaviour they lead to in just this situation. (Cartwright 1983: 92)

If it is the case that no laws then no properties, or if properties and laws are so intertwined that one cannot specify the former without the latter, then some laws had better be true. For if they are not, then we cannot talk of properties either.[8]

This last point, however, is controversial, especially as of late. It relies on a Humean understanding of properties. And Cartwright is a non-Humean, more or less about everything. This observation is crucial. For it is Humeanism that is Cartwright's real opponent. Her capacities are non-Humean tendencies: causal powers. That is, they are irreducible, primary and causally active constituents of the world. Similarly, her properties are non-Humean properties: They are active causal agents, which are identified via their causal role and their powers. So it is not laws that determine what they are; rather, it is properties (capacities, etc.) that determine what, if any, laws hold in the world. With all this in mind, let us turn our attention to her views about capacities. This is just one of her non-Humean themes. But it is perhaps the most central one.

CAPACITIES

Cartwright has devoted two books in the defence of the claim that capacities are prior to laws (Cartwright 1989; 1999). As is well known, she challenges the Humean view that laws are exceptionless regularities, since, she says, there are no such things.[9] How then does it appear that there *are* regularities in nature, e.g., that all planets move in ellipses?

Nomological Machines

Cartwright does not deny that there can be regular behaviour in nature. But she claims that where there is regular behaviour in nature, there is a nomological machine that makes it possible. A "nomological machine" is

> a fixed (enough) arrangement of components, or factors, with stable (enough) capacities that in the right sort of stable (enough) environment will, with repeated operation, give rise to the kind of regular behaviour that we represent in our scientific laws. (Cartwright 1999: 50)

Nomological machines make sure that "all other things are equal". So, they secure the absence of factors, which, were they present, would block the manifestation of a regularity. Take Kepler's law that all planets move in ellipses. This is not a strictly universal and unconditional law. Planets do (approximately) describe ellipses, if we neglect the gravitational pull that is exerted upon them by the other planets, as well as by other bodies in the universe. So, the proper formulation of the law, Cartwright argues, is: *ceteris paribus*, all planets move in ellipses. Now, suppose that the planetary system is a stable enough nomological machine. Suppose, in particular, that as a matter of fact, the planetary system is (for all practical purposes) shielded: It is sufficiently isolated from other bodies in the universe, and the pull that the planets exert on each other is negligible. Under these circumstances, we can leave behind the *ceteris paribus* clause and simply say that all planets move in ellipses. But the regularity holds only so long as the nomological machine backs it up. If the nomological machine were to fail, so would the regularity. As Cartwright has put it: '(L)aws of nature (in this necessary regular association sense of "law") hold only *ceteris paribus*—they hold only relative to the successful repeated operation of a nomological machine' (Cartwright 1999: 49–50).

Nomological machines might occur naturally in nature. The planetary system is such a natural nomological machine. But, according to Cartwright, this is exceptional. As she says: 'more often [the nomological machines] are engineered by us, as in a laboratory experiment' (Cartwright 1999: 49). 'In any case', she adds, 'it takes what I call a nomological machine to get a law of nature' (Cartwright 1999: 49).

For the operation of a nomological machine, it is not enough to have a stable (and shielded) arrangement of components in place. It is not enough, for instance, to have the sun, the planets, and the gravitational force in place in order for the planetary machine to run. Cartwright insists that it is the *capacities* that the components of the machine have that generate regular behaviour. For instance, 'a force has the capacity to change the state of motion of a massive body' (Cartwright 1999: 51). Couldn't the nomological machine itself be taken to be a regularity? No, she answers: 'the point is that the fundamental facts about nature that ensure that regularities can obtain are not again themselves regularities. They are facts about what things can do' (Cartwright 1995: 4). But what exactly are capacities, i.e., the things that things can do?

Cartwright focused her attention on 'what capacities do and why we need them' and *not* on 'what capacities are' (Cartwright 1989: 9). What they are is the job of her *The Dappled World*. Before, however, we examine what they are, let us see the main argument she offers as to why we need capacities.

Why Do We Need Capacities?

The Sellarsian Argument

Sellars's master argument for commitment to the unobservable entities posited by scientific theories is that they play an ineliminable explanatory role (Sellars 1963). In order to formulate it, he had to resist what he aptly called the 'picture of the levels'. According to this picture, the realm of facts is layered. There is the bottom level of observable entities. Then, there is an intermediate (observational) level of empirical generalisations about observable entities. And finally, there is yet another (higher-theoretical) level: unobservable entities and laws about them. It is part of this picture that while the observational framework is explanatory of observable entities, the theoretical framework enters the picture by explaining the inductively established generalisations of the observational framework. But then, Sellars says, an empiricist can protest that the higher level is dispensable. He may argue that all the explanatory work vis-à-vis the bottom level is done by the observational framework and its inductive generalisations. Why then, he may wonder, posit a higher level in the first place?

Sellars's diagnosis is that this picture rests on a myth. His argument against the myth of the levels is that the unobservables posited by a theory explain directly why (the individual) observable entities behave the way they do and obey the empirical laws they do (to the extent that they do obey such laws). So, he resists the idea that the theoretical framework has as its prime function to explain the empirical generalisations of the observational framework. Sellars claimed that unobservable entities are indispensable because they also explain why observational generalisations are, occasionally, violated; why, that is, some observable entities do not behave they way they should, had their behaviour been governed by the observational generalisation.

This is a fine argument and I endorse it fully (Psillos 2004a). Cartwright offers an argument structurally similar to Sellars's in defence of capacities (Cartwright 1989: 163). She has in mind another possible layer cake. The bottom level is the nonmodal level of occurrent regularities; the intermediate level is the level of Humean laws (either deterministic or statistical). The higher level is supposed to be a *sui generis* causal one. *This* layer cake, Cartwright notes, also invites the thought (or the temptation) to do away with the higher level altogether. All the explanatory work, it might be said, is done by Humean laws, endowed with modal force. The higher (causal) level could then be just seen as a higher modal level, with no claim to independent existence: It is just a way to talk about the intermediate level, and in particular a way to set constraints on laws in order to ensure that they have the required modal force. It is *this* layer cake that Cartwright wants to resist. For her, the higher causal level is indispensable for the explanation of what regularities there are (if any) in the world. So we seem to have a solid Sellarsian argument for capacities. But do we?

Capacities and Regularities

Before we proceed to examine this, an exegetical point is in order. Cartwright splits the higher (causal) level into two sublevels: a lower sublevel of *causal laws* and a higher sublevel of ascriptions of capacity. She couches all this in terms of two levels of generality or more accurately of two levels of modality (Cartwright 1989: 142). She says:

> (. . .) the concept of general *sui generis* causal truths—general causal truths not reducible to associations—separates naturally into two distinct concepts, one at a far higher level of generality than the other: at the lower level we have the concept of a causal law; at the higher, the concept of capacity. I speak of two levels of generality, but it would be more accurate to speak of levels of modality, and for all the conventional reasons: the claims at both levels are supposed to be universal in space and through time, they support counterfactuals, license inferences, and so forth. (Cartwright 1989: 142)

Why do we need *two* causal levels? Why, in particular, do we need a level of capacities? To cut a long story short, Cartwright thinks that causal laws are kinds of causal generalisations relative to a particular population

(Cartwright 1989: 144). They are causal, as opposed to Humean laws of association, mostly because, as Cartwright argues, the facts they report (e.g., that aspirins relieve headaches or that smoking causes cancer) cannot be fully captured by probabilistic relations among magnitudes or properties. Causal information is also required to specify the conditions under which they hold. A further thought then is that ascription of capacities is also necessary in order to remove the relativised-to-a-population character of causal laws. We don't just say that smoking causes cancer to population *X*. We also want to say that smoking causes cancer, *simpliciter*. This claim (which is universal in character) is best seen as a claim about capacities: *C* causes *E* means *C* carries the capacity *Q* to produce *E* (Cartwright 1989: 145). Capacities, then, are introduced to explain causal laws and to render them universal in character.[10] This last point is crucial: Causal laws are *ceteris paribus*. After all, it's not invariably the case that aspirin relieves headache. But capacities remove the *ceteris paribus* clause: Aspirin *always* carries the capacity to relieve headache. Capacities, we are told, are *stable*. If something has the capacity *Q*, then it carries it with it from one situation to another (Cartwright 1989: 145).

What then of Cartwright's Sellarsian argument for capacities? I focus on just one central problem. Sellars saves the higher level of electrons, protons, etc. by focusing on the indispensable role this level plays in the explanation of singular observable phenomena or things. Similarly, one would demand of Cartwright's argument to show how capacities are indispensable for the explanation of occurrent regularities, without the intervening framework of Humean laws plus modal force. But it seems that there is a tension in her argument. Whereas in Sellars's case, the entities of the theoretical framework (unobservables) can be identified independently of the entities in the bottom framework, it is debatable that this can happen in Cartwright's case. Here there are conflicting intuitions. One is that we need regularities (or Humean laws) to identify what capacities things carry. Another (Cartwright's, I think) is that this is not the case. I am not entirely certain whose intuitions are right. But it seems to me that the Humean is on a better footing. Capacities might well be posited, but only after there has been a regular association between relevant event types. No one would mind ascribing to aspirin the capacity to relieve headaches, if that was the product (as indeed it is) of a regular association between taking aspirins and headaches going away. "Regular" here does not necessarily mean exceptionless. But, so much the better for positing capacities if the association happens to be exceptionless. To say the least, one could more easily explain how capacities have modal force. So, there is an important disanalogy between Sellars's argument for unobservables and Cartwright's argument for capacities, which casts doubt on the indispensability of positing capacities. That is, in Cartwright's case, we need the lower level (regularities) to identify the entities of the higher level (capacities).

Single Cases

Cartwright insists that capacities might reveal themselves only occasionally or only in a single case. Consider what she says:

> "Aspirins relieve headaches". This does not say that aspirins always relieve headaches, or always do so if the rest of the world is arranged in a particularly felicitous way, or that they relieve headaches most of the time, or more often than not. Rather it says that aspirins have the capacity to relieve headaches, a relatively enduring and stable capacity that they carry with them from situation to situation; a capacity which may if circumstances are right reveal itself by producing a regularity, but which is just as surely seen in one *good* single case. The best sign that aspirins can relieve headaches is that on occasion some of them do. (Cartwright 1989: 3, emphasis added)

This is surely puzzling. Just adding the adjective "good" before the "single case" does not help much. A "good" controlled experiment might persuade the scientist that he has probably identified some causal agent. But surely, commitment to it follows only if the causal agent has a regular behaviour that can be probed in similar experiments. A single finding is no more compelling than a single sighting of a UFO. Single or occasional manifestations cast doubt on the claim that there is a stable and enduring capacity at play (Glennan 1997: 607–608).

Cartwright disagrees. She advances what she calls the "analytic method" in virtue of which capacity ascriptions are established (Cartwright 1999) and later summarises her ideas thus:

> We commonly use the analytic method in science. We perform an experiment in "ideal" conditions, *I*, to uncover the "natural" effect *E* of some quantity, *Q*. We then suppose that *Q* will in some sense "tend" or "try" to produce the same effect in other very different kinds of circumstances. (. . .) This procedure is not justified by the regularity law we establish in the experiment, namely 'In *I*, *Q* → *E*'; rather, to adopt the procedure is to commit oneself to the claim "*Q* has the capacity to *E*". (Cartwright 2002: 435–436)

What is the force of this claim? Note, first, that we don't have a clear idea of what it means to say that *Q* "tends" or "tries" to produce its effects. It seems that either *Q* does produce its effect or it doesn't (if, say, other factors intervene). Second, as Teller notes, it is not clear how the "trying" can be established by looking at a single case only (Teller 2002: 718). One thought here might be that if we have seen *Q* producing its effect at least one time, we can assume that it can produce it; and hence that it has the capacity to produce it. But I don't think this is the right way to view things. Consider the following three questions: (i) what exactly is *Q*'s effect? (ii) how can we know that it was *Q* which brought E about? and (iii) wouldn't it be rather trivial to say that for each effect there is some capacity *X* which produces it? All three questions would be (more easily) answered if we took capacities to be regularly manifested. The "regularity law", "in *I*, *Q* → *E*" makes the positing of a capacity legitimate. It is because (and insofar as) "in *I*, *Q* → *E*" holds that we can say that "*Q* has the capacity to *E*" and not the other way around.[11]

If the capacity *Q* of *x* to bring about *y* was manifested regularly, then one could say that the presence of the capacity could be tested. Hence, one could move on to legitimately attribute this capacity to *x*. But if a capacity can manifest itself in a single case, it is not clear at all how the presence of the capacity can be tested. Why, in other words, should we attribute to *x* the capacity to bring about *y*, instead of claiming that the occurrence of *y* was a matter of chance? So, there seems to be a tension between Cartwright's claim that capacities are manifestable even in single cases and her further claim that capacities are testable.[12]

So far, I have focused on the relation between capacities (the higher level) and regularities (the lower level). But there is also a problem concerning the two sublevels of the higher level, viz., capacities and causal laws.[13] Do claims about the presence of capacities have extra content over the claims made by ordinary causal laws? So, do we really need to posit capacities? Take, for instance, the ordinary causal law that aspirin relieves headaches. If we ascribed to aspirin a capacity to relieve headaches, would we gain in content? There is a sense in which we would. Ordinary causal laws are *ceteris paribus*, whereas capacity claims are not. Because it is only under certain circumstances that aspirin relieves headaches, it is only *ceteris paribus* true that aspirin causes headache relief. But, Cartwright might say, once it is established that aspirin carries the capacity to relieve headaches, the *ceteris paribus* clause is removed: The capacity is always there, even if there may be contravening factors that block, on occasion, its manifestation. The problem with this attempt to introduce capacities is that the strictly universal character of claims about capacities cannot be established. If it is allowed that claims about the presence of capacities might be based on single manifestations, it is not quite clear what kind of inference is involved in the movement from a single manifestation to the presence of the capacity. Surely, it cannot be an inference based on any kind of ordinary inductive argument.[14] If, on the other hand, it is said that claims about capacities are established by ordinary inductive methods, based on several manifestations of the relevant capacity, then all that can be established is a *ceteris paribus* law. Based on cases of uses of aspirin, all that it can be established is that *ceteris paribus*, aspirin relieves headaches. So, it is questionable that talk about capacities has extra content over talk about ordinary causal laws.

Cartwright could argue that claims about capacities are strictly universal in the sense that objects have capacities even if they completely fail to manifest them (Cartwright 2002: 427–428). However, she would then seem to compromise her view that capacities are measurable and testable. There is a deep, if common, reason why we should be wary of unmanifestable capacities: There could be just too many of them, even contradictory ones. Couldn't we just say of any false generalisation (e.g., that bodies rise if they are left unsupported) that the bodies referred to in it have the relevant capacity, though it is never manifested? And couldn't we say that an object carries at the same time the stable capacity to rise if left unsupported and the stable capacity to fall if left unsupported, but that the former is unmanifestable? In other words, what distinguishes between unmanifestable capacities and nonexistent ones?

Moral: if Cartwright insists on single manifestation of capacities, she faces a sticky trilemma. Either talk of capacities does not have extra content over talk in terms of ordinary causal laws; or there is a mysterious method that takes us from a single manifestation to the capacity; or there are unmanifestable capacities. All three options have unpalatable consequences.

Capacities and Interactions

To be fair to Cartwright, she has offered other reasons for commitment to capacities. One of them is that capacities can explain causal interaction. She says that 'causal interactions are interactions of causal capacities, and they cannot be picked out unless capacities themselves are recognised' (Cartwright 1989: 164).

There are cases that fit this model. A venomous snake bites me, and I take an antidote. The venom in my bloodstream has the capacity to kill me, but I don't die because the antidote has the capacity to neutralise the venom. That's a case of causal interaction, where one capacity blocks another. I am not sure this commits us to *sui generis* capacities, as opposed to whatever chemical properties the venom and the antidote have, and a law that connects these properties. But let's not worry about this. For there is a more pressing problem.

Suppose that I take an aspirin while I am still hearing the continuous and desperate screaming of my daughter, who suffers from colic. The aspirin has the capacity to relieve my headache, but the headache does not go away. It persists undiminished. How shall I explain this? Shall I say that this is because the screaming of my daughter has the capacity to cause aspirin-resistant headaches? This would be overly ad hoc. Shall I say that this is because the screaming of my daughter has the capacity to neutralise the capacity of aspirin to relieve headache? This would be very mysterious. Something has indeed happened: There has been an interaction of some sort which made aspirin not work. But why should I attribute this to a capacity of the screaming? If I did that, I would have to attribute to the screaming a number of capacities: the capacity to-let-aspirin-work-if-it-is-mild, the capacity to let-aspirin-work-if-it-is-*not*-mild-but-I-go-away-and-let-my-wife-deal-with-my-daughter, the capacity to block-aspirins'-work-if-it-is-extreme, etc. This is not exactly an argument against the role of capacities in causal interaction (though it might show that there can be causal interaction without reference to capacities). Still, it seems a genuine worry: When trying to account for causal interaction, where do we stop positing capacities and what kinds should we posit?

Cartwright challenges the sceptic about capacities with the following: 'the attempt to "modalise away" the capacities requires some independent characterisation of interactions; and there is no general non-circular account available to do the job' (Cartwright 1989: 164). If we could not characterise interactions without reference to capacities, we had better accept them. But why not follow, for instance, Salmon (1997) or Dowe (2000) in their thoughts that interactions are explained in terms of exchanges of conserved quantities? There is no compelling reason to take *them* to be capacities. We could; (Cartwright, for instance, takes charge to be a capacity). But then again we couldn't. Charge might well be a property (an occurrent property, that is) in virtue of which, and given certain laws, a particle that instantiates it behaves the ways it does.[15]

What Are Capacities?

Suppose that we do need to posit capacities. What exactly is the thing we need to posit? Cartwright is certainly in need of a more detailed account of how capacities are individuated. She tells us that capacities are *of* properties and not *of* individuals: 'the property of being an aspirin carries with it the capacity to cure headaches' (Cartwright 1989: 141). But aspirin is not, strictly speaking, a property. It's something that has a property. And certainly it does not carry its capacity to relieve headaches in the same way in which it carries its shape or colour.

It would be more accurate to say that capacities are properties of properties. That is, that they are second-order properties. But this would create some interesting problems. It would open the way for someone to argue that capacities are functional (or causal) roles. This would imply that there must be occupants of these causal roles, which are not themselves capacities. They could be the properties (maybe many and variable) that occupy this causal role. So, the capacity to relieve pain would be a causal role filled (or realised) by different properties (e.g., the chemical structure of paracetamol or whatever else). If, however, we take capacities to be causal roles, it would be open for someone to argue, along the lines of Prior, Pargeter, and Jackson (Prior et al. 1982) that capacities are causally impotent. The argument is simple. Capacities are distinct from their causal bases (as they are properties *of* them). They must have a causal basis (a realiser) because they are second-order. This causal basis (some properties) are themselves a sufficient set of properties for the causal explanation of the manifestation of the capacity (whenever it is manifested). Hence, the capacity *qua* distinct (second-order) property is causally impotent.

Cartwright wouldn't be willing to accept this conclusion. But then capacities must be *of* properties (or be carried by properties) in a different way. What exactly this way is it is not clear. She asks: 'Does this mean that there are not one but two properties, with the capacity sitting on the shoulder of the property which carries it?' And she answers: 'Surely not' (Cartwright 1989: 9). But no clear picture emerges as to what this relation of "*a* carrying *b*" is. (And is this "carrying" another capacity, as in *a* has the capacity to carry *b*? And if so, isn't there a regress in the offing?) At a different place, we are told that capacities have powers, which they can retain or lose (in causal interactions; Cartwright 1989: 163). Is that then a third-order property? A property (power) of a property (capacity) of a property (aspirin)? I don't think Cartwright wants to argue this. But what does she want to argue?

Cartwright later returns to these issues (Cartwright 1999). Here it seems that another possibility is canvassed, viz., that properties themselves are capacities. It's not clear whether she takes all properties to be capacities, but it seems that she takes at least some to be. We are given examples such as force and charge. I am not sure I have this right, but it seems to follow from expressions such as: 'Coulomb's law describes a capacity that a body has qua charged' (Cartwright 1999: 53). It also seems to follow from considering concepts such as 'attraction, repulsion, resistance, pressure, stress, and so on' as concepts referring to capacities (Cartwright 1999: 66). In any case, it seems that she aligns herself with Shoemaker's view of properties as "conglomerates of powers" (see Cartwright 1999: 70). Capacities then seem to come more or less for free: 'Any world with the same properties as ours would ipso facto have capacities in it, since what a property empowers an object to do is part of what it is to be that property' (Cartwright 1999: 70). So, it seems that Cartwright adopts a causal theory of properties, where properties themselves are causal powers.

Capacities and Laws

A number of questions crop up at this point. First, are all powers with which a property empowers an object constitutive of this property? And if not, how are we to draw a distinction between constitutive powers and nonconstitutive ones? For instance, is the causal power of aspirin to relieve headache on a par with its causal power to produce a pleasing white image? This is not a rhetorical question. For it seems that in order to distinguish these two powers in terms of their causal relevance to something being an aspirin, we need to differentiate between those powers that are causally relevant to a certain effect, e.g., relieving pain, and those powers that are not. Then, we seem to run in the following circle. We need to specify what powers are causally relevant to something being *P*. For this, we need to distinguish the effects which are brought about by *P* in two sorts: those that are the products of causally relevant powers and those that are not. But in order to do this we need first to specify what it *is* for something to be *P*.[16] That is, we need to specify what powers are causally relevant to *P*'s identity and what are not. Ergo, we come back to where we started. (Recall that on the account presently discussed causal powers are the *only* vehicle to specify *P*'s identity).

Second question: why is it the case that some causal powers are held together, and others are not? Why, that is, do certain powers have a certain kind of "causal unity", as Shoemaker (1980: 125) put it? This is a crucial question because even if every property is a cluster of powers, the converse does not hold. Electrons come with the power to attract positively charged particles and the power to resist acceleration, but they don't come with the power to be circular. And the power of a knife to cut wood does not come with the power to be made of paper. This is important because, as Shoemaker himself observes, the concurrence of certain powers might well be the consequence of a law (1980: 125). So, it might well be that laws hold some capacities together. Hence, it seems that we cannot just do with capacities. We also need laws as our building blocks. This issue has a ramification. Why is it the case that nothing has the power to move faster than light? The absence of a certain capacity might also be the consequence of a natural law.

Third question: should we be egalitarian about capacities? Is the capacity to resist acceleration on a par with the capacity to become grandparent? Or with the capacity to be a table-owned-by-George-Washington? This question is different from the first. It relates to what in the literature is called the difference between genuine changes and mere Cambridge changes. The parallel here would be a difference between genuine capacities (properties) and mere Cambridge capacities (properties). Here again, laws are in the offing. For it can be argued that genuine capacities (properties) are those that feature in laws of nature.

I offer these questions as challenges. But they do seem to point to a certain double conclusion. On the one hand, we need to be told more about what capacities are before we start thinking seriously that we should be committed to them. On the other hand, we seem to require laws as well as capacities, even if we accept capacities as building blocks.

Cartwright wants to further advance the view that capacities are metaphysically prior to laws. She says, 'It is capacities that are basic, and laws of nature obtain—to the extent that they do obtain—on account of the capacities' (Cartwright 1999: 49). She offers no formal treatment of the issue how capacities relate to laws. Instead, we are given some examples.

> I say that Newton's and Coulomb's principles describe the capacities to be moved and to produce a motion that a charged particle has, in the first case the capacity it has on account of its gravitational mass and in the second, on account of its charge. (Cartwright 1999: 65)

If laws describe what the entities governed by them can do on account of their capacities, these capacities should be individuated, and ascribed, to entities, independently of the law-like behaviour of the latter. But, as noted above, it is not clear that this can be done. It seems that far from being independent of laws, the property of, say, charge is posited and individuated by reference to the law-like behaviour of certain types of objects: Some attract each other, and others repel each other in a regular fashion. The former are said to have opposite charges, whereas the latter have a similar charge. Cartwright says: 'The capacity is associated with a single feature—charge—which can be ascribed to a body for a variety of reasons independent of its display of the capacity described in the related law' (Cartwright 1999, 54–55).

This may well be true. But it does not follow that the capacity is grounded in no laws at all. Cartwright disagrees. She argues that '[c]apacity claims, about charge, say, are made true by facts about what it is in the nature of an object to do by virtue of being charged' (Cartwright 1999: 72).

Then, one would expect an informative account of what it is in the nature of an object to do. Specifically, one would expect that the nature of an object would determine its capacities, and would delineate what this object can and cannot do. But she goes on to say: 'There is no fact of the matter about what a system can do just by virtue of having a given capacity. What it does depends on its setting . . . (Cartwright 1999: 73).

Why, then, should we bother to attribute capacities? We could just offer an open-ended list of the things that a system does when it is placed in several settings. If, at least, there was a fact of the matter as to what a system could do by virtue of having a given capacity, the capacity could be used to (a) predict what a system can or cannot do and (b) explain why it behaves the way it does. In fact, if Cartwright really means to uphold the strong view that there is no fact of the matter as to what a system can do by having a certain capacity, then the very possibility of prediction and of explanation is threatened. For any kind of behaviour would be compatible with the system's having a certain capacity. No specific behaviour could be predicted, and any kind of behaviour could be explained (by an appeal to context-specific impediments of the system's capacities).[17]

One might object, however, that Cartwright's wording is very careful. It does not imply that there is no fact of the matter about what a system (or an object) can do by virtue of its nature. Yet, one would expect that if the nature of an object placed some substantive constraints on its capacities, there would be a fact of the matter about what this object can do by virtue of its capacities. For instance, one would expect that although a certain particle has the capacity to move, its nature constrains this capacity so that it cannot move with velocity greater than the velocity of light. As this example suggests, it might well be the case that the nature of an object is constrained by what laws it obeys.

In a previous draft of this paper, I tried to examine in some detail what these natures are and how they might relate to capacities. But Paul Teller directed my attention to the following passage, in which Cartwright says:

> My use of the terms *capacity* and *nature* are closely related. When we ascribe to a feature (like charge) a generic capacity (like the Coulomb capacity) by mentioning some canonical behaviour that systems with this capacity would display in ideal circumstances, then I say that that behaviour is *in the nature of* that feature. Most of my arguments about capacities could have been put in terms of natures. . . .
>
> (Cartwright 1999: 84–85)

So, it seems clear that Cartwright thinks there is no significant distinction between capacity and nature. But suppose that she followed many other friends of capacities and distinguished between capacities and natures. Fisk (1970) and Harré (1970), among others, think that an appeal to an entity's nature can explain why this entity has certain capacities. In particular, Harré (1970) argues that (a) discovering the nature of an entity is a matter of empirical investigation; but (b) specifying (or knowing) the exact nature of an entity is not necessary for grounding the ascription of a power to it. He links natures and capacities thus: 'There is a ϕ such that something has φ, and whatever had φ in C*, would have* to G, i.e. if something like α did not have φ in C it would not, indeed *could* not G (Harre 1970: 101).

The nature φ of an entity is thereby linked with its capacity to G. There are many problems with this proposal.[18] But I focus on one. What is it that makes the foregoing counterfactual true? It's not enough to have the circumstances C and the nature φ in order to get G. This is not just because G could be unmanifested. Even if we thought that the power to G were always manifested in circumstances C with a characteristic effect e, there would *still* be room for asking the question: What makes it the case that α's being φ in C makes it produce the characteristic effect e? We need, that is, something to relate (or connect) all these together, and the answer that springs to mind is that it is a *law* that does the trick.[19] This law might well be a brute (Humean) regularity.[20]

An advocate of natures could say that when the nature φ is present, there is no need to posit a law in order to explain why a certain object has a characteristic effect e when the circumstances are C. Yet this move would not really be explanatory. It would amount to taking natures to be collections of powers, and this hardly explains in an interesting way why a certain nature has the capacities it does: It just equates the nature of an object with a collection of its capacities.

A CONCLUDING REMARK

As we have seen, Cartwright has moved from a modest realist position (viz., realism about entities) to a superrealist position (viz., realism about powers and capacities). Part of her motivation for her early, restricted, realism was a certain antifundamentalism, viz., a resistance to the view that there are fundamental laws of nature, which determine what entities do, and which are captured (or should be captured) by scientific theories. It may be ironic that she now replaces this picture by another fundamentalism, viz., the view that capacities are the fundamental building blocks of the world, the things that make things to be what they are and to behave the way they do. Along the way, her early antitheory temper was softened. But her early antilaws temper was hardened.

In contemplating Cartwright's realist toil, we have learned a lot. But it seems that we are still short of a compelling reason to take capacities seriously as fundamental non-Humean constituents of the world. At any rate, even if we granted capacities, we would still need laws to (i) identify them; (ii) connect them with their manifestations; (iii) explain their stability; (iv) explain why some (but not others) occur together; (v) explain why some (but not others) obstruct the manifestation of others. It seems then that both the epistemology and the metaphysics of capacities require laws. Cartwright is to be commended for trying to make a case for the view that *capacities are enough for laws*. If the argument in the later part of this paper has been correct, then the situation is more complicated: *Laws and capacities are necessary for laws*.

NOTES

1. Earlier versions of this chapter were presented at the Workshop in honour of Nancy Cartwright, in Konstanz, December 2002, and in the University of California San Diego Philosophy Colloquium. I thank the participants of these events for many thoughtful comments and criticisms. I especially thank Nancy Cartwright for her comments and encouragement, as well as Craig Callender, Paul Churchland, Gerald Doppelt, Ron Giere, Stephan Hartmann, Carl Hoefer, and Wolfgang Spohn. Paul Teller deserves special mention for giving me many thoughtful written comments on the content as well as the structure of this chapter. Theodore Arabatzis, Steve Clarke, Robin Hendry, Christoff Schmidt-Petri, and David Spurrett must also be thanked for detailed written comments.
2. For more on the relation between scientific realism and metaphysical issues, see Psillos 2005.
3. For readers unfamiliar with these attempts, a brief statement of some major views follows. On Lewis's reading, *c* causally explains *e* if *c* is connected to *e* with a network of causal chains. For him, causal explanation consists in presenting portions of explanatory information captured by the causal network. On Woodward's reading, *c* causally explains *e* if *c* and *e* are connected by a relevant (interventionist) counterfactual of the form 'if *c* hadn't happened, then *e* wouldn't have happened either'. On Salmon's reading, *c* causally explains *e* if *c* is connected with *e* by a suitable continuous causal (i.e. capable of transmitting a mark) process. On the standard deductive-nomological reading of causal explanation, for *c* to causally explain *e*, *c* must be a nomologically sufficient condition for *e*. And for Mackie, for *c* to causally explain *e* there must be event-types *C* and *E* such that *C* is an inus-condition for *E*. For details on all these, see Psillos (2002a).

4. For a different take on the nature of inference to the most likely cause, see Suárez's contribution in this volume.
5. For more on this issue see Psillos (1999: Ch. 4).
6. To see what these worries might be, consider the difference between modest and ambitious transcendental arguments. Is Cartwright's intention to arrive at the modest conclusion that it is rational to believe that there is local knowledge or at the much more ambitious conclusion that *there is* local knowledge?
7. Spurrett defends a similar point in much more detail (Spurrett 2001).
8. A huge issue here concerns the nature of laws. I favour the Mill–Ramsey–Lewis approach, which I defend in some detail in Psillos (2002a: 148–154, 210–211). This approach can identify laws independently of their ability to support counterfactuals. However, it seems to require some prior notion of 'natural property'. But this notion need not equate properties with causal powers or capacities.
9. For an important survey of the debate around *ceteris paribus* laws, as well as a defence of strict laws in physics, see Earman& Roberts (1999). The interested reader should also see the special issue of *Erkenntnis* (2002, Vol. 57, no 3) on the status of *ceteris paribus* laws.
10. This point is also made vividly in Cartwright (2002).
11. Cartwright argues that capacities help explain what makes the design of a single experiment 'a good one': The design is good if it controls for all factors relevant to the effect (Cartwright 2002: 436). But why do we need an appeal to capacities to do this? In a clinical trial what Cartwright demands can be achieved by randomisation. In a physical experiment, in order to control for all factors relevant to the effect we need to appeal to regularities in the following sense: we need to control for all factors that regularly influence effects of this type. Strictly speaking, we cannot control for factors that do not fall under a regularity, since we don't have a clue as to what they might be. When, in an experiment, one does not control for the colour of the experimenter's eyes, it is because there is no regularity that connects the colour of eyes with the result of the experiment. Little (if anything) is gained if we add that the colour of the eyes does not have the *capacity* to alter the effect.
12. One might argue that there are clear cases in which a single case is enough to posit a capacity. An example put to me by Christoph Schmidt-Petri is the capacity to run fast: One case is supposed to prove its existence. I am not so sure this is true. What if I run fast (just once) because I took a certain steroid on a given occasion? Surely, in this case I don't have the capacity to run fast, though the steroid might have the (stable) capacity to make people run fast. But this latter capacity would need regular manifestation in order to be posited. For more criticism of Cartwright's argument that capacities are necessary in the methodology of science, see Giere (this volume).
13. A variant of this problem has been posed by Morrison (1995)
14. This point calls into question Cartwright's claim that capacities show how we can make sense of inductions from single experiments (Cartwright 1999: 90; 2002: 436). Undoubtedly, *if* stable capacities are in operation, then knowing them is enough to generalise from a single experiment. But how is the antecedent of the conditional grounded? It seems that we need regular behaviour (and hence plenty of inductive evidence) in order to posit stable capacities in the first place.
15. For a criticism of causal powers, see Psillos (2006).
16. Compare: something could be an aspirin without having the causal power to produce a white image; but something could *not* be an aspirin without having the power to relieve headaches.

17. A similar complaint is voiced by Earman & Roberts (1999: 456) and Teller (2002: 719).
18. See the criticisms of Fisk's views by Aune (1970) and McMullin (1970)
19. A similar point is made by Menzies (2002). Teller also notes that capacities might well be no different from the OK properties that Cartwright argues should figure in laws (Teller 2002: 720–721).
20. This is just one option, of course; see also Teller (2002: 722). Another option would be to look for a mechanism that connects the nature φ with its power to produce a characteristic effect in certain circumstances. I have a number of objections to mechanisms that I cannot repeat here (see Psillos 2004b). At any rate, it seems enough for the purposes of this chapter that it remains an open option that Humean regularities may get the capacities do whatever they do.

REFERENCES

Aune, B. (1970) 'Fisk on capacities and natures', *Boston Studies in the Philosophy of Science*, 8: 83–87.

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Clarendon Press.

———. (1989) *Nature's Capacities and Their Measurement*, Oxford: Clarendon Press.

———. (1995) 'Précis on *Nature's Capacities and Their Measurement*", *Philosophy and Phenomenological Research 55*: 153–6.

———. (1999) *The Dappled World*, Cambridge: Cambridge University Press.

———. (2002) 'In favour of laws that are not *ceteris paribus* after all', *Erkenntnis*, 57: 425–439.

Dowe, P. (2000) *Physical Causation*, Cambridge: Cambridge University Press.

Earman, J., and J. Roberts. (1999) '*Ceteris Paribus*, there is no problem of provisos', *Synthese*, 118: 439–478.

Fisk, M. (1970) 'Capacities and natures', *Boston Studies in the Philosophy of Science*, 8: 49–62.

Giere, R. (this volume) 'Models, metaphysics and methodology'.

Glennan, S. (1997) 'Capacities, universality, and singularity', *Philosophy of Science*, 64: 605–626.

Harré, R. (1970) 'Powers', *The British Journal for the Philosophy of Science*, 21: 81–101.

Hempel, C. G. (1965) *Aspects of Scientific Explanation*, New York: The Free Press.

Hitchcock, C. R. (1992) 'Causal explanation and scientific realism', *Erkenntnis*, 37: 151–178.

Lipton, P. (1991) *Inference to the Best Explanation*, London: Routledge.

Menzies, P. (2002) 'Capacities, natures and pluralism: A new metaphysics for science?' *Philosophical Books*, 43: 261–270.

McMullin, E. (1970) 'Capacities and natures: An exercise in ontology', *Boston Studies in the Philosophy of Science*, 8: 63–83.

Morrison, M. (1995) 'Capacities, tendencies and the problem of singular causes', *Philosophy and Phenomenological Research*, 55: 163–168.

Prior, E., R. Pargeter, and F. Jackson. (1982) 'Three theses about dispositions' *American Philosophical Quarterly*, 19: 251–256.

Psillos, S. (1999) *Scientific Realism: How Science Tracks Truth*, London: Routledge.

———. (2002a) *Causation and Explanation*, Chesham: Acumen.

———. (2002b) 'Simply the best: A case for abduction', in F. Sadri and A. Kakas (eds) *Computational Logic: From Logic Programming into the Future*, LNAI 2408, Berlin-Heidelberg: Springer-Verlag.

———. (2004a) ‘Tracking the real: Through thick and thin’, *The British Journal for the Philosophy of Science*, 55: 393–409.

———. (2004b) ‘A glimpse of the *Secret Connexion*: Harmonising mechanisms with counterfactuals’, *Perspectives on Science*, 12: 288–319.

———. (2005) ‘Scientific Realism and Metaphysics’, *Ratio*, 18, 385–404—*Ratio, 18*: 385–404.

———. (2006) ‘What do powers do when they are not manifested?’ *Philosophy and Phenomenological Research*, 72: 135–156—‘*Philosophy and Phenomenological Research*, *72*: 135–156.

Salmon, W. (1997) *Causality and Explanation*, Oxford: Oxford University Press.

Sellars, W. (1963) *Science, Perception and Reality*, Atascadero: Ridgeview P. C.

Suárez, M. (this volume) ‘How inference to the most probable cause might be sound’.

Shoemaker, S. (1980) ‘Causality and properties’, in P. van Inwagen (ed.) *Time and Change*, New York: Springer.

Spurret, D. (2001) ‘Cartwright on laws of composition’, *International Studies in the Philosophy of Science*, 15: 253–268.

Teller, P. (2002) ‘Review of *The Dappled World*’, *Nous*, 36: 699–725.

Reply to Stathis Psillos

Stathis Psillos demands a way to identify capacities. It seems we either need laws—'laws individuate properties; properties are what they are because of the laws they participate in'—or a set of behaviours that occur when the capacity is manifested (Psillos this volume: 15). But, he observes, I don't like laws, and I say that some capacities can be manifested in almost any behaviour. Neither of these claims is entirely accurate, however.

The law claims I don't like are those that report regular associations among occurrent properties. But there are other "laws" that I endorse wholeheartedly; for instance, "An object of mass m has a capacity of strength GMm/r^2 to attract a mass of size M a distance r away". This law ascribes a given capacity to a property that we have other ways to identify.[1] Or, "If an object of mass m manifests its capacity to attract an object of mass M a distance r away and nothing interferes, the second object will have an acceleration Gm/r^2." Notice that in this last case we also have a claim about what behaviour occurs when the capacity is manifested. I can thus mimic Psillos: A given capacity is what it is because of the laws it participates in. These laws often involve reference to other capacities, but that is no more an objection to the claim that the laws "individuate" the capacity than the fact that the laws that are supposed to individuate a property refer to other properties.

Some of Psillos's worries about identifying capacities by their manifestation rest on a conflation of the manifestation, or exercise, of the capacity with the occurrence of the canonical behaviour we associate with the capacity.[2] The gravitational capacity, for instance, seems always to manifest itself—a massive object always *attracts* another, yet the canonical behaviour—an acceleration towards that object—may seldom occur. And we know a host of tests that assure us that the manifestation obtains even when the acceleration does not. So the manifestation—"attracting"—is fixed even if the behaviour described in occurent-property language—"acceleration *Y*"—is highly various.[3]

Psillos also worries about prediction and explanation. True, for many capacities almost any behaviour can result from their exercise. But we can still predict because different behaviours result from the exercise of the same capacity in different circumstances. So long as there are rules about how capacities combine or how they respond to variations in circumstance, prediction will be possible.

Psillos share a different worry about testing with Margaret Morrison.[4] I claim that capacities can often be measured and very precisely. They are like forces in that respect. But that does not mean that we can tell by those measurements that what we measure is a capacity. Again, that is like forces. We can measure the acceleration of an object and its mass and multiply to measure the force on it. That does not tell us that there are forces in nature. To defend this, we need an extensive network of empirical, theoretical and philosophical considerations. So too with capacities.[5]

As to what capacities are, I do not object to the putatively untoward consequences of either alternative Psillos offers. Suppose for instance that "... is an interference with ...", "... has the capacity to ...", "... is a trigger for ...", etc., are second-order properties. What matters for capacities is the threefold distinction Hume denied between the obtaining of the capacity (e.g., the capacity to attract obtains whenever an object has a mass), the manifestation or exercise of the capacity (the attracting),[6] and the "occurent-property" behaviour (the motion of the attracted object). It does not matter if the second-order property is inert so long as we can maintain all three distinct features, for instance by admitting exercisings or manifestations as first-order properties—thus allowing first-order properties that are not picked out by what we class as occurrent—property terms.[7]

Like Mill, I recommend capacity talk wherever I find the analytic method in use. But unlike Mill, at least as Schmidt-Petri pictures him, I take this talk literally. The component features have capacities, the capacities are exercised, and the result of their joint operation is what happens. That is how the laws for the components—laws in my sense, ascribing capacities and describing their mode of operation, not laws in the regularity-among occurrent-properties sense—explain the result. What about Psillos? He tells us that the laws for the components "contribute to a full explanation" of what occurs when they operate together, also that these laws "govern" the complex effect without "covering" it. What then do these law claims say, and what sense is there to "govern" or even "explain" once both the covering-law story and the capacities story are rejected?

Alternatively Psillos suggests that complex laws could do the job. There would then have to be an open-ended collection of these laws, enough to cover every arrangement of contributing causes that ever occurs. The notion of regularities here is certainly strained; and if not regularities, what are the truth-makers for these laws? Besides that, I would still argue, as in *The Dappled World*, that even these need a *ceteris paribus* clause in front—"only so long as nothing interferes", where interference is a robust capacity concept.

NOTES

[1] Or ascribes it to any object in the right circumstances instancing that property. Which way one puts it depends on how one wants to understand the metaphysics of capacities.

[2] Note that my usage of these terms here differs from that of Psillos, who seems generally to use 'manifestation' to refer to what I call resultant behaviour.

[3] As I note in *The Dappled World*, sometimes we do not have a nice summarizing word such as "attracts" for the manifestation or exercise of the capacity; hence the resort to "tries to X" where X is a canonical behaviour associated with the capacity.

[4] Early Morrison paper

[5] It was thus, as Psillos points out, a gross exaggeration on my part to say that the best evidence that one feature can cause another is that it does so, in the capacity sense of 'can'. This is good evidence only once we suppose (as in the Gravity Probe) that whatever the cause produces it does out of a stable capacity.

[6] In this case it seems the manifestation occurs whenever the mass does. But that is not necessary—some capacities need triggering or manifest themselves only in special circumstances.

[7] Nor do I object to the existence of "laws" that demand that different specific capacities occur together. On the other hand, I certainly would not admit them in order to explain why they occur together as I don't see why that needs explanation.

9 Invariance, Modularity, and All That

Cartwright on Causation

James Woodward

INTRODUCTION

I think that the first paper of Nancy Cartwright's that I ever read was 'Causal Laws and Effective Strategies' (Cartwright 1979, [1983]) shortly after it appeared nearly twenty-five years ago. At the time I was still in a deep dogmatic Humean slumber about causation, and my first reaction was that the central claims of the paper about the irreducibility of causal laws to what Cartwright called laws of association couldn't possibly be right. But, like a number of other philosophers of science at about this time, I had independently decided to try to learn something about the so-called causal modeling techniques widely used in the social and biomedical sciences. I assumed that these techniques would have something interesting—at that point I didn't know what—to teach philosophers about causation and its relation to probability. As I worked through this literature, and began to appreciate the causal character of the additional assumptions that were required to get causal conclusions out of statistics, it gradually dawned on me—I was a slow study—that Cartwright was absolutely right about the issue of irreducibility. 'Causal Laws and Effective Strategies' was a brilliantly original paper that fundamentally changed the thinking of many of us about causation.

In the intervening decades, I've spent a lot of time thinking about Cartwright's ideas about causation. Even when I haven't been fully persuaded, I've always found them penetrating and insightful. In the remarks that follow, I want to survey some aspects of on an ongoing discussion that she, I, and others, including my sometime coauthor, Dan Hausman, are having on a number of interrelated themes having to do with the interpretation of systems of structural equations, the connection between causation, invariance, and interventions, the status of a condition called modularity, and the merits of accounts that emphasize the diversity of causal relations (Cartwright 2001, 2002, 2003; Hausman & Woodward 1999, forthcoming a, b). My emphasis is on understanding some of the principal points of disagreement, rather than vindicating my own take on things, although I won't be shy about trying to defend myself, where appropriate.

STRUCTURAL EQUATIONS

My aim in this section is to briefly set out some of the main elements of the interventionist interpretation of systems of structural equations that I endorse, with an eye to comparing this interpretation with Cartwright's views in subsequent sections. I claim no particular originality for this interpretation. It has its roots in a tradition in econometrics that goes back to writers such as Haavelmo (1944); Frisch (1938); Strotz and Wold (1960) and has recently been revived by Judea Pearl (2000), among others. Readers who are familiar with these ideas may wish to skip this section.

Consider the following set up which is taken from Hausman (1998). A salt solution flows through pipes and mixing chambers. The concentration of salt in each chamber, measured by the variables $X_1 \ldots X_4$, depends on the upstream chambers to which it is connected, according to the following equations:

2.1 $$X_2 = aX_1$$

2.2 $$X_3 = bX_1$$

2.3 $$X_4 = cX_2 + dX_3$$

Associated with 2.1–2.3 is the graphical structure described in Figure 9.1.

The interpretive convention is that each equation is to be understood as listing on its right hand side all and only the direct causes of the variable on the left-hand side. The complete system of equations is intended to describe all of the direct causal relationships among the variables in the system of interest. The question we want to answer is: Under what conditions will these equations be "causally correct" in the sense that they correctly represent the full set of causal relationships in the system under study, given the convention just described? In other words, what do these equations mean when we interpret them, according to the convention just described, as representing causal relationships? The need for such an interpretation arises because different researchers in the social sciences seem to mean different things, or perhaps nothing very clear at all, by the term *causal relationship*. For this reason, some further specification of what is meant by "cause" is required. The need for such a further specification and the possibility of providing it are, I think, among the issues that Cartwright and I disagree about. As becomes apparent below, Cartwright is a supporter of "causal diversity":

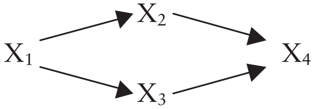

Figure 9.1

She is attracted to the view that causal locutions have a variety of different meanings and may not share any single common content, whether this is framed in manipulationist terms, as I advocate, or anything else.

The contrary idea that I want to defend is that the above equations, as well as causal claims more generally, do have a common content: They should be understood as counterfactual claims about what would happen under hypothetical idealized experimental manipulations—manipulations that I will call interventions, following what has become standard practice in the literature (Spirtes et al. 1993, 2000; Pearl 2000). On this construal the equations will be causally correct if and only if they correctly describe what will happen under some range of interventions—more precisely iff each equation correctly describes what will happen to the value of the variable on its left-hand side under interventions on its right-hand-side variable. Thus, for example, if we were to carry out an intervention on X_2, changing it by one unit, the equation (2.3) will be causally correct iff X_4 will in fact change in just the way described by (2.3)—that is by amount c units.

INTERVENTIONS

How should the notion of an intervention be characterized? There are a number of different characterizations in the literature, including several due to Cartwright herself (Cartwright & Jones 1991; Cartwright 2003; Spirtes et al.1993, 2000; Pearl, 2000). Which characterization is most appropriate depends on our purposes. One use to which we may wish to put this notion is calculational: We may be interested in calculating the effects of various "manipulations", given observations on an unmanipulated system and a certain understanding of what a manipulation involves. This is the purpose that motivates the characterization of an "atomic intervention" in Pearl (2000). As Pearl explains, what is crucial for this purpose is that the manipulation not alter other causal relationships in the system of interest besides the relationship in which the variable intervened on occurs as an effect, or that if the manipulation produces such alterations, it does so in known ways. As long as this condition is met it will be possible, given the right additional information, to predict the effects of such manipulations.

My purpose is different from Pearl's. As explained above, my project is to provide an account of the meaning or content of causal claims in terms of claims about the response of certain variable to interventions on others. A notion of intervention that is suitable for this purpose will need to have certain features that are different from those possessed by Pearl's notion of an atomic intervention. For example, if we are to use the notion of an intervention on X to help to explain what it is for X to cause Y, then we don't want to build into this notion information about the existence or nonexistence of a causal relationship between X and Y—a point to which I return in the section 'Cartwright on interventions' below. By contrast, building in this sort of information is unobjectionable, even desirable, if our purposes are, like Pearl's, calculational, rather than semantic, like mine.

With this consideration in mind, let us return to (2.1)–(2.3). Suppose that we change X_2 by amount one unit, but we do this by changing the value of X_1 by $1/a$ units. Since X_1 is also a cause of X_3, as equation (2.2) indicates, this change in X_1 will also (assuming (2.2) is correct) produce an increase of b/a units in X_3. This in turn will produce an additional change of $(b/a)d$ units in X_4, along the $X_1 \rightarrow X_3 \rightarrow X_4$ route. Thus the total change in X_4 that results from this change in X_2 will not be just c units but rather the sum of contribution that X_2 makes to X_4 (*a* change of c units) plus the contribution of X_1 to X_4 that goes through X_3. In other words, the total change in X_4 will be $c+(b/a)\bullet d$ units.

This illustrates one constraint it is reasonable to impose on the notion of intervention that we are looking for: We want an intervention on X_2 for the purpose of determining whether there is a causal relationship linking X_2 to X_4 (an intervention on X_2 with respect to X_4, as I will call it) to change X_2 in such a way that any change in X_4 will occur only as a result of the change in X_2 and not as a result of a change in some other variable that, like X_3, affects X_4 via some route or path that does not go through X_2. In other words, if we represent an intervention *I* on a variable *X* by means of a directed arrow from *I* to *X*, and if we represent the use to which we wish to put the intervention by means of a dashed arrow punctuated by a question mark from *X* to *Y*, indicating that we wish to learn whether *X* causes *Y*, then one possibility we want to rule out can be represented graphically in Figure 9.2.

The reason for this restriction is obvious: If *I* affects *Y* via a route that does not go through *X*—e.g., via *Z*—then *Y* may change under manipulation of *X* even though there is no causal relationship at all between *X* and *Y*. For similar reasons we also want to rule out structures like those in Figure 9.3.

Intuitively, the change in *X* produced by the intervention *I* should be *exogenous* (it should come from "outer space") in the sense that it changes only *X* and whatever lies on the causal route from *I* to *X* to *Y*. In the salt chamber example, this requirement might be implemented by, for example, building a new pipeline $I\rightarrow X_2$ directed into X_2 (in addition to the $X_1 \rightarrow X_2$ pipe), which operates independently of the $X_1 \rightarrow X_2$ pipe and which allows us to introduce an additional amount of solution into the chamber X_2 in

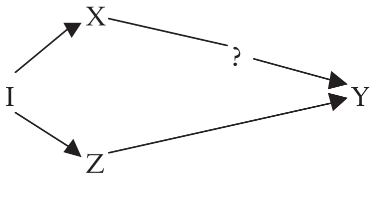

Figure 9.2

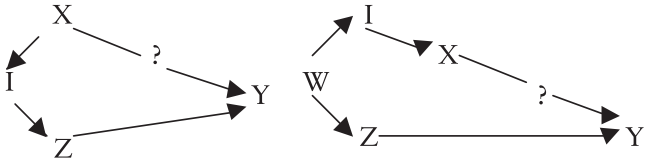

Figure 9.3

a way that is independent of whatever contribution X_1 makes to X_2. If we were to introduce an additional unit of solution into X_2 by means of such a pipe, then it arguably would be reasonable to expect that that if (2.3) is causally correct, the change in X_4 should be just what is indicated by (2.3)—namely *c* units.

I said that it would be reasonable to expect this, but in fact this expectation relies on another assumption that should be made explicit. This is the assumption that in introducing the new exogenous pipe into X_2, we do not alter certain causal relationships that hold elsewhere in the system. For example, if the process of introducing this new pipe into X_2 changes the relationship between X_1 and X_3 by breaking the pipe connecting these two chambers, then obviously even if the new pipe into X_2 leaves the level of solution in X_1 unchanged, the broken pipe between X_1 and X_3 will result in a change in the level of X_4, and this change will not reflect just the contribution of *c* units made by X_2 to X_4. Again a manipulation of X_2 that changes the causal relationships along some other route not involving X_2 but leading to X_4 will be nonideal from the point of view of discovering the causal relationship, if any, between X_2 and X_4.

If we collect these constraints together, we arrive at the following characterization of an intervention variable, which I will label **WIN** (for weak intervention) as it captures a notion of intervention that is weaker than the stronger notion described below.

Let *X* and *Y* be variables, with the different values of *X* and *Y* representing different and incompatible properties possessed by the unit *u*, the intent being to determine whether some intervention on *X* produces changes in *Y*. Then I is a *weak intervention* variable for *X* with respect to *Y* if and only if I meets the following conditions:

WIN:

I1. I causes *X*.

I2. Any directed path from *I* to *Y* goes through *X*. That is I does not directly cause *Y* and is not a cause of any causes of *Y* that are distinct from *X* except, of course, for those causes of *Y*, if any, that are built into the *I*–X–*Y* connection itself; that is, except for (a) any causes of *Y* that are effects of *X* (i.e. variables that are causally between *X* and *Y*) and (b) any causes of *Y* that are between *I* and *X* and have no effect on *Y* independently of *X*.

I3. *I* is probabilistically independent of any variable *Z* that is causally relevant to *Y* and that is on a directed path that does not go through *X*.

I4. *I* does not alter the relationship between *Y* and any of its causes *Z* that are not on any directed path (should such a path exist) from *X* to *Y*.

This characterization still suffers from an important limitation. One way of bringing this out is to observe that even if interventions in the sense just described are always possible, this does not by itself insure that we can exogenously set X_2 to any value that we like. Roughly speaking, if interventions in the sense of **WIN** are possible, it follows that we can produce certain *changes* in the value of X_2 (with respect to its previous value). However, the value of X_2 that results after an intervention will reflect not just the contribution of the intervention but also whatever endogenous contribution is supplied by X_1 as well. If, say, the endogenous value of X_1 is *k* units, then we can use the new pipe into X_2 to add additional units to X_2, but the total value of X_2 then will reflect the *ka* units contributed by X_1 as well as whatever is contributed by the new pipe. In this sort of case, with X_1 remaining connected to X_2, if the value of X_1 should happen to change during the course of the intervention on X_2, and this is not observed, or if it is observed, but we don't know the functional relationship between X_1 and X_2, and we can't directly measure the value of X_2 but only the change in X_2 supplied by the intervention, we may be misled about the actual value of X_2, and this may in turn mislead us about the relationship between X_2 and X_4. We may be similarly misled if X_1 is connected to X_4 via some other route that does not go through X_2, and we are unaware of this fact and the value of X_1 changes.

These observations suggest that there is another feature that it would be desirable for an intervention to have—it would be desirable if we could turn off the pipe linking X_1 to X_2 (but without at the same time turning off or disrupting other relevant pipes such as the pipe connecting X_3 to X_4) so that the *value* of X_2 and not just the *change in value of* X_2 is set entirely by the intervention. If we could do this, we wouldn't have to worry about the possibility that in the course of our intervention the value of X_1 happens to change and so misleads us about the relationship between X_2 and X_4. This idea—that an ideal intervention on *X* should break or disrupt the connection between *X* and its own immediate endogenous causes so that the value of *X* is set entirely by the intervention and is uninfluenced by its other previous causes—has come to be known as the "arrow-breaking" or "equation wipeout" conception of interventions, for reasons that I will come to shortly.

How might we capture this feature or interventions? One natural way of doing so appeals to the idea that the intervention variable *I* acts as a switch. Call the endogenous direct causes (other than *I*) of *X* the parents of *X*. For some values of *I* (when *I* is in the "off" position), the value of *X* is a function of the value of parents (X) alone. For other values of I (when I is "on"), the value of X is completely independent of the value of parents X and depends only on the value of I. In other words, the intervention variable I "interacts" with the variables parents (X) in such a way that when I is in the on position, the previously existing connection between parents (X) and X is "broken".

If we incorporate this feature as well into our characterization of an intervention, we have the following stronger notion which I label **SIN** for strong intervention:

SIN:
I satisfies I1–I4 and in addition

I5. *I* acts as a switch for all the other variables that cause X. That is, certain values of I are such that when I attains those values, X ceases to depend upon the values of other variables that cause X and instead only depends on the value taken by I.

Adopting the graphical representation of causal relationships described above, an intervention I (in the sense of **SIN**) on a variable X with respect to Y will "break" all other arrows directed *into* the variable X (besides the arrow from I to X) but will not break various other arrows, including arrows directed into Y that are not on a path (should this exist) from I to X to Y. For example, the effect of an intervention on X_2 in the structure represented by the equations (2.1)–(2.3) will be to replace this structure with the structure in Figure 9.4.

Represented in terms of equations the effect of an intervention on X_2 will be to wipe out the equation (2.1) in which X_2 occurs as a dependent variable, replacing it with a new equation (2.1*) which indicates that X_2 has been "set" to a new value (e.g., k) by the intervention while the other equations (2.2)–(2.3) remain undisturbed:

(2.1)	$X_2 = aX_1$	(2.1*)	$X_2 = k$
(2.2)	$X_3 = bX_1$	(2.2)	$X_3 = bX_1$
(2.3)	$X_4 = cX_2 + dX_3$	(2.3)	$X_4 = cX_2 + dX_3$

Both the arrow-breaking and the equation wipeout ideas are developed in (Spirtes et al. 1993, 2000; Pearl, 2000).

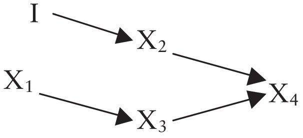

Figure 9.4

The difference between **WIN** and **SIN** is related to (although not identical with) the contrast between what Cartwright has called "value variation" and "causal law variation" conceptions of intervention in a recent paper (Cartwright 2003). The former involves variation in the value of a putative cause that arises because some exogenous variable in a causal system takes a different value. The latter involves a change in the structure of the original causal system, as when, under **WIN**, arrows directed into the variable intervened on are broken. Cartwright worries that although the distinction between the two kinds of interventions may seem clear conceptually, it may be unclear how to apply it to real-world situations (Cartwright 2003: 223). I hope that the characterization of arrow breaking in terms of a switch variable clarifies this notion and that the above remarks help explain why we need the stronger notion of intervention embodied in **SIN** rather than just **WIN**.

Several other features of both **WIN** and **SIN** deserve notice. First, note that the characterization does not make reference to human beings or human activities. Instead the conditions in **W/SIN** are characterized purely in terms of notionssuch as cause and (statistical) independence. Some manipulations carried out by human beings will count as interventions in the sense of **W/SIN** but, if so, will be in virtue of their causal and correlational characteristics. Moreover, an event or process not involving human action at any point will also qualify as an intervention as long as it satisfies **W/SIN**.

Second, a remark about "circularity": the characterization **W/SIN** employs causal language at a number of points. Not only must the intervention variable *I* cause *X*, but *I* must not itself directly cause *Y*, must not be correlated with other causes of *Y* that are independent of the putative $I \rightarrow X \rightarrow Y$ chain, and so on. Because the notion of an intervention is already a causal notion, one cannot use to it to explain what it is for a relationship to be causal in terms of concepts that are themselves noncausal. Thus, a manipulability theory that relies on the notion of an intervention will not allow us to translate or reduce causal claims into noncausal claims. However, there is also an important respect in which a theory that appeals to **W/SIN** to elucidate what it is for there to be a causal relationship between *X* and *Y* is not viciously circular: The characterization of an intervention on *X* with respect to *Y* does not make reference to the presence or absence of a causal relationship between *X* and *Y*. Instead, the causal information required to characterize the notion of intervention on *X* with respect to *Y* is information about *other* causal relationships or their absence: information about the causal relationship between the intervention variable *I* and *X*, information about whether there are other causes of *Y* that are correlated with I, information about whether there is a causal route from *I* to *Y* that does not go through *X*, and so on. The characterizations **W/SIN** also makes reference to the existence of correlational or statistical dependence relationships or their absence.

W/SIN thus fits naturally with a certain picture[1] of the epistemology of causal inference that is familiar from Neurath and Quine: We begin in *media res*, to speak, with a stock of already known causal and correlational information and use this to reach new conclusions about other causal relationships, perhaps using these new conclusions to revise other previously accepted causal beliefs, and so on. That there must be some explication of the notion of an intervention that is not viciously circular in the sense described above is strongly suggested by the fact that we do seem to sometimes find out whether a causal relationship exists between X and Y by manipulating X in an appropriate way and determining whether there is a correlated change in Y. This fact by itself seems to show that we must have some notion of a manipulation of X that would be suitable for finding out whether X is causally linked to Y and that this notion can be characterized without presupposing anything about the causal relationship (if any) between X and Y. It is just this notion that **W/SIN** attempts to capture.

AN INTERVENTIONIST INTERPRETATION OF CAUSATION

Suppose that we adopt **SIN** as our preferred characterization of an intervention. How then should the connection between the behavior of Y under an intervention on X and the existence of a causal relationship between X and Y be formulated? There are several possibilities. We might formulate the connection as a necessary or as a sufficient condition (or both) for causation, and we might formulate the connection as a claim about the response of Y to *all* or alternatively to *some* (range of) interventions on X.

With respect to the first issue, it is important to distinguish between two kinds of causal claims.[2] Consider the following causal structure.

Here X directly causes Y, and X also directly causes Z, which in turn directly causes Y. Assume that these relationships can be represented by means of the following equations

$$Y = aX + cZ \tag{4.1}$$

$$Z = bX \tag{4.2}$$

where a, b, and c are fixed coefficients. Then if $a = -bc$, the direct causal influence of X on Y will be exactly canceled out by the indirect influence of X on Y that is mediated through Z. Thus even though X directly causes and hence (in some relevant sense) causes Y, there are no manipulations of X

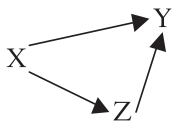

Figure 9.5

alone that will change *Y*. This example involves what Spirtes, Glymour, and Scheines call a failure of "faithfulness" (Spirtes et al. 1993, 2000).

Let us say that *X* is a total cause of *Y* if and only if it has a non-null total effect on *Y*—that is, if and only if a single intervention on *X* alone (with no other interventions on other variables[3]) will change the value of *Y* for some values of the other variables in the system **V** besides *X*. A change in the value of *Y* under interventions on *X* is thus both necessary and sufficient for *X* to be a total cause of *Y*. The notion of a total cause contrasts with the notion of a contributing cause which is meant to capture the intuitive idea of *X* influencing *Y* along some route even if, because of cancellation, *X* has no total effect on *Y*. The example under discussion shows that while it is arguably a sufficient condition for *X* to be a contributing cause of *Y* that the value of *Y* changes, given some intervention on *X*, this condition is *not* necessary. However, as I have discussed elsewhere (Woodward 2003), it is possible to give both necessary and sufficient conditions for *X* to be a contributing cause of *Y* in interventionist terms by invoking combinations of interventions. The basic idea is that *X* will be a contributing cause of *Y* if and only if there is a directed path from *X* to *Y* such that for some set of values of variables that are *not* on this path, if those values were fixed by interventions, there is some (single) intervention on *X* that will change the value of *Y*.[4] The notion of direct causation and hence the notion of a directed path can also be fully specified in interventionist terms.

What about the issue of whether the connection between manipulation and causation should be formulated in terms of what will happen under "some range of" or "all" interventions? In my view, a formulation in terms of "some" is preferable. For one thing, many causal relationships exhibit threshold effects. Some changes in the value of *X* may not change the value of *Y* while other changes in the value of *Y* may produce such a change. In such cases, we don't want to deny that there is a causal relationship between *X* and *Y*. Similarly, if other variables *Z* in addition to *X* are causally relevant to *Y*, it may be that for some values of *Z* no intervention on *X* will change the value of *Y*, while for other values of *Z*, interventions on *X* will change *Y*. Again in such cases it will be true that *X* causes *Y* despite the fact that some interventions on *X* don't change *Y*.

While the "some" formulation thus seems preferable, it also should be clear that the bare claim that *X* causes *Y* (i.e. that there are some interventions on *X* that will change *Y*) is very nonspecific and not very informative. (This is one point at which I'm in at least partial agreement with Cartwright, who also emphasizes the abstractness and nonspecificity of such claims, although for somewhat different reasons). From the perspective of a manipulability account, what one would really like to know is just which interventions on *X* will change *Y* (and in what circumstances) and exactly how they will change *Y*.[5] We may view this more detailed information about what will happen to *Y* under various hypothetical manipulations of *X* as the natural way of spelling out or capturing the detailed content of specific causal claims regarding *X* and *Y* within a manipulability framework. One way of specifying this information is by means of a set of equations like 4.1–4.2 .

INVARIANCE

As I have tried to show elsewhere (Woodward 2000, 2003), the notion of invariance is a useful device for conveying this more specific information. A relationship $Y = F(X)$ linking a vector of cause variables *X* to a effect variable *Y* is invariant if and only it would remain stable or continue to hold under some range of interventions on the variables *X*, where "continue to hold" means simply that the relationship correctly describes what the value of *Y* would be under such interventions. Typical causal relationships—or at least those causal relationships studied in the social and biomedical sciences—will be invariant under some range of interventions and in some background circumstances but not all. We can spell out the content of causal claims more precisely by providing such information about their range of invariance. Thus, in the example involving the salt chambers above, while the equations (2.1)–(2.3) may be invariant under some interventions that inject differing amounts of solution into the chambers, it is likely that they will break down under some other interventions and changes in background circumstances. For example, it may be that if too much solution is injected into the chambers or if it is injected too forcefully, the apparatus will break. Similarly, if the solution is heated to a temperature sufficient to melt the chambers.

According to the framework that I advocate, a set of equations like (2.1)–(2.3) will correctly capture some features of the causal relationships in the system it purports to represent as long as those equations are invariant under some range of interventions, even if they are not invariant under all interventions. The notion of invariance is thus a device for capturing the content of causal claims without appealing to relationships (like the philosophers' notion of a law of nature) that supposedly hold always or universally. A relationship can be locally stable—that is invariant under some range of interventions—without being invariant under all interventions in the way some philosophers believe fundamental physical laws to be.[6]

IN WHAT SENSE MUST INTERVENTIONS BE POSSIBLE?

Let me now turn to an issue that I have been ignoring. The view I have been describing connects causal claims to claims about what would happen under interventions. It is obvious, however, that for many causal claims, the relevant interventions will not or cannot occur. This may be so for any one of a variety of reasons. The claims in question may be about the causes of events that occurred in the past, as when extinctions of various plants and animals are attributed to meteor impacts. The relevant interventions may not be within the technological abilities of human beings, either at present or in the foreseeable future, and may be unlikely to occur naturally. More fundamentally, the kind of fine-grained, surgical change in a putative cause *X* that is required by the notion of an intervention may not be causally or nomologically possible—all possible interventions *I* on *X* with respect to *Y* may be "ham-fisted" or "fat-handed", affecting not just *X* and other variables lying on the route from *I* to *X* to *Y*, but also other variables that are not on this route and that affect *Y*.

Is an appeal to interventions to elucidate causal claims only useful when the relevant interventions can actually be carried out? In my view, the content of causal claims often may be clarified by invoking what would happen under hypothetical interventions even if those interventions cannot or will not be carried out. One role for an interventionist analysis in such cases is normative: It spells out what it is that we are trying to discover or infer. Consider the question of whether smoking causes lung cancer. There are obvious moral reasons why we are unwilling to conduct a randomized experiment (an intervention) in which subjects in a treatment group are forced to smoke while those in a treatment group are prevented from doing so, with the incidence of lung cancer in the two groups then being compared. Instead, we try to infer the effect of smoking on lung cancer in humans on the basis of observational or nonexperimental data (with perhaps some support from experimental data drawn from animal studies). Nonetheless, according to the interventionist account, when we carry out such inferences, we should think of ourselves as trying to determine what the result of such a hypothetical experiment *would be* if we *were* to perform it.

Thought of in this way, the interventionist account helps to clarify certain features of experimental design. When we do an experiment in which only fat-handed manipulations are possible, the interventionist account recommends that we nonetheless try to conduct the experiment so that its results bear on would happen if a non-fat-handed intervention in the sense of **SIN** were performed. An example due to James Bogen illustrates this point (Bogen 2002). Bogen thinks of the example as showing the limitations of interventionist treatments of causal claims, but it seems to me that it instead illustrates their appeal. Compressing greatly, the neurobiologist Karl Lashley was interested in testing the causal claim that primary visual cortex V1 plays a role in the maze-running abilities of rats blinded at birth. Bogen remarks that:

> The ideal way to test this [claim] on rats would be to blind them, train them to run the maze, and then lesion their visual cortices . . . without damaging any other parts of the brain. If performance is impaired then

> the visual cortex must be capable of at least two functions [that is, it must facilitate maze running in blind rats, as well as vision in sighted rats.] (Bogen 2002)

However, the technology available to Lashley did not allow him to perform such an "ideal" experimental intervention. He could not destroy substantial amounts of visual cortex to impair maze performance without destroying other adjacent areas that were not part of visual cortex. As Bogen explains, Lashley attempted to work around this difficulty by 'lesion [ing] the visual cortex in a variety of different ways, each one of which was designed to do collateral damage to a somewhat different adjacent structure'. Thus in one group of rats, the hippocampus was lesioned in addition to the visual cortex, but the auditory cortex was spared, in another group the reverse procedure was carried out, and so on for different groups. The performance of all groups was worse after lesioning.

Why did Lashley proceed in this way? The obvious answer is that he was concerned that in the fat-handed experimental manipulations he was actually able to perform, the effects of destroying the visual cortex would be confounded with the effects, if any, of destroying other parts of the brain. If, say, Lashley's experimental manipulation destroys both visual cortex and some portion of the auditory cortex, then it is possible that the diminished performance is due to the destruction of auditory rather than visual cortex. This is just to say that the experimental manipulation may not be an intervention on the visual cortex with respect to maze performance since it affects something else (auditory cortex) that may, for all that is known, affect maze performance independently of the damage to visual cortex.

Lashley seems to reason that if he finds diminished performance under each of a number of different experimental manipulations which destroy different parts of the brain in addition to the visual cortex, it is unlikely that this total pattern of diminished performance could be due to just to the destruction of these additional brain areas, with the destruction of the visual cortex playing no role. From my point of view, what he was doing was using what happened under various imperfect, probably fat-handed experimental manipulations to triangulate on (or act as a sort of surrogate for) what would have happened to maze performance in an ideal experiment in which only the visual cortex and nothing else was destroyed—i.e. in an experiment in which there is an ideal intervention on visual cortex alone. Bogen himself suggests something very close to this in the passage that I quoted above about what an "ideal test" of Lashley's hypothesis would look like. My suggestion, in other words, is that the interventionist account helps us to understand why Lashley designs his experiment as he does. For example, he is dissatisfied with the results of a single experiment that destroys both the visual and the auditory cortex because he thinks such an experiment fails to tell us what would happen in an ideal experiment in which there is an intervention that destroys the visual cortex alone and because it is the latter, not the former, that is relevant to the causal claim that he wants to assess.[7]

Consider another example bearing on this point—this one due to Cartwright (2001). She observes that in testing the efficacy of a drug experimentally, it is usual to deal with the possibility of placebo effects by

> . . . giv[ing] the patients in the control group some treatment that is outwardly as similar to the treatment under test as possible but is known to have no effect on the outcome under study.
>
> That is, we do not hunt for yet another way to get the medicine into the subjects, a way that does not affect recovery by any other route. Rather we accept that our methods of doing so may affect recovery in the way suggested (or by still other routes) and introduce another factor into the control group that we hope will just balance whatever these effects (if any) may be. (Cartwright 2001: 77)

Let *I* represent the experimenter's manipulations which consist (let us suppose) in injecting the drug (or perhaps something else) into a patient; *D* represents the presence or absence of the drug in the patient's bloodstream at some later time, *R* whether the patient recovers, and *P* whatever is involved in the occurrence (or not) of a placebo effect (e.g., it may be that the patient observes the injection, believes that it contains the drug, is optimistic about recovery as a result and that this exerts a positive influence on immune response (and hence recovery) that is independent of any action of the drug). The causal structure is thus as portrayed in Figure 9.6.

If the placebo effects are real, the experimenter's manipulations will not constitute an intervention. If they are treated as such, this results in a mistaken conclusion about the effect of *D* on *R*. One way of dealing with this possibility would be to redesign the experiment so that those manipulations do constitute interventions—that is by excluding the possibility that *I* affects *R* by a route that does not go through *D*. Depending on the details of the case, this may be easy enough to do—for example, it might suffice to give the patient the drug unawares, by mixing it in with their food or giving it to them while they are asleep. Cartwright's point is that this is *not* what is

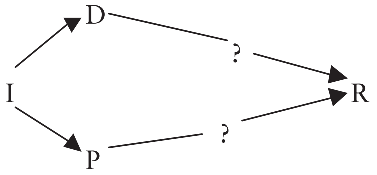

Figure 9.6

standardly done. (Presumably, the objection to proceeding in this way is moral, having to do with the absence of consent, rather than with technological impossibility.) Instead, what is usually done is to subject the patients in both the treatment and control group to the same possible placebo effect (say, by injecting the controls with some substance which may be believed by them to positively affect recovery but does not in fact do so) and then looking for a difference between the groups. In effect, one subtracts out any possible placebo effect from both groups. As Cartwright observes, this second procedure may be possible even if the first procedure—manipulating *D* in such a way that the manipulation has no effect on recovery independent of *R*—is not possible.

I agree with all of this but don't see it as undermining the interventionist account. Again, as I see it, the role of the interventionist account is to elucidate what we are trying to figure out when we ask what the effect of the drug on recovery is. When we carry out the second procedure, we are using it to try to figure out what would happen if an intervention on whether subjects receive the drug were to be performed, and the appropriateness of the second procedure should be assessed by whether it gives reliable information about this. The moral of the example is thus the same as the moral of Bogen's Lashley example: There are other ways of learning about what would happen if an intervention were to be performed besides actually performing the intervention.[8]

IS THE INTERVENTIONALIST ACCOUNT OPERATIONALIST?

In several recent papers Cartwright has criticized the interventionist account (or at least a condition that is closely associated with it, called modularity—see below) as "operationalist" (Cartwright 2001, 2002). I think that if taken literally, this criticism is misplaced. The interventionist account does not, as classical operationalism is alleged to have done, take one procedure for testing a claim and contend that the claim only makes sense or only has a truth value when that procedure can actually be carried out. Cartwright's example of an operationalist is someone who proposes to operationalize "length" using a footruler and then concludes that we cannot "sensibly talk of the size of a molecule" (Cartwright 2002: 421). She complains that the interventionist account 'overlooks the possibility of devising other methods for measuring' causal relationships and also suggests that the account leads us to 'withhold the concept [of cause] from situations that seem the same in all other respects relevant to its application just because our test cannot be applied in those situations' (Cartwright 2002: 422).

I hope that my discussion has made clear that my version of interventionist account does not hold that causal concepts apply or make sense only when interventionist tests for causation can actually be carried out. Nor does it deny that there are other ways of measuring causal influence or testing causal claims besides carrying out interventions. This having been said, I think that it is a fair complaint that the interventionist account is unclear about the sense in which interventions must be "possible" if an interventionist treatment of causation is to be illuminating. The above examples, as well as others I have described elsewhere (Woodward 2002, 2003), suggest that the account can be heuristically useful in illuminating both the content of causal claims and issues of experimental design in cases in which interventions are not technologically possible. There is a very substantial literature in statistics and econometrics that reaches a similar conclusion (see Cook and Campbell, 1979). Nonetheless there are obvious questions about how attenuated we can make the notion of an intervention before it ceases being even heuristically useful. For example, what about causal claims for which the relevant interventions are causally or nomologically impossible? What about cases in which the notion of an intervention is ill-defined for conceptual or metaphysical reasons? (Is there an intervention that will turn a human being into a member of some other species?)

I have explored these issues elsewhere (Woodward, 2003: Ch. 3). My inclination is to think that the notion of an intervention can be usefully employed in elucidating causal claims whenever contentions about what would happen under such interventions "make sense" and have definite answers, even if these answers can only be supplied by theory or by nonexperimental evidence rather than by direct experiment. Thus, in my view, it makes perfectly good sense to ask what the gravitational effect of the moon on the earth's tides would be under an intervention in which the distance between the moon and the earth is doubled, even if, as it may turn out, any physically possible process that would change the orbit of the moon would affect the tides in some other way, not involving the moon's gravitational attraction, and hence will not count as an intervention. Newtonian mechanics allows one to calculate what the effect on the tides would be if the distance from the earth to the moon were to double and nothing else relevant to the tides were to change; in my view that is enough for the question of what would happen under the above intervention to make sense.

Similarly, it seems to me that we are perfectly capable of making sense of and reasoning about what would follow from an intervention that introduces the medicine into the patient in a way that is independent of the placebo response in the example described above, even if available medical technology does not permit such an intervention. By way of contrast, there are other cases, such as a supposed intervention that transforms members of one species into another or which puts Julius Caesar in charge of U. N. forces in Korea (to use Quine's famous example), where we arguably have no coherent idea of what would be involved or what would happen if such an intervention were to occur.

INTERVENTIONISM AS A MONOCRITERIAL THEORY

There is another way of bringing out the difference between the inteventionist account and Cartwright's views about causation that does not invoke the charge of operationalism but nonetheless may help to isolate what Cartwright does not like about it. The interventionist account is monocriterial or, as Cartwright calls it, "monolithic": It takes just one of the criteria commonly thought to be relevant to whether a relationship is causal—whether it is potentially exploitable for purposes of manipulation—and gives it a privileged or preeminent place. When this criterion comes into conflict with other proposed criteria for causation (like spatiotemporal contiguity or transmission of energy-momentum), the account takes manipulability related considerations to trump the others. Although the matter deserves more detailed attention than I can give it here, I believe that this captures our judgments about particular examples—when there is conflict between different criteria for causation, our judgments are guided by considerations of manipulability (Woodward 2003). By contrast, as I understand her view, Cartwright thinks of causation as a "cluster concept"—a variety of different criteria are relevant to whether a relationship is causal and which of these are most appropriate or important will depend on the causal claim at issue. This in turn is related to her ideas about the diversity of different kinds of causal relationships, which I discuss in the final section, 'Causal diversity'.

CARTWRIGHT ON INVARIANCE AND CAUSAL CORRECTNESS

Let me now try to compare the account I have been sketching with some of Cartwright's claims about the connection between causation and invariance under interventions. Cartwright agrees with me that these notions are interconnected (at least for some kinds of causal systems) but understands the connections as well as the notions of intervention and invariance somewhat differently than I do. In particular, in several recent papers Cartwright offers proofs that for systems of equations satisfying certain additional assumptions, those equations will be causally correct (according to her understanding of causal correctness, about which more below) if and only they are invariant in a sense, different from mine, which she specifies (Cartwright 2002, 2003). She suggests that these proofs provide a more precise and formal explication of a connection that I have argued for more loosely and informally.

I won't try to discuss the details of these proofs—this would require a chapter in itself—but I do want to comment on a larger issue concerning argumentative strategy. As noted above, there a number of different things that people have meant, informally or pre-analytically, by "causal correctness".[9] The interpretive task, as I see it, is to flesh out, making more precise or at least more specific, what might be meant by this notion. A very rough analogy is provided by the informal notion within statistics of an estimator being good or having desirable characteristics. We have some rough idea, pre-analytically, of what this might mean but for the notion to be useful we need to decide on some more precise specification. Thus we arrive at the proposal that a good estimator should possess various properties—e.g., it should be unbiased, it should be a minimum-variance estimator among the class of unbiased estimators, and so on. Once we have made the notion of an estimator having good properties more precise in this way, we can then go on to ask various questions that have well-defined answers: e.g., under what conditions will such and such an estimator be unbiased? I take it to be clear that it would be misguided to complain of someone who proceeds in this way that he has failed to provide a mathematical "proof" that a good estimator must be unbiased; instead, "unbiasedness" is part of the explication we adopt for "good estimator". Or, to put the point more cautiously, the only sense I can attach to such a proof would involve first providing some alternative explication of the properties that make an estimator good and then proving that these properties imply unbiasedness. And if one proceeds in this way, one will then have to take the connection between being a good estimator and these alternative properties as not themselves something that can be proved. Moreover, one would also need to provide some reason for thinking that an explication of good estimator in terms of these alternative properties was in some way a better or more intuitive starting point than the explication in terms of unbiasedness.

As I see it, the connection between causal correctness and invariance under interventions has something like the same status. Once we agree to adopt invariance under intervention as an explication of causal correctness, we can then go on to raise questions about the relationship between causal correctness, understood along interventionist lines, and other notions of causal correctness, such as the notion of Granger-causation described in endnote 9. We can also ask such questions as: what sorts of evidence and additional assumptions are relevant to establishing that an equation is causally correct in the sense of invariance under interventions? But in general, providing a proof of the sort Cartwright seeks, connecting correctness and invariance requires providing some alternative specification of what correctness means and showing that this implies invariance. My inclination is to think that any alternative specification will be no more intuitive or preferable as a starting point than a specification in terms of invariance under interventions.

What is the alternative characterization of causal correctness that Cartwright assumes in her proofs? In part her characterization is informal. Some aspects—such as her claim that variables on the right-hand side of a causally correct equation should be "genuine causes" of the left-hand-side variable and that the equation should "get the weights right" (Cartwright 2002: 418)—are, in one sense, completely unexceptionable. They are conditions that virtually everyone will accept, regardless of the interpretation of causation they favor. But because they are not accompanied by any further explication of what it is for a cause to be "genuine" and so on, it seems to me that they do not afford much independent purchase on what is for an equation to be causally correct.

Cartwright also imposes additional formal constraints that a causally correct system must satisfy—for example, she assumes that what she calls nature's system (that is, the causal system as it is in the world, as opposed to our representation of it) must be antisymmetric and must satisfy "numerical transitivity". The latter condition requires that 'Causally correct equations remain causally correct if we substitute for any right-hand-side factor any function in its causes that is among nature's causal laws' (Cartwright 2003). "Numerical transitivity" has the consequence that the reduced form equations associated with a set of equations that are causally correct will themselves be causally correct, as will equations formed from causally correct equations by omitting causally relevant exogenous variables. One forms the reduced form equations associated with a system by first identifying the exogenous variables in the system—i.e. the variables that are not themselves caused by any of the other variables in the system and do not have arrows directed into them in the graphical representation of the system. One then substitutes into the equations in the system in such a way that one is left with a set of equations, one for each endogenous variable, which have the endogenous variable on their left-hand sides and only exogenous variables (and an error term) on their right-hand sides. It is always possible to do this and the resulting reduced form system will always be observationally equivalent to the original system.

As an illustration, suppose, following Cartwright, that the following system is causally correct, according to whatever standard of causal correctness is thought appropriate (Cartwright 2003: 216):

9.1 $$q_1 = u_1$$

9.2 $$q_2 = a_{21}\, q_1 + u_2$$

9.3 $$q_3 = u_3$$

9.4 $$q_4 = a_{41}q_1 + a_{42}q_2 + a_{43}q_3 + u_4$$

Cartwright tells us that according to her notion of causal correctness the following equation also will be causally correct.

9.5 $$q_4 = (a_{41} + a_{42}a_{21})q_1 + R$$

where R is presumably just $R = a_{42}u_2 + a_{43}u_3 + u_4$

(9.5) is the reduced form equation associated with the original system. By contrast, in my version of the interventionist account, (9.5) is not causally correct, at least if, as I assume, it is intended to stand alone as a correct representation of the original system. The reason is that it fails to represent what will happen under interventions on endogenous variables like q_2 and

q_3 and that it fails to represent direct causal relationships. In contrast, if the original system is causally correct, it will provide this information.

Although Cartwright is not fully explicit about this point, my guess is her endorsement of numerical transitivity is motivated in part by a more general skepticism about the idea that there is a matter of fact about which causal relationships in nature are direct rather than indirect. I have some sympathy with this skepticism, since which causal relationships are direct will be relative to a choice of representation or a variable set. On the other hand, we seem to need the notion of direct causation for a variety of purposes: to represent what will happen under combinations of interventions, some of which involve endogenous variables; to allow us to track the consequences of changes that disturb only portions of a system while leaving other portions intact; and to formulate principles connecting causation and probabilities (Woodward 2003). Quite a bit thus appears to be lost if, like Cartwright, we adopt an understanding of causal correctness that does not require the representation of direct causal relationships and relationships between endogenous variables. It is presumably for such reasons that researchers generally are not content to work just with reduced form equations.

CARTWRIGHT ON INTERVENTIONS

Another difference between Cartwright's views and mine concerns the characterization of the notion of an intervention. As noted above, in the characterizations **WIN** and **SIN**, an intervention on *X* with respect to *Y* makes no reference to the causal relationship, if any, that exists between *X* and *Y*. This contrasts with an alternative way of characterizing the notion of an intervention, much more common in the recent literature, that does build into the notion reference to the impact of the intervention on the causal relationship, if any, between *X* and *Y*. Cartwright advocates a version of this proposal in several recent papers. She writes:

> *I* is an intervention on *X* if *I* is an *HW* intervention on *X* and all the causal equations remain the same. (Except that when we intervene on *X* by changing the causal equations that govern *X*, then those equations that have *X* as an effect or that have *X* as cause and effects of *X* as effect must be dropped or altered appropriately). (Cartwright 2002: 416) [An *HW* or Hausman–Woodward intervention is, roughly, an intervention that satisfies the exogeneity conditions in **WIN**.]

The full details of this proposal do not matter for what follows. What does matter is that Cartwright clearly intends that the "causal equations that stay the same" under an intervention on *X* should include the equations, if any, linking *X* to its effects. In other words, one of the conditions that a manipulation of *X* must satisfy if it is to count as an intervention in

Cartwright's sense is that the manipulation should not disrupt any causal relationship between *X* and its effects. Call this the preservation requirement. A number of other writers, including Judea Pearl, also endorse this requirement—for example, it is built into Pearl's notion of an atomic intervention (Pearl 2000).[10] By way of contrast, **W/SIN** does not impose the preservation requirement. Condition I4 in **W/SIN** does require that the intervention *I* on *X* with respect to its putative effect *Y* not alter the relationship between *Y* and any of its causes *Z* that are not on any directed path (should such a path exist) from *X* to *Y*, but this condition says nothing about whether *I* alters the connection between *X* and *Y*.

Why impose the preservation requirement? One motivation may seem obvious: If our manipulation of *X* destroys the causal relationship that connects *X* to *Y*, so that *Y* does not change under manipulation of *X*, then we may be misled into thinking that there is no causal relationship between *X* and *Y*, when in fact such a relationship exists. Note, however, that we will make this mistaken inference only if we formulate the connection between causation and intervention as a claim about what will happen under "all" rather than, as recommended earlier, under "some range of" interventions. If we formulate the connection in terms of "some" interventions, we will not be justified in concluding that *X* does not cause *Y* just because there is some intervention on *X* (a manipulation that destroys the causal relationship connecting *X* to *Y*) that does not change *Y*. Instead, there will be a causal relationship between *X* and *Y* as long as there is some range of interventions on *X* which are associated with a change in *Y*. Moreover, as argued above, there are independent reasons for adopting the "some" interventions formulation.

We may further explore what is at issue here by considering another example: a spring that (within a certain range of extensions) conforms to Hooke's law, $F = -kX$. Imagine a manipulation of the extension that satisfies the conditions for an intervention on *X* with respect to *F* in **SIN** but that stretches the spring so much that it breaks. I take Cartwright's (and Pearl's) view to be that such a manipulation should not count as a bona fide intervention, at least if we take the relevant mechanism to be what is described by Hooke's law.[11] By contrast, according to **W/SIN**, such a manipulation is an intervention. It is true that once an intervention occurs that breaks the spring, no subsequent changes in the extension will change the restoring force—hence there is no causal connection between *X* and *F* once the spring is broken. But before the spring-breaking intervention occurs, there was (according to the interventionist conception) a causal relationship between *X* and *F*, since there were other more moderate non-spring-breaking interventions on the extension that would have changed the restoring force. In general, as suggested above, we may specify the causal relationships between *X* and *F* more precisely by specifying the range of interventions over which it is invariant and the interventions over which it fails to be invariant. The latter are, to repeat, genuine interventions on my view.

One apparent consequence of adopting the preservation requirement is that in order to determine whether we have carried out an intervention on *X*, we must have some basis for determining whether our manipulation of *X* has disrupted the causal relationship, if any, connecting *X* to *Y*, or, more precisely, the function that we take to characterize this relationship. This in turn seems to require that we already have some information about the causal relationship, if any, between *X* and *Y* and introduces a worry about a kind of "circularity" that seems to be potentially much more vicious than the circularity built into **SIN** and **WIN**. While **W/SIN** builds information about other causal relationships, besides the relationship between *X* and *Y*, into the characterization of an intervention on *X*, the mechanism-preserving requirement builds into that characterization information about very thing that we want to characterize—the causal relationship between *X* and *Y*. In other words, if we adopt the mechanism-preserving requirement, then we seem to lose the possibility of carrying out the interpretive project described earlier: using the response of *Y* to an intervention on *X* to characterize what it is for *X* to cause *Y*.[12]

Commenting on this concern, Cartwright denies that circularity associated with the mechanism preserving requirement is always or automatically vicious. She writes:

> we may often be in a position to assume that what we do to change *X* has very little chance of changing the laws about what *X* causes even if we do not know exactly what those laws are. If we are not in that position, we are not able to rely on our test [linking what happens under an intervention on *X* to the presence of a causal relationship between *X* and *Y*]. (Cartwright 2002: 416)

I agree that one can *sometimes* recognize that a contemplated manipulation of *X* is likely to disrupt any causal relationship between *X* and *Y*, should any exist, without knowing whether there is in fact such a relationship.[13] But I think that it is also true that we sometimes find out about whether there is a causal relationship between *X* and *Y* and about its characteristics by means of relatively "black box" experiments—by manipulating *X* in circumstances in which we have little, if any, prior information about its causal relationship (if any) with *Y*. In the case of the spring, we find out whether there is a causal relationship between *X* and *F* and what its characteristics are—its functional form, the range of interventions under which it is invariant, and so on—by manipulating *X*. This in turn suggests that there must be some legitimate characterization of the notion of an intervention on *X* that builds on little or no information of this sort. **W/SIN** attempts to provide such a characterization.

For all of these reasons, I conclude that if our purposes are to provide an account of the content of causal claims in interventionist terms, we should adopt a conception of intervention like **WIN** that does not impose the preservation requirement. This view also allows us to talk, as we did in the spring example, in terms of a generalization being invariant under some range of interventions and breaking down under others.[14] It also has the virtue of making it possible to provide a noncircular interventionist explication of what it is for *X* to cause *Y*.

MODULARITY

I turn now to some remarks on the notion of modularity. Consider the following system of equations:

11.1 $$Y = aX + U$$

11.2 $$Z = bX + V$$

and the associated directed graph in Figure 9.7.

According to our earlier discussion if (11.1)–(11.2) correctly represents the causal facts, then each individual equation (11.1) and (11.2) must be invariant under some range of interventions on the right-hand-side variables of that equation.

Suppose, however, that the situation is this: although (11.1) and (11.2) continue to hold under some range of manipulations of *X*, all possible ways of changing *Y* disrupt (11.2) in the sense of changing the relationship between *X* and *Z* described by (11.2). Given the way that we defined the notion of an intervention in **SIN**, this will be a situation in which, although it is possible to carry out interventions on *X* with respect to *Y*, one cannot carry out an intervention on *Y* with respect to *Z*, since any manipulation of *Y* will disrupt the relationship between *Z* and its cause *X* in violation of clause 4 in the characterization of an intervention which, it may be recalled, says:

> 14 I does not alter the relationship between *Y* and any of its causes Z that are not on any directed path (should such a path exist) from X to Y.

The requirement of modularity is designed to rule out the possibility just described. It says, in effect, that for each variable in the system, including all of the endogenous variables, it is possible (in the attenuated "makes sense" notion of "possible" gestured at in the 'Is the interventionalist account

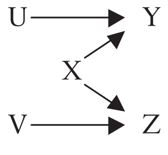

Figure 9.7

operationalist?' section) to carry out an intervention in the sense of **SIN** on that variable. Slightly more precisely, the modularity requirement may be formulated as follows:[15]

> MODULARITY. A system of equations is modular iff (i) each equation is invariant under some range of interventions on its independent variables and (ii) for each equation, it is possible to intervene on the dependent variable in that equation in such a way that only the equation in which that dependent variable occurs is disrupted while the other equations in the system are left unchanged.

One way of motivating this modularity requirement appeals to an idea about distinctness of causal mechanisms: Ideally, each equation in a system of equations should represent a distinct causal mechanism, where the criterion for distinctness of mechanisms is that distinct mechanisms should be changeable (in principle) independently of one another. This in turn means that it should be possible to alter or disrupt each of the equations in the system (i.e. by manipulating the dependent variable in such a way that the mechanism or relationship that the equation represents is disrupted) without altering or disrupting the other equations. Applied to (11.1)–(11.2) what this means is that if these equations correctly describe the causal relationships in the system they represent, then the mechanism or causal relationship by which *X* affects *Y* should be distinct from the mechanism by which *X* affects *Z* and this in turn means that each mechanism should be disruptable independently of the other.

As another illustration, consider a causal structure in which atmospheric pressure *A* is a common cause of the occurrence/nonoccurrence of a storm *S* and the reading *B* of a barometer, with no causal connection between *S* and *B*, as represented in Figure 9.8.

Modularity is intended to capture the idea that if the mechanism or causal relationship connecting *A* to *B* is genuinely distinct from the mechanism connection *A* to *S*, then it should be possible in principle to intervene to disrupt the connection between *A* and *B* (say by manually manipulating the position of the barometer dial in a way that is independent of *A*) without disrupting the relationship between *A* and *S* and vice-versa. If we cannot, even in principle, do this (if, e.g., the system is a quantum mechanical one and *B* and *S* are in an entangled state), then we should not think of the system as having a common cause structure.

$$S = aA$$

$$B = bA.$$

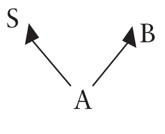

Figure 9.8

In several previous papers (Hausman & Woodward 1999; Woodward 1999), Dan Hausman and I claimed that when a system of equations is not modular it will fail to accurately and completely represent the causal relationships it models. One way of putting the argument is this: consider a process that fixes *Y* to some value *y* in such a way that *Y* is no longer influenced by *X*. We may represent this by replacing (11.1) with a new equation (11.1*) $Y = y$. This captures the fact that the value of *Y* is now fixed at y rather than, as was previously the case, being determined by *X*. If, under all such changes in *Y*, equation (11.2) is disrupted, then the system (11.1)–(11.2) will not be modular. Moreover, if (11.2) is disrupted whenever (11.1) is replaced by an equation of form (11.1*), then the value of *Z* will change under these changes in *Y* even when the value of *X* does not change, and even though (11.1)–(11.2) claims there is no causal connection between *Y* and *Z*. This in turn suggests that the representation (11.1)–(11.2) is inadequate or incomplete in some way—e.g., perhaps there is a causal relationship connecting *Y* to *Z* that is not represented by (11.1)–(11.2). More generally, we can say that if all changes that alter one equation also alter some other equations in a system, then the system will be misspecified in the sense that it will fail to correctly and completely represent the causal structure that it purports to model—variables will change in response to changes in other variables even though the equations represent the variables as causally unrelated or, alternatively, will fail to change under changes in the values of other variables in the way that the equations suggest that they should.

We can bring out more clearly what modularity involves by considering the following system of equations

11.3 $$Y = aX$$

11.4 $$Z = bX + cY$$

Let us now rewrite (11. 3) and (11.4) as follows

11.3 $$Y = aX$$

11.5 $$Z = dX$$

where $d = b$+ac.

Since (11.5) is obtained by substituting (11.3) into (11.4), the system (11.3)–(11.5) has exactly the same solutions in *X*, *Y*, and *Z* as the system (11.3)–(11.4). If we assume that *X*, *Y* and *Z* are the only measured variables in our system, then there is an obvious sense in which (11.3)–(11.4) and (11.3)–(11.5) are "observationally equivalent"—they imply or represent exactly the same facts about the patterns of correlations that obtain so far among these measured variables. Nonetheless by the rules given above for interpreting systems of equations, these two systems correspond to different causal structures. (11.3)–(11.4) says that *X* is a direct cause of *Y* and that *X* and *Y* are direct causes of *Z*. By contrast, (11.3)–(11.5) says that *X* is a direct cause of *Y* and that *X* is a direct cause of *Z* but says nothing about a causal relation between Y and Z. This difference is also reflected in what the two systems imply about would happen under interventions on their endogenous variables. (11.3)–(11.4) implies that under an intervention on Y, Z will change, while (11.3)–(11.5) denies this. Thus the two systems are "observationally equivalent" in the sense that they agree about the pattern of correlations that will be observed as long alterations in the structure of the system generating those correlations do not occur, but they disagree about what would be observed were such an alteration in structure to occur.

Despite this "observational equivalence", if (11.3)–(11.4) is modular, then (11.3)–(11.5) cannot be (and vice versa). To see this, consider an intervention on the variable Y in (11.3) which replaces (11.3) with the new equation (11.3*) $Y = y$. In effect, what this intervention does is to set the coefficient a in (11.3) equal to zero. If the system (11.3)–(11.4) is modular, (11.4) will continue to hold under at least one such intervention that replaces (11.3) with (11.3*). But, if (11.3)–(11 .4) is modular, (11.5) must change under this intervention since, as we have seen, its effect is to change the value of the coefficient a in (11.3) and the coefficient d in (11.5) is a function of a. Thus changing a in (11.3) will change d and hence (11.5). This corresponds to our judgment that if (11.3)–(11.4) is a correct representation of the causal facts, then (11.3)–(11.5) collapses or mixes together distinct mechanisms or causal routes—the influence of X on Z that occurs because X directly influences Z (this is represented by the coefficient b) and the influence which occurs because X influences Y which in turn influences Z (this is represented by the product ac)—into a single overall mechanism linking X and Z, which is represented by the coefficient d. This failure to correctly segregate the system being modeled into distinct mechanisms is directly reflected in the nonmodularity of (11.3)–(11.5).

In fact there are a number of other systems of equations besides (11.3)–(11.5) that can be obtained from (11.3)–(11.4) by algebraic transformations and which correspond to distinct systems of causal relationships. If one accepts that, despite their observational equivalence, at most one of these systems can correctly represent the causal facts, there must be some additional constraint that is satisfied by the correct system. Modularity is the natural candidate for this constraint. The idea is that among all of the observationally equivalent representations we should prefer the one that is modular because it will be the one that correctly and fully represents causal relationships and mechanisms. As Alderich puts it, the constraint of modularity (or as he calls it "autonomy") picks out a "privileged parameterization" (Alderich 1989).

The equations (11.3)–(11.5) are the reduced form equations (in the sense described in the section 'Cartwright on invariance and causal correctness') associated with the system (11.3)–(11.4). Although, as noted above, the reduced form equations will always be observationally equivalent to the original system from which they are formed, the former will not be modular if the latter is. To the extent that researchers are often not content with nonmodular reduced form equations, and instead prefer a modular set of equations, this is presumably because they regard such equations as providing a more adequate and complete representation of causal relationships.

These remarks provide, I hope, some motivation for the modularity requirement. As noted above Cartwright imposes a requirement ("numerical transitivity") that has the apparent consequence that if a system of equations is "causally correct", the associated reduced form system is also "causally correct". By way of contrast, if, as I would argue, a causally correct system must be modular, then in my view the reduced equations will not be causally correct if the original system is. It is thus not surprising that Cartwright rejects the modularity requirement. One of her arguments appeals to a contention about the "job" of a set of equations. Commenting on the contention that a set of equations that is nonmodular fails to fully capture causal relationships, she writes that this contention

> . . . is not true. We have a job that we want equations [11.3] and [11.4] to do—give a full non-redundant set of causes for and set out the true causal equations between these causes and their effects. These equations are not supposed to give information about why they are the true causal equations for the situation, nor about what causal equations might hold if they did not hold. Why on the occasion is it impossible to change one without changing the other? Such information may exist but it is not the job of the equations to convey it. Nor can we assume that both jobs together can be done by the same equation. (Cartwright 2002: 418)

Elsewhere she writes:

> The equations in question already have a job to do. The normal understanding is that we are discussing equations that (a) pick out for the given effect the full non-redundant set of causes and (b) lay out the functional form of the (true) causal law that holds between these causes and the effect. We can if we want change the subject. We can talk instead about sets of equations that represent relations, each of which can be interfered with separately [which is what Modularity requires.—J.W.]. But there is no reason to think that equations (if there are any) that do this new job will have the characteristics usually connected with sets of equations that do the original job. (Cartwright 2002: 418)

I agree with part of these remarks. No equation provides, by itself, information about why it holds. Presumably, this sort of information will only be provided by some other more general equation, typically not part of the same system as the original equation, as when General Relativity explains why and under what circumstances Newtonian gravitational theory holds. Modularity does not claim otherwise. Nor does modularity claim that a system of equations should tell us what would happen under all possible ways of disrupting those equations, including those that do not involve interventions. Instead to say that a system of equations is modular is to make a claim about what would happen under a very specific situation in which one of the equations does not hold: If the system is modular, then for each equation there is at least one situation (involving an intervention on the dependent variable in that equation) in which it fails to hold and the other equations do hold.

In the remarks quoted above, Cartwright maintains that we "change the subject" away from representing causal relationships when we impose a requirement like modularity and that the information associated with this requirement is independent of or different from the job of specifying causes. Whether imposing such a requirement really represents a change of job or subject in this way is, as I see it, exactly the point at issue. Unlike Cartwright, I see the satisfaction of a modularity requirement as closely bound up in the task of fully and accurately representing causal relationships. As argued above, if we are willing to assume that an interpretation of a system of equations should tell us what will happen under interventions on all of variables in the system, including endogenous variables, some version of the modularity requirement is very natural. It seems to me the real import of this part of Cartwright's objection is that she doesn't think that causal claims need to (or should be) given an interpretation in terms of claims about the outcomes of hypothetical manipulations.

If this is what the "two-jobs" criticisms really comes to, it seems to me that it would be more convincing if Cartwright were able to provide a clear alternative interpretation of what equations like (11.3)–(11.4) mean—an interpretation that does not appeal to or have implications concerning the outcomes of hypothetical manipulations. This could then be used to show that the two jobs that Cartwright claims are distinct really are distinct and that one can specify causes without committing oneself to claims about what would happen under hypothetical interventions. In my view, Cartwright does not really provide such an alternative interpretation, at least in the passages quoted above. While no one will disagree with her contention that causally correct equations should "give a full non-redundant set of causes for and set out the true causal equations between these causes and their effects", such remarks do not deliver what we are really looking for, which is some independent purchase on what the quoted phrase means. An interventionist theory represents one attempt to provide such an independent purchase. I take it that Cartwright is reluctant to provide such an alternative interpretation at the level of generality of the interventionist theory because of her views about the diversity of causal relationships and the inadequacy of any single monolithic account.

Cartwright has other criticisms of modularity besides the "two jobs" objection. Among other things, she holds that it is false as a matter of empirical fact that whenever two causal mechanisms or relationships are distinct, it will be possible to disrupt one relationship without disrupting the other (or to intervene to alter the value of the dependent variable in one of the relationships without altering the other). One of her most suggestive claims is that what seem intuitively to be distinct mechanisms may be realized together in the same spatiotemporal location in such a way that intervening to disrupt one of the mechanisms without disrupting the other is impossible. Of course, in evaluating this claim much will depend on the sense in which we demand that interventions be "possible". I take Cartwright to insist that the relevant notion of possibility must be a robust, nonattenuated one.

Cartwright offers several examples of mechanical devices in which she claims that modularity fails. One that figures centrally involves a toaster that works in the following way:

> The expansion of the sensor due to the heat produces a contact between the trip plate and the sensor. This completes the circuit, allowing the solenoid to attract the catch, which releases the lever. The lever moves forward and pushes the toast rack open.
>
> I would say that the movement of the lever causes the movement of the rack. It also causes a break in the circuit. Where then is the special cause that affects only the movement of the rack? Indeed, where is there space for it? The rack is bolted to the lever. The rack must move exactly as the lever dictates. So long as the toaster stays intact and functions as it is supposed to, the movement of the rack must be fixed by the movement of the lever to which it is bolted. (Cartwright 2001: 72)

I take her point to be that, given the design of the toaster, there is no way to alter the position of the rack, except by moving the lever, and this will also affect the circuit, which is intuitively a different mechanism. The obvious response to this is that there certainly *is* a way of moving the rack independently of the lever—all one has to do is detach or "unbolt" (her word) the rack from the lever and move it independently.

If I have understood her correctly, Cartwright's objection to this is that the bolting of the lever to the rack is not a cause of the movement of the rack in addition to the lever itself. Rather the bolting pertains to the "identification" of the system whose behavior we are trying to understand—it is part of what makes the system a toaster of a particular design (Cartwright 2002: 46). 'Without the specific design under consideration the question of causal connection, or lack of it, between levers and racks, is meaningless' (Cartwright 2002: 80).

But whether or not we choose to describe the bolting as a cause, it is nonetheless true that unbolting the lever from the rack (or sawing it off or whatever) will permit the independent movement of the rack and this is all that modularity requires. Whether or not the toaster would remain a toaster of a particular design under this sort of manipulation is irrelevant to whether the modularity requirement is satisfied. Moreover, while it may be that there is no general answer to the question of what the causal connection is between levers and racks, it is dubious that this question only makes sense within a toaster of a particular design. Assuming that the lever and rack are rigid bodies, as long as they are bolted together, it is a causal truth that moving one will move the other. This claim will remain true if this structure is taken outside of the toaster or placed in some completely different machine, as long as both components are able to move freely. Similarly, solenoids are components of many different machines, and the principles governing their operation are relatively stable across such different contexts—this is what makes engineering, conceived as a generalizing discipline, possible.

The failure of this particular example to undermine modularity does not of course rule out the possibility that there are other counterexamples. As noted above, Cartwright suggests that one important class of counterexamples involves cases in which mechanisms that we are prepared to count as distinct on some intuitive basis are not spatiotemporally distinct in a way that allows for separate interference. Obviously, in evaluating this claim much will depend on just what the proposed alternative intuitive basis for individuating mechanisms involves. One apparently natural possibility is that mechanisms should be individuated functionally or in terms of the tasks they perform—different tasks or functions mean different mechanisms. (This "functional" notion of mechanism is commonly assumed in several disciplines—e.g., cognitive psychology.) It is easy to imagine cases in which distinct functional mechanisms draw for *part* of their operation on a shared structure occupying a single spatiotemporal location. Think, for example, of a single neuron or set of neurons that plays a role in (what seems to count as) two distinct functional or computational tasks, each involving larger complexes of neurons. For example, there are reasons to believe that a common neural structure is involved in both in the performance of certain actions (e.g., grasping a peanut) and in the perception of the performance of this action by conspecifics—this is the so-called "mirror neuron" phenomenon. If we disrupt this shared neural structure, we disrupt both these tasks, and if each task corresponds to a distinct functional mechanism, both functional mechanisms will be disrupted. But in such cases, each task also will make use of additional structures that are unique to that task and are in different spatiotemporal locations. For example, animals that perceive the performance of an action by a conspecific do not always perform the same action themselves, and this is presumably because some additional structure (besides the mirror neurons) must be activated (or perhaps inhibited) for performance. In principle, if we interfere with this unique, nonshared structure, we should be able to disrupt one functional mechanism without disrupting the other.

Suppose, however, that both putative functional mechanisms make use of exactly the same structure in exactly the same spatiotemporal location. Then (at least in this case and perhaps more generally) one might wonder why a purely functional criterion for the individuation of mechanisms is appropriate. Why not conclude instead that the case is one in which a single mechanism performs several distinct functions or tasks or, alternatively, reconceptualize the tasks so that they are really a single task (characterized at some abstract level) after all? For example, it is not implausible that there is some single computational task that is performed by a given kind of mirror neuron, regardless of whether they are involved in motor activity or perception.

As this example brings out, to show convincingly that it is possible to have distinct mechanisms that are not separately disruptable, we need some principled, alternative criterion of mechanism individuation that does not involve independent disruptability and an argument that this criterion should take precedence over the independent disruptability criterion when the two conflict. To the extent that the alternative criterion is understood purely functionally, it may not be easy to defend.

This having been said, I readily acknowledge that there are many other considerations that are relevant to the assessment of the modularity condition that I have left unexplored and that a more detailed look at these may well convince us that there are cases in which the condition is violated. Suppose that this happens. What would this show? There has been a strong tendency in philosophical discussion—a tendency which I've sometimes unreflectively shared—to think that interesting claims about causation should be true universally, or even *a priori* (they should be "built into" the concept of causation), so that a single counterexample or perhaps even the logical possibility of a counterexample, discredits the claim entirely.

As I say, I've argued in this fashion myself, but I now wonder whether our habit of thinking that the only interesting alternatives are "universally true" and "has at least one exception" is always the most helpful way of looking at matters. In a recent talk at *PSA 2002* and in a forthcoming paper (Gopnik et al.), the psychologist Alison Gopnik suggests an analogy between a principle sometimes claimed to govern causation—the Causal Markov principle[16]—and "default principles" governing the operation of the visual system. A basic problem faced by the visual system is to infer full three-dimensional images from more fragmentary two-dimensional information falling on the retina. In carrying out such inferences the visual system is guided by certain general principles connecting these two—for example, the principle that illumination is from above, that sharp changes in visual properties signal object boundaries, that apparent size is correlated with distance from the observer, and so on. These principles are not infallible—there are such things as visual illusions—but they are reliable in many typical circumstances in which the visual system operates, and if they were not so reliable our visual systems would, at the very least, need to be designed differently. The default assumption of the visual system is that it is in circumstances in which such principles are true, at least in the absence of specific evidence to the contrary.

Gopnik suggests that the Causal Markov condition has a similar status. Hausman and I have defended that condition elsewhere (Hausman & Woodward 1999, forthcoming a, b), and I will not repeat our arguments here. However, I do want to suggest that the default analogy or something like it may be a useful way to think about the status of assumptions like modularity and perhaps other assumptions that go into the interventionist account of causality. That is, while there may indeed be circumstances in which modularity fails, it is a reliable default assumption (in at least many) typical circumstances in which we engage in causal reasoning. Moreover, if this assumption were really to fail generically, we might well think about causation in a different way, scaling back the significance of considerations having to do with manipulation, which I take to be central at present to our concept, and replacing them with something else. In this way, that modularity generally holds may be, as it were, a central part of the background we assume when we employ the notion of causation, even if it is not part of the content of this concept itself.[17]

Cartwright would of course deny this last claim. Her contention is not merely that there are exceptions to modularity but that this condition fails "generally". Here I will just say that even if her particular counterexamples to this condition are accepted, they fall well short of establishing this more general contention.

Perhaps the following remarks will go some way toward isolating the differences between the interventionist account and Cartwright's views about causation. One of the central ideas of the interventionist account is a "Galilean" idea about the function of experiments: One can learn about the causal structure of a complex system by disrupting some parts of it while leaving other parts intact, by taking the system apart, trying to understand whatever principles govern its components, taken in isolation, and then understanding the overall behavior of the system as the result of the principles governing these individual components. In effect, one learns about the original system by considering new systems in which some but not all of the causal connections in the original system hold. Part of the motivation for this way of proceeding is that the correlations that obtain in the original system may mislead us about the causal connections that hold in that system. For example, correlations may suggest causal connections where none obtain, as when the joint effects *X* and *Y* of a common cause *C* are correlelated despite the absence of a causal connection between them. Alternatively, causal connections may fail to reveal themselves in correlations, as in failures of faithfulness. In such cases, by removing or altering some causal connections in the original system, one allows other causal connections (or the absence of such connections) to reveal themselves in correlations (or the absence of correlations). Thus intervening on one of the joint effects *X* in the common cause example breaks the causal connection between *X* and *C* and makes the correlation between *X* and *Y* disappear, revealing that *X* doesn't cause *Y*.

Obviously, this whole way of proceeding requires that there be some systematic connections between the behavior of the components of the original intact system and the behavior of those components when other parts of the original system are interfered with—under some possible interferences, we should expect the components to continue to operate in the same way as they do in the intact system. This is the intuition behind modularity.

Of course whether any given manipulation which disrupts some component allows others to continue to operate as before is an empirical matter. But when Cartwright claims that it is not the job of a system of equations to tell us what would happen if one of the equations were altered, she seems to me to come close to challenging the whole thrust of the methodology sketched above. If causal claims about the behavior of one part of a system really have no implications at all for how that part would continue to behave under interference with other parts, what story, if any, can be told about the take-things-apart-to-see-how-they-work strategy?

This Galilean strategy contrasts with a more "Aristotelian" picture to which I suspect Cartwright may be attracted. According to this picture, understanding the behavior of a complex system is a matter of understanding how those components perform *in situ*, with the system remaining intact. When we take the system apart or otherwise disturb the integrity of some of its parts, we should not expect there to be any systematic connections between the unmanipulated and manipulated systems. (Recall Cartwright's remarks in the passage quoted above about the irrelevance of the behavior of the components of a partially dismantled toaster to understanding the operation of an intact toaster.) A set of equations that purports to describe the causal relationships in an unmanipulated system thus carries with it no general implications at all for what we should expect the causal relationships to be in the manipulated system. There may or may not be such connections but is no part of the business of the original system of equations to convey them. On this view of the matter, the modal or counterfactual commitments carried by causal claims are weaker or less extensive than the commitments that the Galilean picture thinks they carry—they don't tell us anything about what would happen if we were to change the causal structure of the original system in various ways. Modularity expresses the denial of this view.[18]

CAUSAL DIVERSITY

In her more recent work, Cartwright has strongly emphasized the theme of causal diversity and the abstractness and nonspecificity of locutions like "*X* causes *Y*". Following Elizabeth Anscombe (Anscombe 1971), she draws our attention to the great variety of specific causal locutions in ordinary language, and suggests that there is no such thing as "the" causal relationship, and that instead we should think in terms of a multiplicity of different causal relations:

> Causes make their effects happen. That is more than, and different from, mere association. But it need not be one single different thing. One factor can contribute to the production of another in a great variety of ways. There are standing conditions, auxiliary conditions, precipitating conditions, agents, interventions. (Cartwright 1999: 18)

While I am sympathetic to some of these themes, I would put matters differently. If the interventionist analysis is on the right track, then, as Cartwright maintains, claims of the form "*X* causes *Y*" are very nonspecific and not very informative. They say only that there are some interventions on *X* that will change the value of *Y*, and, for many purposes, we want much more specific information: information about just which interventions on *X* will change *Y*, in what way or by how much, and so on. The specific causal verbs to which Anscombe and Cartwright draw our attention—"attracts", "repels", "raises", "lowers", "pumps", "breaks", "knocks over", and so on, provide some additional information about these matters but it is worth noting that they are still qualitative and imprecise, still abstract and "nonspecific", although less so than "causes". In my view, there is no reason why these locutions should be accorded a privileged status in discussions of causation, as though causation really happens only at the "level" of pushes and pulls, and not at more specific or generic levels. It seems to me that what one really wants, especially in scientific contexts, is more detailed information (than is conveyed by locutions like "push", "pull", etc.) about the character of the relationship between *X* and *Y*—information of a sort that is expressed in a precise functional relationship or, failing that, qualitative information about the behavior of this relationship. Thus, rather than just being told that *X*s attract or repel *Y*s, one would like to know what the form of the force law linking *X* to *Y* is, or failing that, qualitative information of, for example, the following sort: does the force fall off rapidly or slowly with distance, is there a range outside of which it can be neglected for certain purposes, how does this force compare in strength with other forces? Even when causal explanation or analysis is relatively qualitative and nonmathematical (e.g., as it often is in molecular biology), this sort of qualitative information can be crucial and vaguer information to the effect that one thing attracts or repels another insufficient for understanding what is going on.

But now consider the implications of these observations for causal diversity. If attraction and repulsion are different "kinds" of causal relationship, how about attractive forces that obey different force laws? Is a force that falls off inversely with the square of distance a different "kind" of cause than a force that obeys an inverse fifth power law? Indeed, if, as in the case of electrostatic attraction/repulsion, there is a common force law that governs both, why not see regard them as involving the same kind of cause after all? In other words, why stop at the level of generality represented by "attracts" and "repels"? Why not a different kind of cause for each different functional relationship?

Assuming that this is not an attractive option, here is an alternative suggestion that seems to me to respect many of Cartwright's observations about causal diversity: Different cause–effect relationships conform to different functional relationships and of course those relationships have different qualitative characteristics as well. However, in each case those functional relationships will have a common inverventionist interpretation. Thus a gravitational force that obeys an inverse square law, electrostatic repulsion obeys a different inverse square law, and intermolecular forces are often modeled in terms of a force that falls off with the fifth power of the difference. But in each case, these laws may be interpreted as telling us how the force variable on their left-hand sides will respond to interventions on the variables on their right-hand sides. The differences among these force laws does, as Cartwright emphasizes, mean that different tests will be appropriate for them and that they will have different implications for policy/manipulation. However, neither of these points shows that that there isn't some nontrivial common causal interpretation shared by all these laws. Put slightly differently, I think that we should resist the inference from the observation that different causes have different modes of action, obey different functional relationships, and should be tested in different ways to the conclusion that there is no such thing as "the" causal relationship in the sense that there is no common interpretation of this relationship, whether along interventionist or any other lines.

I have emphasized in passing some of the benefits of such a general interpretation and the costs of giving up on the search for it. A general interpretation disciplines our thinking about causation, specifying what it is that we are trying to discover when we ask whether a relationship is causal, and distinguishing this from other equally general possibilities that may be on offer. For example, it ought to provide a general answer to the question of how causal claims differ from claims about the obtaining of correlations. This is important in areas of inquiry, like the social sciences, in which there may be considerable disagreement about just what we are committing ourselves to when we claim that a relationship is causal. A general interpretation also provides a principled basis for decisions about formal constraints, such as transitivity, that many have wanted to impose on causation.[19] It can also link causal claims to general, nonsubject-matter-specific ideas about testing and experimental design. So a general interpretation seems to me worth pursuing, especially if, as I have suggested, we can capture what is defensible in the notion of causal diversity by focusing on the diversity of functional relationships.

NOTES

1. A picture that also is suggested by Cartwright's well-known slogan, "no causes in, no causes out".
2. For a more detailed discussion of the difference between these two sorts of causal claims, see the treatment of direct and total effects in Pearl (2000) and the discussions in Hitchcock (2001b) and Woodward (2003).
3. Why the restriction to a single intervention on *X*? Suppose that *X* does not cause *Y* but that whenever an intervention occurs that changes *X* to some value, a second intervention also occurs that changes *Y* to some particular value. Then *Y* changes systematically under interventions on *X*, even though there is no causal relationship between *X* and *Y* .Cartwright (2002) appeals to just such an example to criticize a proposal about the connection between causation and manipulation in Hausman & Woodward (1999). We may exclude such counterexamples by requiring that for *X* to be a total cause of *Y*, *Y* must change under a single intervention on *X* and no others.
4. It is *not* true, however, that *X* is a contributing cause of *Y* as long as there is a directed path from *X* to *Y*. If there are intermediate variables V_i along this path between *X* and *Y*, the functions linking *X* to these variables V_i and V_i to *Y* may compose in such a way that there is no overall sensitivity of changes in the value of *Y* to changes in the value of *X*; see Hitchcock (2001a) and Woodward (2003: 57) for additional discussion.
5. We also value the discovery of causal generalizations that hold not just over any old arbitrary set of interventions, but instead hold over "large" ranges of interventions that satisfy additional constraints having to do with "continuity" and "importance"—(see Woodward 2000, 2003; Hitchcock & Woodward 2003).
6. In this respect, at least, the notion of a locally (but not universally) invariant generalization ought to be congenial to those, like Cartwright, who are skeptical about the universality of even fundamental physical laws.
7. Iain Martel has objected that Lashley's procedure can be explained on many alternative accounts of causation besides the interventionist account. For example, counterfactual accounts can think of Lashley as looking at what happens in situations that are as "close" as practically achievable to an ideal situation in which the cause alone does not occur; accounts that treat causes as necessary conditions for their effects will see Lashley as trying to determine whether visual cortex is necessary for the maze task, etc. My reply: we know on independent grounds that the necessary condition analysis of cause is unsatisfactory. With respect to the counterfactual account, one may think of the interventionist account as spelling out more precisely by means of the characterization of interventions just what it means to say that one possible world is "close" to another (see Woodward 2003: Ch. 3, for additional discussion).
8. There is an additional consideration that is relevant here. The design favored by Cartwright assumes that the effect of the drug when the manipulation potentially associated with the placebo effect is present is the same as the effect of the drug when this manipulation is absent. It is possible that this assumption is false—that is, that the drug is only effective when the patient believes that he or she has received it. If so, the factor that causes recovery is not the drug alone but rather the combination of the drug and the patient's belief. It is a limitation of the design Cartwright describes that it will not detect this possibility if it is present. The design recommended by the interventionist framework, in which the drug is introduced into the subjects in a way that does not affect recovery by any other route, does not have this limitation and is in this respect normatively preferable. More generally, we can see from this

consideration that the design in which the potential placebo effect is present in both the treatment and control is relevant to the assessment of the efficacy of the drug only insofar as it is correct to think of this design as telling us what would happen if the placebo were absent from both the treatment and control groups.

9. An illustration is provided by the notion of "Granger causation" which is widely employed in econometrics. Roughly speaking, *X* Granger causes *Y* if *X* is temporally prior to *Y* and information about *X* improves our ability (relative to some baseline) to predict whether *Y* will occur. Interestingly, Granger causation turns out to be a different notion of cause (and hence to be associated with a different notion of causal correctness) than the interventionist notion. *X* can be a Granger cause of *Y* even though it is not a cause in the interventionist sense (see Hoover 1988; Woodward 1995, 2003). It thus is a live question whether we should adopt this notion of cause instead of the interventionist notion.
10. Pearl characterizes this notion as follows:

    > The simplest type of external intervention is one in which a single variable, say X_i, is forced to take on some fixed value x_i. Such an intervention, which we call "atomic" amounts to lifting X_i from the influence of the old functional mechanism $x_i = f_i(pa_i, u_i)$ and placing it under the influence of a new mechanism that sets the value x_i while leaving all other mechanisms unperturbed. (Pearl 2000: 70)

11. There is a subtlety here that is worth noting. If we think of the mechanism associated with the spring as what is described by Hooke's law, then an extension that breaks the spring will disrupt this mechanism and hence will not count as an intervention, given the mechanism-preserving requirement. On the other hand, we might also think of the behavior of the spring as described by a more complicated generalization *G* that specifies that if it is ever extended beyond a certain length, the restoring force is subsequently always zero, that if it has never been so extended, it conforms to Hooke's law within a certain range of extensions, etc. If we think of the mechanism associated with the spring as given by *G*, an extension that breaks the spring does not disrupt this mechanism and hence may count as an intervention. Hence, on the mechanism-preserving requirement, whether a manipulation of *X* is an intervention will depend on the level of detail characterizing the function specifying the mechanism in which *X* figures as a cause. This illustrates one respect in which the mechanism-preserving requirement seems to build information about the relationship between *X* and its effects into the characterization of an intervention on *X*.
12. This is apparent in Cartwright's proof in her Theorem 1 (Cartwright 2003) that invariance under interventions implies causal correctness. Because she in effect requires that an intervention on *X* preserve the causal relationship, if any, between *X* and *Y*, the invariance of this relationship under interventions follows trivially from her characterization of interventions.A similar point seems to me to hold for Pearl. He *defines* the notion of an intervention by taking the notion of a "functional mechanism" as primitive (see endnote 10) thus losing any possibility of using the former to characterize the latter.
13. An example: You see a wire running from a switch to a light and wonder whether flipping the switch causes the light to go on and off. You may not know whether this causal claim is true—that is what you want to find out—but it is a very plausible guess that *if* the position of the switch causally affects the light, it does so via the wire. Thus an experimental manipulation of the switch that involves severing the wire will not be illuminating for the purposes of determining whether the position of the switch affects the light.

14. This is the appropriate place to acknowledge a potential complication. Suppose that our manipulation of *X* is of such a character that it *creates* (or it looks as though it creates) a causal connection between *X* and *Y* where none existed previously—i.e the manipulation alters the causal powers of *X*, leading us to mistakenly conclude that *X* causes *Y*, even though it does not, at least in the original system of interest. Although the matter deserves more attention than I can give it here, my inclination is to think that alleged cases of this sort will fall into one of two categories. First, (i) some will be cases in which the manipulation causally influences *Y* via a path that does not go through *X* so that the manipulation does not count as an intervention on *X* with respect to *Y*, according to **W/SIN**. As an illustration, suppose one attempts to test the claim that an (uncharged) mass exerts a gravitational on influence on another mass (which happens to be charged) by adding massive charged particles to the former. There is a sense in which this endows the first mass with a new causal power, but this is accomplished by means of a manipulation that exerts an influence (an electromagnetic force) on the second body which is independent of any gravitational attraction between the two—hence the manipulation does not qualify as an intervention. A second possibility (ii) is that the manipulation introduces a new mechanism between *X* and *Y* where none previously existed but not in a way that is naturally describable as falling into category (i). For example, a government might choose to observe the value of *X* (sea levels in Venice) and depending on this value alter *Y* (bread prices in Britain) in some systematic way (e.g., by imposing a tax). Or an experimenter might introduce a pipe between X_2 and X_3 in the salt chamber example where none previously existed. To the extent that such cases do not fall under category (i), one way of dealing with them would be to impose an additional restriction: An intervention on *X* with respect to *Y* should not introduce new variables along the causal path (if any) between *X* and *Y*. The background to this restriction is that we employ a certain stock of variables to model the causal system of interest—these will correspond to the possibilities of change or intervention that we are willing to take seriously. The British government does not at present set taxes on bread in accord with Venice sea levels, and we do not regard it as a serious possibility that it will ever do so. Thus in the representation of the original system associated with these variables, it would be illegitimate to introduce a variable *T* representing British bread tax policy and draw an arrow from *X* to *T*. If such a tax is introduced in such a way that it is causally between *X* and *Y*, this will represent a new causal structure, different from the original one.
15. Again, I claim no originality for this requirement. The basic idea is closely related to the econometric notion of autonomy (see Alderich 1989). The idea that each equation in a system of structural equations should correspond to a distinct mechanism and that distinct mechanisms should be independently disruptable has also been defended by Judea Pearl in a number of papers—(see Pearl 2000).
16. The Causal Markov condition says that conditional on its direct causes, every variable is independent of every other, except possibly its effects. Cartwright has been a persistent critic of this condition—see Cartwright (2002), and for additional discussion, Spirtes et al. (2000); Hausman & Woodward (1999, 2004, 2005).
17. Or alternatively, it may be that there is no sharp distinction between what is part of the concept of causation and what is part of the background for its usual application.
18. A similar dialectic appears to be at work in Cartwright's criticisms of the Causal Markov Condition (CM) and her well-known (hypothetical) example

involving a chemical factory in which (she claims) CM fails (see Cartwright 2002). In this example, the state *C* of the factory is the common cause of both a certain chemical (*X*) and a pollutant (*Y*) which is produced as a byproduct, but according to Cartwright even full information about *C* fails to screen off *X* from *Y*. One way of specifying what is distinctive about that example is this: the behavior of *C* in producing each of the joint effects *X* and *Y* in isolation (i.e. information about the probability with which *C* produces *X* (*Y*) when the production of Y (*X*) is suppressed) does not tell us how *C* will behave in the "intact' system in which both *X* and *Y* are produced together. Moreover, this is not an ordinary case of "emergence" in which one causal factor produces new effects because of its interaction with a second causal factor—the change in behavior of *C* when it produces both effects together is not the result of some internal change in *C* or the result of the presence of some new causal factor with which *C* interacts. Interestingly, no one has been able to produce an uncontroversial actual example of a macroscopic systems having this sort of structure. Moreover, it is not clear that the microphysical systems that are sometimes alleged to violate CM (such as the atomic decay example described by Iain Martel (this volume) are true counterexamples because it is not clear that they really have a common cause structure. In particular because the decay products will be in an entangled state or a superposition, there is no possibility of intervening to disrupt the relationship between the candidate common cause and one of its effects without disrupting the relationship between the common cause and the other effect. The modularity criterion for the distinctness of these causal relationships is thus violated (see section 'Modularity'). Martel holds that such microphysical examples do have a common cause structure and, in illustration of the point that one man's modus *ponens* is another's *modus tollens*, takes the examples to show that the modularity criterion is mistaken. However, Martel provides no basis for his claim that the examples have a common cause structure, other than that this seems "clear" to him and some others. My view is that philosopher's intuitions are a dubious guide to causal structure in the microphysical realm. In the case under discussion there are good reasons, rooted in quantum mechanics itself, for denying that the occurrence of the decay products and the processes generating them are distinct in a way that supports the claims that the examples have a common cause structure. Martel's claims would be more persuasive if he were to articulate and defend an alternative set of criteria for the presence of a common cause structure and show how the physics of the situation supports this analysis.

[19] For an argument that the imposition of a transitivity requirement on causation is mistaken, see Woodward (2003).

REFERENCES

Alderich, J. (1989) 'Autonomy', *Oxford Economic Papers* 41: 15–34.

Anscombe, G. (1971) 'Causality and determination', reprinted in E. Sosa and M. Tooley (eds) (1993) *Causation*, Oxford: Oxford University Press.

Bogen, J. (2002) 'The opposite of counterfactual is factual', paper read at PSA workshop. Available HTTP: <philsciarchive@philsci-archive.pitt.edu.

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Clarendon Press.

———. (1979) 'Causal laws and effective strategies' reprinted (1983) in *Nous*, 13: 419–437.

———. (1999) 'Causal diversity and the Markov Condition', *Synthese*, 121: 3–27.

———. (2001) ‘Modularity: It can—and generally does—fail’, in M. Galavotti et al. (eds) *Stochastic Causality*, Stanford: CSLI Publications.

———. (2002) ‘Against modularity, the causal Markov condition, and any link between the two: Comments on Hausman and Woodward’, *British Journal for the Philosophy of Science*, 53: 411–453.

———. (2003) ‘Two theorems on invariance and causality’, *Philosophy of Science*, 70: 203–224.

Cartwright, N. and Jones, M. (1991) ‘How to Hunt Quantum Causes’, *Erkenntnis* 35: 205–31.

Cook, T. and Campbell, D. (1979) *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Boston, Houghton Mifflin.

Frisch, R. (1938, 1995) ‘Autonomy of Economic Relations’, Reprint, *The Foundations of Econometric Analysis*, ed. D. Hendry and M. Morgan, Cambridge, UK: Cambridge University Press.

Gopnik, A. et al. (forthcoming) ‘A theory of causal learning in children: Causal maps and Bayes nets’, *Psychological Review*.

Haavelmo, T. (1944) ‘The Probability Approach in Econometrics’ *Econometrica* 12: 1–118.

Hausman, D. and J. Woodward. (1999) ‘Independence, invariance, and the causal Markov condition’, *British Journal for the Philosophy of Science*, 50: 521–583.

———. (2004) ‘Modularity and the causal Markov condition: A restatement’, *British Journal for the Philosophy of Science*, 55: 147–161.

———. (2005) ‘Manipulation and the causal Markov condition’ *PSA 2002*, vol 2: 846–856.

Hitchcock, C. (2001a) ‘The intransitivity of causation revealed in equations and graphs’, *The Journal of Philosophy*, 98: 273–299.

———. (2001b) ‘A tale of two effects’, *Philosophical Review*, 110: 361–396.

Hitchcock, C., and J. Woodward. (2003) ‘Explanatory generalizations, part II: Plumbing explanatory depth’, *Nous*, 37: 181–199.

Hoover, K. (1988) *The New Classical Macroeconomics* Oxford: Basil Blackwell.

Pearl, J. (2000) *Causality.: Models, Reasoning and Inference*, Cambridge: Cambridge University Press.

Spirtes, P. et al. [1993] (2000) *Causation, Prediction, and Search*, 1st edn, New York: Springer-Verlag.

Stotz, R. and Wold, H. (1960) ‘Recursive vs. Non-Recursive Systems: An Attempt at a Synthesis’ *Econometrica* 28: 417–27.

Woodward, J. (1995) ‘Causality and Explanation in Econometrics’ In On the Reliability of Economic Models: Essays in the Philosophy of Economics, ed. D. Little, Dordrecht: Kluwer, 9–61.

———. (1999) ‘Causal Interpretation in Systems of Equations’, *Synthese* 121: 199–257.

———. (2000) ‘Explanation and invariance in the special sciences’, *British Journal for the Philosophy of Science*, 51: 197–254.

———. (2002) ‘Counterfactuals and causal explanation’, paper read at PSA workshop. Available HTTP: <philsciarchive@philsci-archive.pitt.edu.

———. (2003) *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.

Reply to James Woodward

Jim Woodward has provided us a monumental review of his views on causality and of many of the disagreements he and I have had about them. I suppose that the most efficient thing I can do in reply is to explain directly what I believe about many of the same issues. As with Woodward's discussion here, I will consider only systems of linear equations and their causal interpretation. In particular I will focus on what I call "epistemically convenient linear deterministic systems". They look like this:

$$x_1 \ c = u_1$$

$$x_2 \ c = a_{21}x_1 + u_2$$

$$x_n \ c = \Sigma a_{n1}x_n + u_n$$

where causes are on the right and effects on the left and where each *u* can take any values in its range simultaneously with any combination of values for the other *u*s. This means that Woodward's interventions are always possible since each right-hand-side variable can vary independently of all others.

With respect to systems like these, I learned an exciting fact from Woodward—it seems that the invariance of an equation under interventions on its right-hand-side variables is sufficient for an equation to be causally correct. I didn't know that, and it seems an important fact for testing for causality. Woodward asserted this claim repeatedly, but his arguments tended to be by example.[1] He constructed a few cases where linear combinations of causally correct equations give rise to causally incorrect ones, and in those cases the causally incorrect ones were not invariant.[2] But how do we know that *every* such linear combination will fail to be invariant? That there is no way for an equation to be invariant if it is a nontrivial linear combination of the basic, or "causal", equations? Woodward says that that is just what causality is. But his own method of constructing the examples suggests not. Suppose we did find cases where funny linear combinations of the equations we regard as causal, cases like the barometer and the storm, turned out to be invariant. Would we then wish to regard them as causal? Happily this cannot happen. That is shown by my representation theorem, which Woodward mentions.[3]

The theorem supposes a number of axioms, like linearity, irreflexivity, and antisymmetry, which allow the equations to be written in the familiar triangular form above. Woodward complains that apart from the axioms I don't say anything about what causality means other than that causes are on the right and effects on the left. We need to say the latter to connect with Woodward's notion of intervention. But we don't need to say anything more. The point of such a proof is that for *any* system that looks like this, invariance under intervention on right-hand-side variables is sufficient to ensure that an equation is one of the basic equations of the system and not a linear combination, as is the equation relating the barometer drop and the storm. Nicely, one of the other different criteria that Woodward stresses can also be shown to hold: If the value of y varies when x varies by intervention, then x must appear nontrivially on the right-hand side of some basic equation in the system—i.e. it must appear as a cause in one of the "causal" equations.

To repeat, for *any* system that satisfies the axioms, both criteria will hold. The results depend on the abstract structure of the system and not on any more specific meaning one wants to attribute to causality.

We must be careful how to read these results. Woodward thinks that this special kind of invariance constitutes causality. On the other hand he also talks about the fact that effects should vary only when their causes vary, as he should, since this is a standard test for causality. It would be nice from his point of view therefore if we could show that whenever each of a set of equations otherwise unrestricted is invariant under right-hand-side interventions—and is hence causal on Woodward's definition—the second criterion holds as well. That is not what I have shown. Rather my representation theorems show that whenever this specific set of axioms is satisfied, both criteria hold—both depend on the structure postulated in the axioms and not on each other.

Most of the axioms simply ensure that the equations have the triangular (acyclic) structure we see above. Two do not. The first is to the effect that the only equations that are taken to hold are either the basic ones or linear combinations of these. This means that there are no functional relationships that do not have their origins in the basic equations, the equations we label "causal". This may not be true in reality in many situations, but it is clearly necessary if the invariance test is to be used.

The second is one that Woodward does not like—numerical transitivity. Consider a causal equation, $z \ c= f(\ldots, y)$ and also the causal equation for y itself, $y \ c= g(x, \ldots)$. Numerical transitivity also counts as causally correct the result of substituting the equation for y into that for z: $z \ c= f(\ldots, g(x, \ldots))$. I adopt this for three reasons. First, as Woodward suggests, I think that for many situations the idea of direct causation *in the world* doesn't make sense—for many situations between any cause and its effect there will be an intermediate. The second is because even where there are direct causes we are very often not in possession of them. In the above formula, we might never have thought about *y*. Still, it is important to know the equation relating *x* and *z* and to know that it has a very different status than that relating the barometer drop and the storm. Focus only on direct causation leaves no place for this important law. Third, not allowing the law relating *z* and *x* to count as a causal law undermines Woodward's own very important contribution—for it is invariant under interventions on right-hand-side variables![4]

Let us turn now to other kinds of invariance. My own view is that most causal relations we study are not absolutely fundamental. Rather they hold because something else is true. Both Woodward and I have for years discussed Tyrgve Haavelmo's example: Stepping on the throttle makes the car go faster. That is a causal relation if ever I have heard of one. But God did not make it true. It holds on account of the way the car is built. I call the arrangement of the parts of the car a *nomological machine*. It is responsible for the fact that the causal relation between throttle pressure and acceleration holds.

One of the important lessons that Woodward teaches, along with Sandra Mitchell and economists like David Hendry and Kevin Hoover, is that if we want to use causal knowledge to predict what will happen under some given manipulation, it had better be stable under that manipulation. That means that, among other things, the manipulation had better not disrupt in the wrong ways the arrangements responsible for that causal relation to hold in the first place.

This lesson points out an important distinction that sometimes gets blurred in Woodward's discussions. He talks of a "manipulation" account of causality, but "manipulation" in two senses. The first involves the kind of manipulations that we need to have to test a causal claim—what he calls "interventions". In this case it must be part of the definition of the manipulation that it not disturb the underlying arrangements that make the causal relation hold if it does hold or fail to hold if it does not. Otherwise true causal relations can fail the test and false ones pass it. But for use we are interested in real manipulations. What happens if we do what we actually intend to do? These may very well disturb the underlying arrangements. So maybe we will not be able to make use of our causal knowledge in the way we had hoped.

What matters if we are going to use a claim for prediction, whether the claim is a causal one, or a conservation "law" or a statement of co-association, or whatever, and whether we are actually going to make changes ourselves or just let matters flow as usual, is that the very claim we use for prediction be true in the situation that will transpire. That is what Sandra Mitchell teaches and it is sometimes what Woodward stresses. But that is clearly neither intervention on right-hand-side variables nor what he calls "modularity".

What then of modularity? It seems to me it has nothing to do with causality or with real manipulability. It requires that each causal law in a set be changeable without changing any other.[5] This is not enough to guarantee that we can use any of the causal laws for real manipulation or that they will be stable for prediction. Yet it seems too strong for causation.

Of course sometimes such failures are due to the fact that one law cannot be changed without changing the underlying arrangements, which changes then percolate back up to affect other laws. These are the kinds of cases Woodward and I have dwelt on and the kinds that have formed the basis for the famous Lucas critique of using well-established macroeconomic relations as a basis for prediction of what will happen under new policies. Woodward and I both agree that if we hope to use our causal knowledge in ways that might affect the very arrangement responsible for its being true, it would sure help to know about these arrangements. Also I see that if a second law changes when a first does that can be important to know, and it can be important even if the reason is not that they emanate from an underlying structure that is itself affected by the change in the one law. But I do not see why that is so privileged or why it is special to causality. Woodward himself teaches what Mitchell teaches: *Anytime* we want to use a claim *of any kind* (causal or not) for prediction, it is important to know that that claim will hold in the circumstances where we make the prediction.

I conclude from this that we should pay more attention to the claims of Mitchell and Woodward about invariance in general and worry less about causality. It is nice to have causal knowledge.[6] But it won't help for prediction unless it remains true in the circumstances the predictions are about. But then any claim that will remain true in the targeted circumstances can be used for predictions about those circumstances whether it is causal or not.

NOTES

1. Of course he made lots of independent arguments about the general importance of "invariance", but there are huge numbers of different kinds of invariance, and it is this special one that is needed for causality.
2. For instance where joint effects of a common cause are functionally related but the functional relation is not invariant under intervention. Because they are both effects of low pressure, the storm occurs iff the barometer drops. But this correlation breaks down if we smash the barometer.
3. REF—my PSA paper on two theorems + paper on Suppes.
4. Woodward urges that it is not invariant under interventions on y. But what if we have never heard of *y*? I do not see exactly how he hopes to formulate the invariance condition. Should an equation be invariant under interventions on all variables we have ever thought of? On all variables that are *in fact* in the system? What would that mean? All variables that are causes or effects of all the variables we are interested in? Or . . . ?
5. I am also always puzzled about what the set is supposed to have in it.
6. Why? For one thing, it is important for the ascription of responsibility. For another, it can tell us how to build things, but in this case, in my view, it better be knowledge of capacities and not just of causal laws. Also, it can tell us how to fix things, for instance, failures of the cause to produce the effect often result from some break in the process connecting them; or it can tell us how to attenuate effects we do not like or enhance those we do.

10 The Principle of the Common Cause, the Causal Markov Condition, and Quantum Mechanics

Comments on Cartwright[1]

Iain Martel

Nancy Cartwright believes that we live in a Dappled World—a world in which theories, principles, and methods applicable in one domain may be inapplicable in others and in which there are no universal principles. One of the targets of Cartwright's arguments for this conclusion is the Causal Markov condition, a condition which has been proposed as a universal condition on causal structures.[2] The Causal Markov condition, Cartwright argues, is applicable only in a limited domain of special cases and thus cannot be used as a universal principle in causal discovery. I have no dispute with any of these claims here. Rather, I wish to argue for a very limited thesis: that the Causal Markov condition *is* applicable in the specific domain of microscopic quantum mechanical systems; further, that the condition can fruitfully be applied to the much discussed EPR setup. This is perhaps a surprising conclusion, for it is precisely in this domain that Cartwright's arguments against the Causal Markov condition have been considered to be the most successful.

I begin with a review of the Causal Markov Condition (hereafter **CM**) and Cartwright's main argument against it. Cartwright's argument is seen to be primarily an attack on the related Principle of the Common Cause (hereafter **PCC**), which is often thought to be entailed by **CM**. I then show that, when suitably generalized to accommodate continuous models, **CM** does not, in fact, entail **PCC**. Next, I show how this generalized version of **CM** can be used to build a common cause model for the EPR setup, which does not falsely entail the Bell inequalities. Finally, I argue (against Cartwright) that the propagation condition built into these models is appropriate for the quantum domain and that the proposed model is thus superior to Cartwright's own, nonpropagating common cause models for EPR.

CM and PCC

The Causal Markov condition is a generalization of the screening-off condition common in probabilistic theories of causation. As defined by Daniel

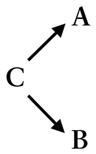

Figure 10.1

Hausman and James Woodward (following Spirtes et al. 1993) in a recent paper, **CM** says the following:

> **CM** (The Causal Markov Condition): For any system of acyclic causal relations, there exists some set of variables **V** such that for all distinct variables X and Y in **V**, if X does not cause Y, then Pr(X/Y & **Parents**(X)) = Pr(X/**Parents**(X)).[3] (Hausman & Woodward 1999: 529)

This condition states that, conditional on the set of its immediate causes (its parents), an event X is probabilistically independent of all other events except for its effects. As Cartwright puts it, this condition (as it stands) combines two demands: one, 'a prohibition against causes exerting influences across temporal gaps', the other a screening-off condition which is a generalization of Reichenbach's Principle of the Common Cause—'a full set of parents screens off the joint effects of any of those parents from each other' (Cartwright 1999: 107). It is this last aspect of **CM** which attracts most of Cartwright's attention, so let us look at it more closely.

Consider the causal structure illustrated in Figure 10.1. In this diagram, C is the only parent of both A and B. Thus, if this is the complete causal story, **CM** entails that A is independent of B, conditional on C. That is, the common cause C screens off the correlation between A and B. Thus, for simple causal structures, **CM** entails **PCC**: unconditionally correlated events are uncorrelated, conditional on their common cause. A counterexample to **PCC** involving a simple causal structure such as that in Figure 10.1 will thus also count as a counterexample to **CM** (again, as formulated). And the counterexamples Cartwright presents are of precisely this form. To make the later presentation simpler, however, the counterexample I present will not be Cartwright's, though it has the same form.[4]

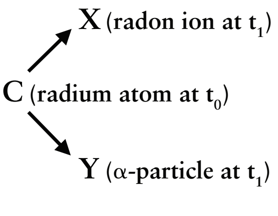

Figure 10.2

Consider the causal fork in Figure 10.2. A radium atom at time t_0 (*C*) decays to produce a radon ion at t_1 (*X*), while emitting an alpha particle (*Y*). We assume that there are no further causes of the decay of the radium atom: that it was entirely indeterministic. Suppose that the probability of the radium atom decaying is 0.5. Then $\Pr(X/C) \approx \Pr(X/C) \approx 0.5$. (The approximate equalities here are important, though normally they are overlooked. Even if the radium atom decays at t_0, it is not certain that there will be a radon ion at t_1, for the radon ion itself could decay. Similarly, the alpha particle could be absorbed by some other atom, or destroyed.) Given these conditions, **CM** requires that Pr(*X*/*Y* *&* *C*) is also approximately 0.5. But this is not the case: if there is an alpha particle at t_1, then the radium atom must have decayed, in which case there will almost certainly be a radon ion at t_1 also. Thus, contrary to **CM**, $\Pr(X/Y \,\&\, C) \approx 1 \neq \Pr(X/C)$. The common cause does not screen off the correlation between its effects.

Before proceeding, I should note some common objections to this sort of counterexample. In their paper, Hausman and Woodward suggest four different strategies for responding to alleged violations of **CM**. The first strategy is to deny that the common cause has been characterized in sufficient detail. If the specification of the common cause leaves out relevant factors, then screening off will fail. A more precise specification of the details of a process may, Hausman and Woodward argue, be needed to uncover a screening-off common cause (Hausman & Woodward 1999: 561–562). But this kind of response is not available in quantum mechanical cases of this sort, for the decay process is supposed to be entirely indeterministic—any radium atom has the same probability of decay in a given time interval, and no further specification can change that probability.

The second strategy is to claim that the alleged common cause is not, in fact, the true common cause but, rather, a precursor to the common cause. In such a case, the precursor will not screen off the correlation between its effect, but the true common cause might. In the present case, the claim would be that it is the decay of the radium atom that counts as the common cause, rather than the existence of the atom before the decay. Given that the decay takes place, the probability of finding a radon atom is close to one, and this probability is independent of whether or not an alpha particle is found. The problem with this response is that it rests on an equivocation on what it is to be an *event*. There is, of course, a perfectly good intuitive sense in which the decay of an atom counts as an event. But this is not the sense of "event" that is relevant to assessments of **CM**.[5] For **CM** is stated in terms of *variables*, or momentary states of systems. The events relevant to discussions of **CM** are thus the states of distinct spatiotemporal regions. And we will never find a momentary state of the system which is *the* decay of the radium atom—at any moment, either there is an undecayed radium atom, or there is a radon ion and an alpha particle. In the former case, the situation is just as originally described, and screening off fails. In the latter case, the causal fork has already occurred, so we have a correlated pair of events themselves in need of a causal account. In neither case will we find the elusive screening-off common cause.

The third strategy used by Hausman and Woodward to counter putative counterexamples to **CM** is to argue that the two effects of the common cause are not really distinct events, that is, that there really is only one effect, differently described. We clearly would not expect *any* event to screen off a correlation between an event and *itself*! In the present case, the idea would be that the emission of an alpha particle and the emission of a radon atom are simply different aspects of a single event ('the decay of the radium atom'?). But this response involves the same equivocation over events as the last: By referring to *emissions*, the objection invokes the active notion of "event", rather than the momentary state notion relevant to **CM**. If we stick to the latter notion, it is hard to see how the existence of a radon ion at a particular time and location could be counted as identical to the existence of an alpha particle at that time and (some different) location![6]

Finally, Hausman and Woodward's fourth strategy for responding to alleged counterexamples to **CM** is to accept that there is an unscreened correlation between two events, but to hold that the correlation is not due to the operation of a common cause, but to some kind of 'non-causal (but non-accidental) relation to one another' (Hausman & Woodward 1999: 565). Indeed, this is their response to the kinds of correlations found in the EPR case, to be discussed in a moment. But in the present case, and this is one reason for starting with this case, this response is hardly credible. For surely there can be no clearer case of a causal fork than the present example. The radium atom existing at time t_0 is clearly the cause of the radon ion existing at time t_1, and it is also clearly the cause of the alpha particle's existence at that time. So it is not possible to deny that the correlation in this case is due to a common cause.[7] I conclude that the counterexample is a genuine one: **PCC** is false.

CM WITHOUT PCC

The foregoing discussion shows, I believe, that Cartwright is right to insist that **PCC** fails for causal structures such as that in our example, where an indeterministic common cause produces its effects together, subject to constraints such as conservation laws. I pass no judgment on how serious a problem this is for those, like Spirtes et al., who believe the problem can be limited to the quantum domain, and is not found in the macroscopic world—I am interested specifically in the quantum domain, and here at least Cartwright's criticism definitely applies. But does this also mean that the Causal Markov condition cannot be applied at the microlevel? I wish to show that the answer is no, at least once **CM** has been modified to allow for continuous models.

To see how **CM** can be separated from **PCC**, it is important to note that the screening-off relations implied by **CM** are, in general, mostly of a

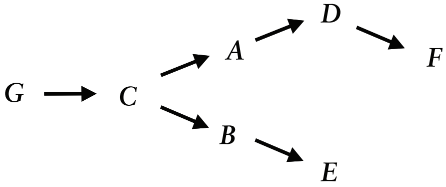

Figure 10.3

different kind to the common cause structures we have been considering. Consider the more complicated causal diagram in Figure 10.3. **CM** entails a number of screening off relations between events in this diagram. For example, **CM** entails that *A* screens *D* off from *C*, and also from *B*, *E*, and *G*. But *A* is not a common cause in any of these cases. In particular, we should note that **CM** *does not* directly entail that *C*, which *is* a common cause, screens off the correlation between *D* and *E*, for *C* is not a Parent of *D*. Rather, **CM** entails that *A*, the causal intermediary between the common cause *C* and effect *D*, screens off the correlation between *D* and *E*. As must *B*, the causal intermediary between *C* and *E*.

This too-infrequently noted fact about the precise entailments of **CM** suggests a way of restoring screening off, even if **PCC** fails. The idea is to look for events other than the common cause to provide the screening off between the effects of a common cause. Cartwright herself outlines one way of attempting this:

> Imagine that a particular cause, *C*, operates at a time *t* to produce two effects, *X* and *Y*, in correlation, effects that occur each at some later time and some distance away, where we have represented this on a DAG [causal graph] with *C* as the causal parent of *X* and *Y*. We know there must be some continuous causal process that connects *C* with *X* and one that connects *C* with *Y*, and the state of those processes at any later time *t2* must contain all of the information from *C* that is relevant at that time about *X* and *Y*. Call these states P_X and P_Y. We are then justified in drawing a more refined graph in which P_X and P_Y appear as the parents of *X* and *Y*, and on this graph the independence condition will

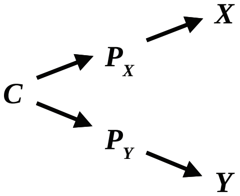

Figure 10.4

> be satisfied for X and for Y (although not for P_X and P_Y). Generalising this line of argument we conclude that any time a set of effects on an accurate graph does not satisfy the independence condition, it is possible to embed that graph into another accurate graph that does satisfy independence for that set of effects.[8] (Cartwright 1999: 115–116)

Before considering Cartwright's objection to this line of argument, let us see how it is supposed to work for the decay example we have been considering. As before, take C to be the existence of the radium atom at t_0, the moment before it actually decays. Take X to be the existence of a radon ion at time t_1, after the decay. And take Y to be the existence of an alpha particle at time t_1. We have already seen that $\Pr(X\&Y/C) > \Pr(X/C).\Pr(Y/C)$; that is, C does not screen off X from Y. But now, applying the strategy just discussed, let us consider the event $X_{1/2}$: the existence of a radon ion at time $t_{1/2}$—again, after the decay (see Figure 10.5).[9]

Given this event, the fact that the radon atom was produced by the decay of a radium atom is no longer relevant to the probability of finding a radon atom at time t_1—all that is relevant to this probability is the probability of the radon atom itself decaying between $t_{1/2}$ and t_1. For the same reason, the existence of the alpha particle at t_1 is also no longer relevant to the probability of finding the radon atom at t_1, for its only relevance was as an indicator of the decay of the radium atom. Thus, $X_{1/2}$ screens off the correlation between X and Y. Furthermore, since $X_{1/2}$ is a parent (cause) of X, **CM** is satisfied: $\Pr(X/Y\&\text{Parents}(X)) = \Pr(X/\text{Parents}(X))$.

Now, as Cartwright notes, this strategy will need to be applied repeatedly. For note that there is a correlation between our event $X_{1/2}$ and event Y. And, for the same reasons as before, it should be clear that C will not screen off this correlation. To still satisfy **CM**, there will thus have to be some further event, $X_{1/4}$, between C and $X_{1/2}$, which screens off the correlation between $X_{1/2}$ and Y. This, of course, leads us to consider the correlation between $X_{1/4}$ and Y, and so on. At every stage, we can add a further event which screens off the unscreened correlation found at the previous stage.

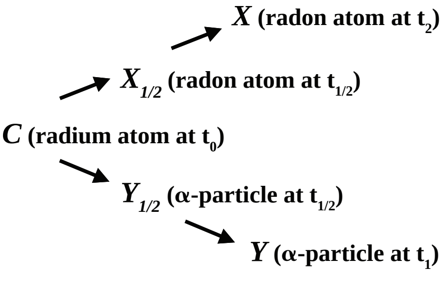

Figure 10.5

We should note, however, that this strategy does not yet rescue **CM** from counterexample, even supposing Cartwright's objections to be averted; see below. For each new model only removes one counterexample at the expense of the addition of a new one—at each stage, there remains an unscreened correlation between effects of a common cause, such that **CM** is violated. Indeed, so long as the model is finite, it is clear that there must always be an event X_n such that C is the parent of X_n in the model and that **CM** will fail for this event. To solve this problem, we need to move to a continuous (dense) model, in which an infinite sequence of events connects C with X, such that for every pair of events in the sequence, there is a further event between them. In such a model there will be no first event after C, just as, in the sequence of real numbers, there is no first number after zero. Thus, for any event in the sequence there will be an earlier event which screens it from Y.

Moving to continuous models of this kind, however, requires a modification of the Causal Markov Condition. For in a dense sequence, the notion of a parent no longer applies—there is no event immediately prior to any one given event. Rather, each event has an infinite sequence of ancestors, with none of them counting as *the* parent. So **CM**, as stated, will fail for such a model—since the class, Parents(X), is empty for all X in such a model, every correlated pair of effects of a common cause will count as a counterexample! But it is clear, intuitively, what the solution should be. Having replaced the parent of the finite model with an infinite sequence of ancestors in the dense model, we should reformulate **CM** in terms of some subset of these ancestors. Here is a proposal for a reformulated Causal Markov Condition:[10]

> **CM***: For any (microscopic) system of acyclic causal relations, there exists some (possibly infinite) set of variables **V** such that for all distinct variables X and Y in **V**, if X does not cause Y, then there exists some set of events **Ancestors** (X), such that **Ancestors** (X) $\in$ **Cause**(X) and Pr(X/Y & **Ancestors** (X)) = Pr($X/$ **Ancestors** (X))[11]

An event Z is a member of **Cause**(X) iff there is a path from Z to X in the causal graph.

My claim, then, is this: for cases such as the decay example discussed, **CM*** may still be satisfied, even if **PCC** fails. All we have to do is to construct a continuous model in which there is a dense, infinite sequence of events X_i forming the process from common cause C to effect X, such that for any X_n in the causal process connecting C and X, there is an X_m, m<n, such that $\Pr(Y/X_n \& X_m) = \Pr(Y/X_m)$.

Let us turn, now, to Cartwright's criticism of the sort of strategy we have been considering. The problem, Cartwright argues, is that 'it confuses claims about individual events that occur at specific times and places . . . with general claims about causal relations between *kinds* of events' (Cartwright 1999: 116). The claim that there must be continuous processes connecting each event concerns singular causal relations, whereas **CM** is stated in terms of generic causal relations, where, Cartwright argues, such connections may not be found. As an example, Cartwright cites the causal regularity between the signing of a cheque at one time and location and the giving out of cash by a bank at a different time and location. Though these events are cause and effect, there may be indefinitely varied ways of realizing the link between them, such that there need be no law-like regularity connecting the link-type events with the cause-kind events or the effect-type events. In such a case, Cartwright argues, a continuous model of the kind proposed will not be possible.

This criticism is directed primarily at the kind of macroscopic causal processes in which Cartwright and Spirtes et al. are interested. But does it also apply to the microscopic cases we have been discussing? Cartwright perhaps does not think so: 'For some causal laws the cause itself may initiate the connecting process, and in a regular law-like way. For these laws there will be intermediate vertices on a more refined graph' (Cartwright 1999: 116). I think that this is precisely the case in our decay example: The state of the particles at any given moment is connected by a law-like process to the states at later points along the process. And the experimental setup can be arranged such that there is only one possible causal path connecting initial and final states. In such a case, the requirement of causal continuity can appropriately be applied at the type level, and the defense of the applicability of **CM** to this case goes through.[12]

THE EPR EXPERIMENT

So far, all we have shown is that a modified form of **CM** can be applied in at least some experimental setups involving purely indeterministic common causes. By itself, this result is not particularly interesting. After all, even Cartwright acknowledges that **CM** may be an appropriate condition in *some* situations—her target is only the claim of universal applicability. But Cartwright has also defended the particular claim that screening-off conditions such as **CM** are inapplicable in the much-discussed EPR case in quantum mechanics. Using the ideas already developed, I will now argue that this particular claim is false: **CM** (**CM***) *is* an appropriate condition even for the EPR setup and can be used to produce a common cause model for the EPR correlations.

In the standard presentation of Bell's theorem, we are asked to consider a pair of particles prepared in what is known as the "singlet" state: $\Psi = 1/\sqrt{2}\ (z_{+}\otimes z_{-} - z_{-}\otimes z_{+})$. This is a state in which the total spin of the particle pair is zero. This tells us that if we measure the spin of one particle along a particular axis, then a measurement of the spin of the other particle along that axis will always yield the opposite answer. And, for measurements along different axes, quantum mechanics tells us that the two outcomes will be correlated in a certain strong way.[13] Furthermore—and this is what causes all the fuss—this correlation will appear no matter how far apart the particles are when they are measured. Given these facts, the expected causal structure is one in which the event of particle production (Z) counts as the common cause of the later correlated measurement outcomes (X & Y). What Bell showed was that the predicted correlations between the two measurements outcomes are too strong for any interpretation according to which the correlation is screened off by the event Z—the event of pair production—even if we allow for hidden variables giving a finer specification of event Z than that given by the quantum mechanical formalism. Using some version of Reichenbach's Principle of the Common Cause (**PCC**), it is then argued that quantum mechanics does not allow of a normal causal interpretation.

We have already seen, however, that this use of **PCC** is illegitimate. Any common cause in this case must act to produce its effects subject to the law of conservation of angular momentum. In such circumstances, as Cartwright argued, we should not expect the common cause to screen off the correlation between its effects. Perfectly normal cases of common causes do not satisfy the requirement of screening off, and so the failure in this case of that condition does not in itself reveal anything about the possibility of a causal interpretation of quantum mechanics—despite many arguments to the contrary. From what has been said so far, there is, structurally, no difference between the EPR case and the decay case we have been considering, and it is quite implausible to deny a causal structure there.

The standard arguments against a common cause account of EPR, then, rest on the unsound assumption of **PCC**. But this does not yet settle whether there really is a common cause account and if there is, whether it satisfies **CM**. To settle this question, we need to see whether the techniques used to model the decay example can be applied to EPR. In the decay case, the key was to look at intermediaries between the common cause and its distal effects. Suppose that we apply this idea to the EPR case, assuming that the common cause is at the source of the particle pair: See Figure 10.6.

Consider event $X_{1/2}$. This is the state of particle *a*, halfway from the source to the measurement apparatus. Can this state screen off the correlation between the measurement outcomes? If it did, then we would be able to give exactly the same common cause model for EPR as we did for the decay case. Unfortunately, the same arguments which are used to show that the state at the source cannot screen off the measurement correlation apply just as well here—given that $X_{1/2}$ may occur before the measurement settings have been

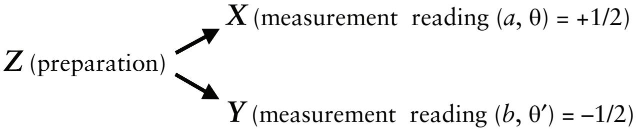
X (measurement reading (a, θ) = +1/2)
Z (preparation)
Y (measurement reading (b, θ′) = –1/2)

Figure 10.6

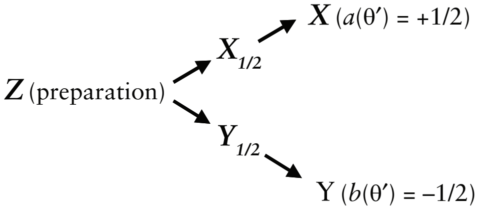

Figure 10.7

made, the state at $X_{1/2}$ could only screen off the measurement correlation if it determined probabilities for many different combinations of measurement settings, and this cannot be done without violating the Bell inequalities.

What this shows is that the argument for Bell's theorem need not depend on the false assumption of **PCC**. We can show that no common cause located at the particle source can account for the distant correlations, even assuming that **PCC** fails, simply by assuming a causal intermediary for that common cause and showing that it cannot satisfy **CM**.[14] But this does not get us what we were looking for—a working common cause model. We need to search again for our common cause.

Another look at our intermediary event $X_{1/2}$ provides the clue. For note that $X_{1/2}$ *does* screen off the correlation between *Z* and *X*; this is just the propagation part of **CM**—the part that Cartwright calls a Markov condition proper. The quantum state at time $t_{1/2}$ encodes all of the information of the quantum state at time t_0, thus screening that state off from the state at t_1. But we must be careful about what this screening-off state is. For quantum mechanics tells us that, when two systems become entangled in the way the two particles are in the EPR setup, it is incorrect to treat the two systems as having separate quantum states. There are not two states $\phi_{1/2}(a)$ and $\phi_{1/2}(b)$, but a single state $\Phi_{1/2}(a \,\&\, b)$. Looked at this way, the state at $X_{1/2}$ which screens off the correlation between *Z* and *X* is the same state (i.e. $\Phi_{1/2}(a \,\&\, b)$) as that at $Y_{1/2}$—which screens off the correlation between *Z* and *Y*. Given the existence of the entangled state $\Phi_{1/2}(a \,\&\, b)$ at time $t_{1/2}$, the preparation state *Z* is irrelevant to *both* the occurrence of measurement outcome *X* and the occurrence of measurement outcome *Y*. Thus the diagram should not be that in Figure 10.7, but that in Figure 10.8.

And this suggests that we have been looking in the wrong place for our common cause: The common cause is located *after* $t_{1/2}$, not before it!

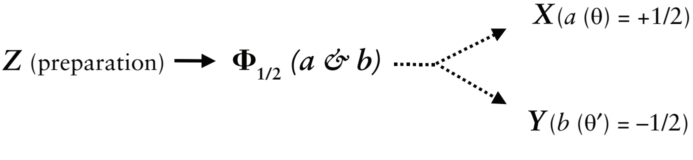

Figure 10.8

Moreover, this explains the failure of screening off: It can be proved that a *distal* common cause—the cause of the common cause—cannot screen off the correlation between its effects.[15]

With the hint that the common cause comes later in the story, and the idea that the intermediary states include joint superpositional states involving both particles, we are now in a position to find the sought-after common cause model. The idea is that the causal account is simply that given by the naive textbook presentation of the quantum mechanical story[16]: the preparation of the particles produces a particle pair in the singlet state; this joint superpositional state evolves over time as the particles separate; finally, a measurement is performed on one of the particles, and, thus, on the single superpositional state. This causes each of the particles to enter an eigenstate of spin along the axis of the measurement apparatus, and these now separate states cause the correlated measurement outcomes at the recording devices; see Figure 10.9.

In this picture, **CM*** is satisfied at every stage. As a propagation condition, it is satisfied in the evolution of the joint quantum state from the time of the state preparation to the time of the first measurement. At this point, the joint state collapses into two separate states, each of which evolves independently. The causal structure of this part of the picture is thus just the same as that of the decay case—while the common cause Z_1 itself still does not screen off the correlated post-collapse states, any post-collapse state of particle *a* is screened off from any post-collapse state of particle *b* by some earlier post-collapse state of particle *a*[17] (e.g., *X* is screened off from *Y* by $X_{1+\in}$, which is itself screened off from *Y* by some still earlier event $X_{1+\in'}$, and so on). Once again, **CM*** is satisfied. We thus have, as promised, a common cause model for EPR satisfying **CM***.

Now, there will of course be a number of objections to this model—it wouldn't be a model for quantum mechanics if it didn't have its share of "absurd" consequences! I will briefly consider three main objections, and compare the account here with a different common cause model for EPR proposed by Cartwright herself.

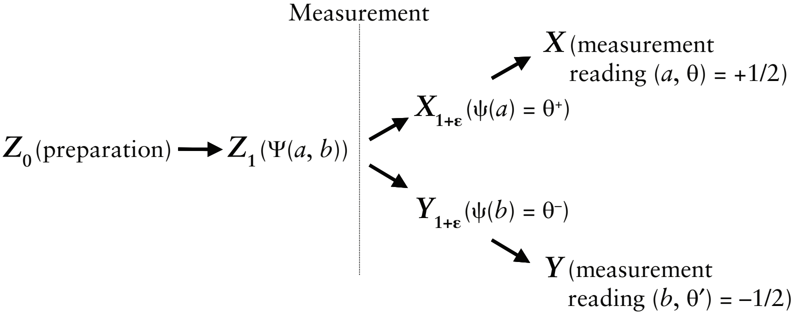
Measurement
X (measurement reading $(a, \theta) = +1/2$)
$X_{1+\varepsilon}$ $(\psi(a) = \theta^+)$
Z_0 (preparation)
Z_1 $(\Psi(a, b))$
$Y_{1+\varepsilon}$ $(\psi(b) = \theta^-)$
Y (measurement reading $(b, \theta') = -1/2$)

Figure 10.9

The first objection concerns the taking of the quantum state itself as an event. The event I am describing as $\Phi_{1/2}(a \,\&\, b)$ is "located" at two quite distinct spatial points; namely, the locations of the two particles.[18] But the standard notion of an event, a momentary event, not an action-event, holds that events occupy single spatiotemporal locations, or at least, contiguous spatiotemporal regions. On this notion of an event, a nonlocal quantum state is disqualified. Rather, we must treat the momentary states of each particle, considered independently, as distinct events, or else take the values of the quantum wave function in the vicinity of each particle as the events. Either way, we get two distinct events rather than the single event $\Phi_{1/2}(a \,\&\, b)$ in our model. In response to this objection, I would argue that it is physical theory itself which should tell us what to count as a single event, not some metaphysical preconceptions. In this view, the standard notion of an event illegitimately assumes that physical theory always involves localized causes. And this assumption is precisely what is rejected by quantum mechanics. The nonlocality of the quantum state is, I would say, an ontological *nonlocalizability*, which violates our preconceptions about the identity conditions for entities and events. I would thus argue that the true metaphysical lessons from quantum mechanics concern the ontological structure of the world, not the causal structure.[19]

The second objection I anticipate is a theoretical objection, not a metaphysical one. The objection here is that the model just presented is inconsistent with relativity theory. For the model presupposes that there is a single preferred time coordinate, *t*, such that we can talk meaningfully of the state of particle *a* changing at *the same time* as the state of particle *b*. But the theory of relativity says that simultaneity is frame-dependent—that distant events that occur at the same time in one reference frame can be greatly separated in time in other reference frames. In what reference frame is our event $\Phi_{1/2}(a \,\&\, b)$ an *instantaneous* state? Any answer, it seems, would establish an absolute rest frame, which is supposed to be disallowed by the theory of relativity.[20]

Now I think that this kind of objection is perhaps inevitable, no matter what account we give of EPR—the objection is not specific to the model I propose. For the model is stated within the framework of *nonrelativistic* quantum mechanics, and this theory is well known to be inconsistent with the theory of relativity! And it has proven extremely difficult to account for measurement within a Lorentz invariant quantum theory. So any story we tell about EPR based on current theory is going to run into problems with the theory of relativity at some point.[21] The problem, as I see it, comes from the quantum mechanical equations themselves, not from the attempt to give them a causal interpretation. We might say, then, that the causal model I provide does its job by pinpointing precisely where the problem with reconciling quantum mechanics and relativity lies. And this may be the best that we can do at the current stage of theorizing.[22]

But perhaps we need not be so pessimistic about the prospects for reconciliation with relativity. For there is one proposal for a relativistically acceptable account of quantum measurement which might be adaptable to our causal model. What I have in mind is Gordon Fleming's proposal to adopt a radical hyperplane dependence for physical states such as the position of a particle.[23] Fleming suggests that in a relativistic quantum theory, a particle may be spatially localized only relative to a particular hyperplane. So, while a particle may be localized at a particular spatiotemporal point relative to one hyperplane passing through that point, it may be radically nonlocalized relative to another hyperplane passing through the same point. Let us suppose that this is correct. Then we might also claim that causal structure may be equally hyperplane dependent—that the causal structure I have suggested may be correct, relative to one hyperplane, and a different causal structure be correct relative to a different hyperplane. If this idea were right, then my claim about **CM** would also have to be modified: The claim would then be that, *relative to each hyperplane*, there is a causal structure satisfying **CM**.

The final objection I wish to consider comes from Cartwright, by herself and in joint work with Hasok Chang (Cartwright 1989, 1990; Cartwright & Chang 1993). This objection returns us to the issue of propagation, this time focused specifically on the issue of propagation in quantum mechanics and EPR. The model I have offered gives a common cause structure and satisfies **CM** by the assumption of propagating causes. But this assumption leads us to invoke nonlocalized events. Cartwright, who rejects **CM**, nevertheless believes that a common cause model for EPR can be given. Cartwright's model avoids nonlocalized events, but it does so by rejecting the requirement of propagation. Indeed, Cartwright believes that we have good reason to deny propagation in the quantum realm.

Cartwright's argument against propagation is a variation on a familiar argument concerning the two-slit experiment. In the two-slit experiment, a beam of particles is emitted from a source and passed through a barrier containing two narrow slits. The slits may be opened and closed, and the effects of this on the beam are registered in the patterns observed on a detector screen. With only one slit open, the particles are registered on the screen mostly in a direct-line path from the source through the open slit. But with both slits open, an interference pattern is observed. This interference pattern shows regions where there is a very low probability of finding a particle, but which would have had a very high probability had only one slit been open. Such a conclusion is incompatible with the idea that each of the particles traverses a well-defined trajectory through either one slit or the other.[24]

Cartwright claims that this result is inconsistent with causal propagation. But the argument she gives assumes that a propagating cause is a localized one—causal influence may pass through both slits, but it must do so as two independent causal processes, one for each slit. If this is the case, then it can be shown that the probability of finding a particle in any given region cannot be lower with two slits open than it was with one slit open, which contradicts the observed frequencies. However, this argument fails if we allow nonlocalized events, such as a complete quantum state. For then we can hold that the causal influence propagates through *both* slits at once, as a single causal process. And in this way, we can replicate the observed frequencies with a causal model. Propagation of this sort *is* possible in the two-slit experiment, and Cartwright gives us no reason to reject the demand that causes propagate, at least in this nonlocalized fashion—which is precisely the manner of propagation to which I appeal in the causal model for EPR above.

But what of Cartwright's own common-cause model for EPR? Although we have seen that Cartwright's argument against propagation does not apply against the kind of propagation proposed here, this still does not show that propagation *must* be assumed in EPR. Cartwright's model denies propagation and so avoids the nonlocalized events that must be postulated to maintain propagation. Other things being equal, such a model should be preferred. Unfortunately, other things are not equal: Cartwright's position is committed to an even more mysterious kind of nonlocality.[25]

Cartwright's model is, in many ways, similar to the model just defended. The quantum state itself is taken as the common cause of the correlated effects, operating together with the measurement settings. However, in Cartwright's model it is the quantum state at the *source* that is taken to be the common cause, not the quantum state at the moment before measurement. Here is her model, stated in the form of causal equations (Cartwright 1989: 239):[26]

$$x_L(\theta) = \hat{a}_L(\theta).x_1 \ \mathrm{v}\ u_L(\theta)$$

$$x_R(\theta) = \hat{a}_R(\theta).x_1 \ \mathrm{v}\ u_R(\theta)$$

$$P(\hat{a}_L(\theta).\hat{a}_R(\theta')) = \tfrac{1}{2}\sin^2((\theta - \theta')/2)$$

As Cartwright herself points out, it is essential to her causal model that the common cause, x_1, operates to produce its effects *together*. There is a correlation between the action variables $\hat{a}_L(\theta)$ and $\hat{a}_R(\theta)$, representing the constraint imposed by conservation of angular momentum. For this to happen, Cartwright argues, the common cause 'must operate at the time the particles leave the source, say t_0, to produce an effect at some later time, $t_0 + \Delta t$, and with nothing to carry the causal influence between the two during the period $\Delta t'$ (Cartwright 1989: 127). The common cause must operate across a temporal gap—it does not propagate.

We have just seen that the crucial feature of Cartwright's common cause is that it operates across a temporal gap. How is it that this allows the common cause to reproduce the quantum mechanical correlations when locally acting causes cannot? The answer, it seems, must have something to do with the fact that the measurement settings are available at the time of operation of Cartwright's common cause but not at the time of the existence of that cause. But if this suspicion is correct, then Cartwright's model contains nonlocal action of a very disturbing nature. In determining its operation to produce a particular outcome for one particle, it seems that the common cause must be constrained both by the outcome it produces at the other wing and by the angle at which that outcome is produced, for this is the only way it seems that it could determine its action with the right probabilities. Is Cartwright really committed to this?

The problem is clearly hidden in the action variables. We are given that the common cause, x_1, acts to produce the outcome 'spin-up' on the left and right particles if the action variables, $\hat{a}_L(\theta)$ and $\hat{a}_R(\theta')$ respectively, take the value 1. We are further given a probability distribution for each pair of measurement settings, θ and θ′. But Cartwright is short on details as to how we are to interpret these elements. It is not clear whether there is supposed to be a value for $\hat{a}_L(\theta)$ for each value of θ, or just for the one value that the measurement device takes. But in either case, the theory runs into problems.

Suppose, for instance, the common cause, x_1, determines a value for the action variables for each possible value of θ. But then the common cause determines a joint probability distribution for every combination of settings, which is impossible; that is part of the content of the 'non-locality' proofs. So x_1 cannot determine values for the action variables for any possible value of θ.

Suppose, on the other hand, that x_1 only acts to give a value for one pair of values θ and θ′. If it does this for a random pair of angles, then it cannot duplicate the quantum mechanical correlations.[27] So the common cause must act to determine values for the action variables $\hat{a}_L(\theta)$ and $\hat{a}_R(\theta')$ for the exact pair of angles that are chosen for measurement. Thus, the common cause must 'know' what angles will be chosen for the measurement of the two particles before it can determine its joint action on the two particles.

We see, then, that the only way that Cartwright's causal model can give the right results is if we interpret the common cause as determining a joint operation on the spatially separated particles across a temporal gap, under the constraint of the device settings at the distant measurement sites. This involves nonlocality on a quite unacceptable level. Cartwright's common cause model cannot, therefore, be accepted.

But there is an even more serious problem with Cartwright's idea of non-propagating causes. This is that it does not seem to allow for interference with one of the particles in between the source and the measurement device. If the causal influence does not propagate with the particles, then it would seem that there would be no way, for example, to account for the effects of "rotating" one particle in midstream. Consider the experimental setup shown in Figure 10.10.[28] Here, the particle pair is produced in the singlet state, as before. But after the particles have been allowed to separate (and before the measurements), one of the particles is passed through a uniform magnetic field, suitably arranged to "rotate" the spin of the particle by 180°. The result of the "rotation" is that the singlet state is transformed, without collapse into the state $\Psi' = 1/\sqrt{2}\ (z_+ \otimes z_+ - z_- \otimes z_-)$. This transformation does not affect the spin of the other particle, but it does change the probabilities

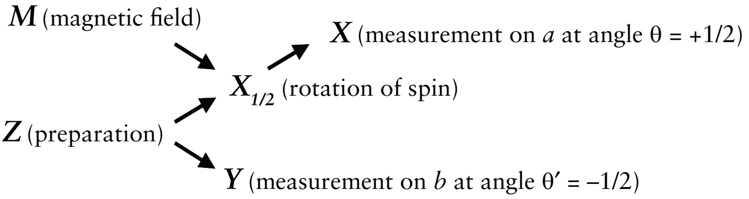
M (magnetic field)
X (measurement on a at angle θ = +1/2)
X1/2 (rotation of spin)
Z (preparation)
Y (measurement on b at angle θ′ = –1/2)

Figure 10.10

for the measurement outcomes—rather than a perfect anticorrelation when measured at the same angle, the spins of the two particles will now yield a perfect correlation.

How are we to account for this situation, on Cartwright's account? The quantum state at the particle source is the same state as that in the original EPR experiment, that is, one that gives probability zero to correlated results for spin measurement at the same angle. Yet this state is held to be the cause of the perfectly correlated measurement outcomes at the same angle in the modified setup. The cause must obviously be affected in its operations by the existence of the magnetic field operating on particle *a*. But how could this be, since on Cartwright's account the causal process never passes through the magnetic field?! The cause, when it comes to operate, must somehow have picked up information along the journey, even though it never took that journey. And this, I would argue, is incoherent. Furthermore, it does no good to suggest that the particle, which Cartwright does allow to propagate, may carry this information, for the particle has no definite spin until caused to do so.

Note, on the other hand, how easy it is to understand the modified EPR experiment on the model I have proposed. On this model, the quantum state propagates through the interval between source and measurement. It exists at each point on the trajectory of either particle. On passing through the magnetic field, that state is changed, such that the state just prior to measurement is very different to that which left the source. This altered quantum state then causes the measurement outcomes with a different set of probabilities than those for the unaltered state in the original setup.

CONCLUSION

In this chapter , I have tried to show three things. First, I have argued that the Causal Markov condition, in the modified form **CM***, can be satisfied, even when the Principle of the Common Cause is false. Second, I have argued that this is precisely the case in standard sorts of cases of indeterministic common causes, such as the example of radioactive decay I discussed. The upshot of this is that such standard counterexamples to the Principle of the Common Cause are not also counterexamples to the Causal Markov condition. The third conclusion of my chapter is that the same ideas may be used to give a common cause model satisfying the Causal Markov condition even for the EPR experiment, so long as we allow for nonlocalizable events as common causes. My final conclusion, then, is that the Causal Markov condition can, despite Cartwright's strong objections, still be a highly useful guide to causal structure at the microscopic level.

NOTES

1. Work for this chapter was supported by the Alexander von Humboldt Foundation, the Federal Ministry of Education and Research, and the Program for the Investment in the Future (ZIP) of the German Government. I would like to thank Nancy Cartwright, Stephan Hartmann, Christopher Hitchcock, Veiko Palge, Miklos Redei, Paul Thorn, Jim Woodward, and especially Paul Teller for insightful comments on versions of this chapter.
2. See, for example, Spirtes et al. (1993), Hausman & Woodward (1999), and Pearl (2000). It should be noted that, largely in response to criticisms by Cartwright and others, these supporters of the Causal Markov Condition generally express doubt about the applicability of the condition in quantum mechanical cases such as the EPR case, to be discussed below. Cartwright's claim is that the problem is far more pervasive than is allowed by these supporters, and I am inclined to agree that the phenomenon Cartwright focuses on is widespread.
3. This is the more refined principle, **CM′**, which Hausman and Woodward introduce in response to criticisms of a simpler principle, in which the variable set **V** is not quantified over. It is the more complex principle which Hausman &Woodward claim is defensible, so my discussion will focus on it.
4. The example I give is one adapted from an example used by Hausman & Woodward (1999), but the basic idea goes back to van Fraassen (1980).
5. It is clear that the episodic sense of "event" is inappropriate here, for in talking, say, of the *decay*, of an atom, we are illicitly building the causal relation to an effect into the specification of the cause.
6. Another response along these lines was suggested by Woodward at the workshop. The claim was that there is no causal fork in this case, for there are not two distinct causal *processes*. The idea is that it is impossible to act to manipulate the causal mechanism responsible for producing the alpha particle without also manipulating the causal mechanism responsible for the production of the radon ion. Thus, on the manipulability view of causation held by Woodward, there is only one causal mechanism at work and, so, only one causal process. But this response is as implausible as the others. For once the decay products have separated they form two independent and archetypical causal processes—even on a manipulability account of causal processes. So, the decay of a radium atom begins with one causal process and ends with two, in any account of a causal process. And if this isn't a causal fork, then what is! My conclusion concerning this response is that, if anything, it shows the untenability of Woodward's manipulability view of causation. The counterexample to **PCC** remains.
7. It is, in any case, unclear how this response is supposed to solve the problem, for **CM** makes a universal claim, concerning *all* correlations, and so is refuted by the failure of screening off, whether or not the case involves a common cause.

8. Interestingly, Cartwright's discussion is the only time I have seen this proposal in the literature, although I independently proposed such an idea in my dissertation (Martel 2000).
9. Note that I am assuming here a quasi-classical model of radioactive decay, in which the decay, while indeterministic, nevertheless still determinately occurs at a particular point in space and time. The correct, quantum mechanical picture would be more complicated, for quantum mechanics entails that the system is in a superposition of both decayed and undecayed states until the point of measurement (or collapse, or whatever). The correct model would thus have to account for the same kind of nonlocalized, nonseparable causes as we are about to discuss in the context of the EPR paradox. What this shows, I think, is that the sort of weirdness found in the EPR setup cannot be marginalized to a few odd cases—if quantum mechanics is right, then the weirdness is found everywhere, including atomic decay! At this point in the discussion, however, my aim is simply to show that, ignoring nonseparability and superpositions, the existence of a purely indeterministic common cause by itself presents no obstacle to **CM**.
10. Note the inserted restriction on the claims of **CM** to microscopic causal relations. This restriction is added as, in this context, I do not wish to maintain the claim of universality usually made by **CM**'s supporters.
11. In the continuous case, we may wish to strengthen the claim, to restore the first demand of **CM**, the 'Causal Markov Condition proper'. As it stands, **CM*** does not require that distal causes be screened off by proximal causes, for the set **Ancestors** (X) could include distal cause Y. To rule this out, we might restrict the condition to dense systems and add the requirement that $Y \notin$ **Ancestors** (X). This stronger condition requires that, between any two causally related events, there is a screening-off causal intermediary.
12. But what about the general case, of macroscopic causal forks such as in Cartwright's original example? Here too, I am somewhat doubtful about Cartwright's criticism. Consider Cartwright's banking example, and suppose that there are two ways of getting from her signing of the cheque to her nanny's getting the cash: One way involves a sorting machine at the post office; the other involves a child carrying an envelope by hand. Now, *given* the presence at the post office at a time, *t*, of a cheque written by Cartwright to the nanny, on a particular date, what can we say about the probability of the money being received at some later time? Well, given that the cheque is sitting in the post office, the option involving hand delivery by the child is *irrelevant*. Thus, the probability of the money being received, given that the cheque was written and that the cheque is sitting at the post office is equal to the probability of the money being received given simply that the cheque is at the post office. Similarly for the alternative path: if we are given that the child is carrying the cheque, then the post office possibility is ruled out, and the child's still having the cheque in its hand at time *t* screens off the earlier cheque signing from the receipt of the money. The presence of many alternative paths makes no difference to the screening-off relations.
13. In general, the correlation is given by the formula: $\mathrm{Pr}(X/Y\&Z) = \mathrm{Pr}(Y/X\&Z) \approx 1 - \sin^2((\theta - \theta')/2)$, where θ and θ' are the angles of measurement for the two spin states.
14. Given standard assumptions ruling out backwards causation, pre-established harmony, etc.
15. See Sober (1988)
16. Of course, this interpretation of quantum mechanics is rejected by most, for a variety of reasons. Here, I assume this interpretation for simplicity. But it should be possible to plug in any variant of the collapse view to the model

I propose. Many, perhaps, will wish to reject all collapse views. But I do not need to get into this debate here (see Note 21), for I am only defending the claim that it is *possible* to give a causal interpretation of EPR.

17. An important technical note should be made here. In treating the post-collapse causal structure as the same as that in the radioactive decay case, I am assuming that there is a last moment before measurement-induced collapse, but not a first moment after collapse. That is, in the model, the pre-collapse events form a set which is closed at the moment of collapse, whereas the set of post-collapse events is open. This is required in order that there not be an unscreened correlation between the initial measurement products. But why should we not model the situation the other way round, with the set of post-decay events closed and the set of pre-decay events open? Then **CM*** would fail. There are two responses to this objection. First, I think that the way I model the collapse is the more natural, as it parallels the structure of the closely related decay case. Second, even if this were not so, the model I suggest would not be ruled out, and my point is merely that it is *possible* to provide a causal model for EPR satisfying **CM***, and this is so even if there are other possible models which do not satisfy **CM***. My thanks to Paul Teller for stressing the importance of this point.
18. In fact, the problem, if it is a problem, is far worse. The wave function of a particle is smeared out over the whole of space, even if we have just measured its position (though the magnitude will be vanishingly small for any point which is away from the classical location of the article. This is the well-known "tails" problem of quantum mechanics.) So the state $\Phi_{1/2}(a \;\&\; b)$ in fact "exists" at *every* spatial location!
19. This is, of course, not the place to defend such grand ontological generalizations. Note, though, that similar ontological claims about the lesson of quantum mechanics have been made concerning our intuitive notions of individuality and objecthood.
20. I take it that Cartwright herself would not find this objection very persuasive, as she has expressed doubts about such relativistic constraints in this context. Indeed, Cartwright offers another causal model for EPR in which there is direct causation between space-like separated events.
21. Here I include all of those accounts which hide the problem by denying talk of causation in EPR, or invoking notions of "passion" at a distance, for there still remains the problem of *when* the passion takes place, or when the noncausal correlation is produced.
22. A point about the dialectical situation is in order here. As a defender of **CM**, *I* do not have to provide the definitive answer to these very difficult questions about the interpretation of quantum mechanics. For I merely have to show that there is no good argument showing the *impossibility* of a causal model for EPR satisfying **CM**. For recall that it is the opponent of **CM** who began by making the bold claim that *no* such causal model could be given. By showing that one candidate for an interpretation of quantum mechanics seems to allow such a causal model, I have already answered the objection. I do not have to go further, to show that this is the right interpretation or that it solves all of the problems. I should note also that I might have chosen a different interpretation to base my model on—for example, the Bohm theory seems to allow for a straightforward nonlocal causal model satisfying **CM**, and, indeed, **PCC**, for it is a deterministic theory.
23. See Fleming (1966). I am very grateful to Paul Teller for suggesting the possibility of such a response to me.
24. Except in a theory like Bohm's, which has definite particle trajectories, but which also has nonlocalized quantum waves passing through both slits. Here

the particles have trajectories, but the causal influence does not. But Bohm's view has a kind of propagation, nevertheless: the kind of propagation I defend for my view.

[25] Cartwright's model has further problems which I shall not discuss; see Cachro & Placek (2002).

[26] I have eliminated two equations and implemented some changes (suggested by Cartwright) to make the common cause, $x1$, the quantum state itself.

[27] This is seen most easily if the randomly selected angles are chosen to be different but the measuring devices are set to the same angle. Then there will be a probability for the two measurements yielding the same result, which QM prohibits.

[28] Note that this diagram should be read as schematic only—neither Cartwright nor I would hold that the diagram correctly represents the causal structure!

REFERENCES

Cachro, J., and T. Placek. (2002) 'On Cartwright's models for EPR', Studies in History and Philosophy of Modern Physics, 33B: 413–433.

Cartwright, N. (1989) *Nature's Capacities and their Measurement*, Oxford: Oxford University Press.

———. (1990) 'Quantum causes: The lesson of the Bell inequalities', *Proceedings of the 13th International Wittgenstein Symposium*, Vienna: Holerlin-Pichler-Tempsky.

———. (1999) *The Dappled World*, Cambridge: Cambridge University Press.

Cartwright, N., and H. Chang. (1993) 'Causality and realism in the EPR experiment', *Erkenntnis*, 38: 169–190.

Fleming, G. (1966), "A Manifestly Covariant Description of Arbitrary Dynamical Variables in Relativistic Quantum Mechanics", Journal of Mathematical Physics 7:1959–1981.

Hausman, D., and J. Woodward. (1999) 'Independence, invariance, and the causal Markov condition', *British Journal for the Philosophy of Science*, 50: 521–583.

Martel (2000) Probabilistic Empiricism: In Defence of a Reichenbachian Theory of Causation and the Direction of Time, unpublished dissertation.

Pearl, J. (2000) *Causality: Models, Reasoning, and Inference*, Cambridge: Cambridge University Press.

Sober, E. (1988) 'The principle of the common cause', in Fetzer J. (ed.) *Probability and Causality*, Dordrecht: Reidel.

Spirtes, P., Glymour, C., and Scheines, R. (1993) *Causation, Prediction, and Search*, New York: Springer-Verlag.

van Fraassen, B. (1980) *The Scientific Image*, Oxford: Clarendon Press.

Reply to Iain Martel

Iain Martel offers a very nice model for EPR, an especially natural one for quantum realists because it takes the quantum state entirely seriously. The entangled quantum state of the two particles exists at the instant of measurement in both the left-hand and right-hand wings of the experiment. It, in conjunction with the operating apparatus, produces new nonentangled states for the two particles in each wing; it produces one state "with its left hand" and the other "with its right hand". All that is strange is that a spatially extended cause acts "all at once" in two different places.

My model is strange in a different way, which Martel points out. He describes the first simple version of it, which represses the variables describing the measurement apparatuses. The "enlarged model" includes these, making explicit the fact that the entangled state at the source is a partial cause that acts in conjunction with the state of the measurement apparatus in a given wing to produce the result in that wing (Cartwright 1989: 241). Martel's model uses the evolved quantum state of the particles to carry the influence of their state at the source through time and space to the time and place of measurement. Why have many people not leaped to such a model? Because, I reckoned, the effect depends on details about the entire entangled state at the time of measurement and hence on features outside its backward light cone. So my model dispensed with this.

The enlarged model simply supposes that the two causes—the entangled state at the source and the apparatus state in the wing where the result occurs—together produce the effect according to the formula recorded in equation (1′). The formula is of course valid only in the absence of further causes. If there is a "rotation" after the first partial cause and before the second, a new formula is required that describes the effect produced conjointly by the three partial causes.

Locality in this model means that nothing outside the backward light cone is relevant to the production of the effect, in particular neither the setting of the distant apparatus nor the distant outcome.[1] Without the joint common cause in the proper past of both outcomes, the outcomes will not be correlated. Yet, as Martel points out, we cannot suppose that the joint cause produces some trace associated with a specific result for each wing;

rather, the action to produce the effect can occur only once all relevant partial causes have occurred. This means that causes that no longer exist contribute to what is produced later, across a time gap, and with no trace existing in between.

As Martel says, 'the common cause must be constrained both by the outcome it produces at the other wing and by the angle at which that outcome is produced. Is Cartwright really committed to this?'[2] The answer is 'yes'. Perhaps I am not upset about action like this, even cooperative action, because I have learned from Bertrand Russell that this is characteristic of all causal models (Russell 1913). Whether the stages of the causal process are discrete or continuous, the cause always passes out of existence before the effect occurs.

Wesley Salmon's "at-at" theory of the causal process was designed to circumvent this difficulty (Salmon 1984). The cause (an interaction) will be gone before the effect (another interaction) occurs, but a causal influence appears at every instant in between. What happened in the end? A causal influence appears at a certain time and place and makes the effect appear simultaneously. But how? I believe that whether we are regularity theorists or power theorists, nature must have a formula: "x at t produces/has the power to produce/occurs with y at t". But then why cannot the formula be "x at t – δ produces y at t" or "x at t – δ and z at t – αδ produce y at t"? Maybe we think that the effect draws something out of the cause; for instance, the cause hands over some energy—or in the case of EPR, a bit of spin—to the effect. Then it seems the transferred quantity must exist at times in between.

Perhaps I am too Humean here. The idea that there is a bucketful of something and the cause doles out bits of it to the effect does not seem to me a plausible story. It does not seem plausible for how conservation works, let alone causation, either inside or outside of quantum mechanics. Even given the bucket story, nature must use a formula to "decide" what is doled out and when. But if the formula holds, there is no need for the bucket to begin with. And I don't see any important differences between a formula that connects events at different times from one that connects events at one and the same time.

Before turning to EPR, Martel also reminds us of an important fact about causal Markov-type conditions that is generally true. If between a cause of type *C* and its effect of type *X* there is a temporally continuous path such that at any instant in between there is a cause of *X*, suppose it to be of type P_x, that screens off *X* from *C* ($P(X/P_x\&C) = P(C/P_x)$), then there will be a cause of *X*—to wit, P_x—that screens off *X* from joint effects of *C* (i.e. his **CM***). I myself am wary of assuming that there always are such processes when "is a cause of" is intended to mean, roughly, "is in the antecedent of a causal law for". I am wary of this even if we allow that any two cause–effect tokens may be connected by temporally continuous causal processes.

Martel hypothesizes that if there are processes for the tokens, then **CM*** should hold. My hypothesis is that often the intermediate steps in the token process will not fall under relevant types that figure in the right kind of laws; that is, under a type P_x such that there is a causal law (or a series of 'intermediate' laws) connecting C-type events and P_x-type events and also a causal law (or series of causal laws) connecting P_x-type and X-type events. In the banking example, for instance, I doubt that there is a causal law connecting writing checks with posting them. But that of course has to do with the nature and sources of causal laws and not with the main topic of EPR.

NOTES

1. This is guaranteed by conditions (1) and (2) in Appendix I, *Nature's Capacities and Their Measurement*.
2. Martel also worries about how the 'action variables' get determined. But these do not represent what I would count as real events, and I think Martel would agree given his discussion of Hausman and Woodward. They merely encode information about probabilities. In particular, neither their values nor the probability distributions over them are fixed by the state of the entangled particles at the source.

REFERENCES

Cartwright, N. (1989) *Nature's Capacities and Their Measurement*, Oxford: Clarendon.

Russell, B. (1913) 'On the Notion of Cause', *Proceedings of the Aristotelian Society*, 13: 1–26.

Salmon, W. (1984) *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.

11 Social Capacities

Julian Reiss

INTRODUCTION

Nancy Cartwright is commonly held to advocate the capacities concept as a central tool for the philosophical analysis of practice in natural and social science alike. But it would be wrong to ascribe to her the view that social phenomena are governed by causal factors with stable capacities (or *social capacities* in short). Her point is rather that the methods many social scientists use presuppose, in order to be successful, the existence of capacities. But since in her view the record of success in employing these methods is at best mixed, to be consistent she cannot believe that the social world is actually for the most part governed by capacities.

Already in *Nature's Capacities and Their Measurement*, which originally introduced the capacity concept and, in fact, used econometric practice more than once to prove a point, Cartwright employs John Maynard Keynes in order to express scepticism about the reality of social capacities. According to Keynes, the universe consists of bodies 'such that each of them exercised its own separate, independent and invariable effect' (Keynes 1957/1921: 249, quoted from Cartwright 1989: 156). However:

> We do not have an invariable relation between particular bodies, but nevertheless each has on the others its own separate and invariable effect, which does not change with changing circumstances, although, of course, the total effect may be changed to almost any extent if all the other accompanying causes are different. (Keynes 1957/1921: 249, quoted from Cartwright 1989: 156)

The contribution a factor makes to a situation is thus dependent on the arrangement of all other factors. In other words, the Keynesian world is "holistic". As a consequence, John Stuart Mill's methodology of analysis and synthesis cannot be applied.

In a different context, Cartwright uses ideas of members of the German Historical School to shed doubt on the reality of capacities in social phenomena. In an encyclopaedia entry on capacities, she writes:

> Just as the analytic method of Newtonian physics was challenged by Goethe and others in his particular German tradition, the analytic method of classical economics laid out so clearly in Mill was rejected by the Historical School dominant in German political economy at the end of the nineteenth and beginning of the twentieth century. [. . .] Gustav Schmoller, in particular, is famous in methodology for his insistence against Carl Menger that history and political economy could not employ exact universal laws as physics does. For Schmoller, as for Mill, economic phenomena are brought about on each occasion by a myriad of interacting causes. But for Schmoller, the role each cause plays depends on the total context in which it is set. So although separate causal factors can be identified and reidentified from one context to another, the separate causes do not have stable capacities. (Cartwright 1998: 45, emphasis added)

The aim of this chapter is to investigate how well-founded Cartwright's scepticism is. In order to do so, I first review some essential aspects of the capacities concept and its application to social science in the next section, 'Capacities'. In the following section, 'Are there social capacities?' I play devil's advocate and present three case studies that give good reason to believe that our prospects for finding social capacities are very grim indeed. In the penultimate section 'How well-founded is scepticism about social capacities?', I then discuss whether Cartwright is right. For that I distinguish two forms of scepticism, atheism and agnosticism. I argue that there is good reason to be agnostic but little for being a fully fledged atheist. I conclude by pointing out a number of methodological modifications social science could adopt to make it more likely to find social capacities.

CAPACITIES

Cartwright is a causalist in the sense of believing in the reality of causal relations or the reality of properties with genuine causal efficacy. In her conceptual framework, causal concepts are primitive: They cannot be further analysed in terms of laws of nature, relations of counterfactual dependence, or the like. Among the causal concepts, the capacity concept adds to the idea of causality the ideas of potentiality and stability. Saying that some X has the capacity to Ψ tells us something about what X does potentially: When X operates unimpeded, it produces Ψ. However, even when this process is interfered with, X will tend to or try to do Ψ. In other words, if there are causal factors present that impede on X's action to do Ψ, X will still contribute to the overall result. Secondly, the ability of X to Ψ must be stable across some range of circumstances if it is to count as a capacity.

To give a hypothetical example from social science, let us assume that economists have established that the growth of money in an economy has the capacity to raise the general level of prices. According to the conception that Cartwright defends, this means (a) that the growth of money is not only correlated with the level of prices but it actively produces its increase, (b) that this says nothing about the actual behaviour of the price level, since there can always be factors such as technology shocks or international trade which can interfere with the operation of the money stock; however, even in such a situation, money will *contribute* to the actual behaviour of the price level, and (c) the ability of the growth of money to increase the price level is stable across some range of situations, for example across different capitalist economies, under different monetary regimes, etc.; however, it is possible that there are situations (in systems with radically different economic constitutions, say) where money does not have this capacity.

What is the relevance of capacities for social science? The answer is conditional: If there are social phenomena that are governed by factors with capacities and they are epistemically accessible, then the social scientist and/or engineer is helped in realising his epistemic and pragmatic aims. The aims of social science, I take it, can be described by the tripartite expression explanation-prediction-control (see, for example, Menger 1963). Knowledge about capacities helps in explaining social phenomena. This is evident from their causal nature and the wide acceptance of causal models of scientific explanation. Capacities can also help in predicting phenomena. It is an analytic truth that capacities allow to make true conditional predictions about phenomena of the form: 'If nothing interferes, doing X will result in Ψ.' This is an exact prediction but nonvacuously true only if the antecedent is fulfilled. On the other hand, if disturbing factors do occur, the prediction will be inexact and about the contribution of X to the overall result.

Knowledge about capacities may also help in planning and control. The matter is, however, slightly more complicated here. When Mill originally introduced the tendencies concept, after which Cartwright modelled her own concept, he did so in order to save the truth or universality of natural laws in the light of apparent disconfirmations due to intervening factors.[1] Suppose we have a particle which is subject to two forces, one that pulls in x-direction and another that pulls in a 135° angle to it. When each force operates on its own, the law that describes its behaviour is true. But when both forces operate jointly, the law, understood as a function from forces to actual *results* (in this case, motions), is literally false. This is because part of the motion in x-direction that would have occurred had the first force operated on its own is upset by the second force and vice versa. The motions do not occur in the way the (individual) laws predict. But if we understand the law as an ascription of a tendency or capacity, its truth is salvaged. This is because, although the motion that actually occurs is not the one predicted by either law, each law is true of the *tendency* to produce its characteristic result, and each force contributes to the overall result.[2]

Therefore, the kind of stability the concept requires is a stability "under interferences". But this does not imply that the X-Ψ relation must be stable under *all* interferences and, in particular, under interferences which destroy the causal structure on account of which the capacity arises. An economic example will illustrate this claim. Let us assume that the Phillips curve has a true causal reading in the capacities sense. That is to say, let us assume that the unemployment gap (the difference between the actual and the "natural" unemployment rate) has the capacity to decrease inflation. This means (a) that the unemployment gap causes inflation, (b) that any precise formulation about the exact relation between the two quantities is true only potentially, that is, in the absence of interferences, and (c) that when other factors capable of affecting the inflation rate are co-present, the unemployment gap still contributes to the overall result. But it does not mean that the capacity must still be there if the economy has been intervened on, for example by changing the wage-setting process. Imagine, for example, an intervention that changes the labour market from a purely centralised and unionised system to a localised and free system. This will surely have effects on the rate with which the unemployment gap changes the inflation rate, i.e. on the strength of the capacity, but possibly even on its existence (i.e. it may change its strength to zero).

The ability to plan and control presupposes knowledge not only of capacities but also of a second kind of stability, which in econometrics is called autonomy. An autonomous relation is essentially one which remains stable under (some range of) interventions. Stability across some range of circumstances is part of the concept of capacity, but the difference between the circumstances does not have to involve interventions. All interventions change circumstances though, and hence, all autonomous relations involve a capacity but not necessarily the other way around. However, thus construed, knowledge of capacities again helps in realising (this time, pragmatic) scientific aims; it is just that the economic planner needs to know something more too.

ARE THERE SOCIAL CAPACITIES?

If there are stable capacities in a domain of interest, research is aided in a number of ways. Most importantly, claims that have been established with respect to a certain test situation *X* are exportable outside *X*. For example, if we judge on the basis of the Stanford/NASA gyro experiment (see Cartwright 1989: Ch. 2) that coupling has the *capacity* to affect precession in the way the experiment tells us, then we assume that coupling affects precession also outside the experimental situation. True, outside the test situation coupling may *result* in no precession at all. But that means that there is an inhibiting cause which prevents the capacity from being exercised.

Although Cartwright has voiced her scepticism in a number of papers, talks, and in personal communication, she never really defends it with respect to modern social science (with one exception that I will discuss below). The purpose of this section is to examine a number of significant methods of causal inference in social science that indeed give reason to believe that the causal claims established by them are not claims about capacities. My argument then is straightforward. Were there (knowable) social capacities, we would (probably) be able to find out about them with our best methods. However, analysis of our best methods in social science shows that these methods are not capable of finding capacities. Therefore, (probably) there are no capacities.

Exhibit I: The Vanity of Rigour

In the "Vanity" paper (Cartwright 1999), Cartwright argues that the thought experiments or "toy models" we find every so often in theoretical economics do not provide evidence for capacities. This is due to the fact that these models employ many "non-Galilean" idealizations, which implies that one cannot attribute the effect to the cause of interest as its "characteristic effect".

The concepts of "Galilean" idealization and capacity are closely linked. An idealization is Galilean if it helps in learning about operation of a causal factor free from disturbances. Galileo's own thought experiments on falling bodies are good examples. In one thought experiment, Galileo asks us to imagine two bodies falling from a tower without air resistance, a heavier cannon ball and a lighter musket ball. According to the Aristotelian tradition, the heavier ball falls at a faster rate and hits the ground first. However, if we now suppose that we join the two balls with a string, the Aristotelian theory falls into a contradiction. This is because we can derive that the amalgam falls both faster as well as slower. On the one hand, it falls slower because the lighter ball pulls the heavier one upwards and thus slows down the ensemble. On the other hand, the two balls together are heavier than the heavy ball alone and thus should fall faster. Hence, Galileo argues, in a vacuum all bodies fall at the same rate.

The assumption of no air resistance, then, is a Galilean idealization as it helps us learning what the Earth's pull does to falling bodies in the absence of disturbing factors. In other words, it helps us in learning about the capacity of the Earth to attract heavy bodies. Ernan McMullin, in his paper 'Galileian Idealization' (McMullin 1985), distinguished a number of kinds of idealization, but in the present context the kind he calls 'causal idealization' is most relevant. McMullin writes,

> And it is this sort of idealization that is most distinctively "Galilean" in origin. His insight was that complex causal situations can only be understood by first taking the causal lines separately and then combining them. [. . .]
>
> The move from the complexity of Nature to the specially contrived order of the experiment is a form of idealization. The diversity of causes found in Nature is reduced and made manageable. The influence of impediments, i.e. causal factors which affect the process under study in

> ways not at present of interest, is eliminated or lessened sufficiently that it may be ignored. Or the effect of the impediment is calculated by a specially designed experiment and then allowed for in order to determine what the "pure case" would look like. [. . .]
>
> Galileo is convinced that he has discovered the motion that "nature employs for descending heavy things". [. . .] It is "natural" in the sense that it defines what the body would do on its own, apart from the effects of causes (like the resistance of air) external to it. These latter are to be treated as "impediments", as barriers to an understanding of what the "natural" tendency of body is.
>
> (McMullin 1985: 265)

It is a commonplace that the models characteristic of theoretical economics are highly idealised. Cartwright points out that many of the idealizations employed are not of the Galilean kind. Consider Akerlof's famous lemons model (Akerlof 1970).[3] Akerlof's aim was to explain the phenomenon of a large price differential between new cars and cars that have just left the showroom, or, in more general terms, that markets where quality matters often experience lower than expected prices and exchanged quantities. The second-hand car market is an instance of this more general phenomenon.

Akerlof explained the phenomenon by pointing out that in such markets there is an asymmetry in the information distribution: sellers know more than buyers. After they learn about the quality of their cars, owners of lemons (bad-quality cars) will want to sell their cars and exchange them for new ones, whereas owners of good cars will keep their cars. Because the quality of the car is not observable to buyers, cars are priced at some average rate, which further increases the incentives of owners of bad cars to sell their bad cars and of owners of good cars to keep their good ones. Hence, quality, prices, and exchanged quantities drop.

To lend credibility to his story, Akerlof provides a mathematical derivation of the result in addition to a more intuitive thought experiment. As is very common in investigations of this kind, Akerlof makes a large number of assumptions in order to derive the result in the mathematical model. Cartwright points out that making these assumption is in fact a methodological prerequisite. This is, she claims, because the basic principles of economics (the equivalent to "laws" in physics) are both few in number and meagre.[4] As a consequence, there is not a lot of deductive power built into them. But this in turn means that many additional structural assumptions must be made if results are to be deduced mathematically.

Among the assumptions Akerlof makes is that there are two types of traders with distinct utility functions, and both types are von Neumann–Morgenstern maximisers of expected utility, that the cars' quality is distributed uniformly between zero and two, that goods are infinitely divisible, and that the price of "other goods" is one. Few of these are Galilean in nature. That is, few of the assumptions help us learn what asymmetric information does on its own. If one attempts to trace back responsibility for a result, one finds that not only the factor of interest is to blame, but so are all of the assumptions made—otherwise no result would have been obtained. But this in turn means that we have not isolated a tendency.

So what did Akerlof establish? In my view, he measured the causal effect of asymmetric information on quality and volume in the system he envisaged. To see this, note that his method of proof resembles very closely Mill's "method of difference". Remember that the method of difference infers the causal effect of a factor *F* by comparing two situations which are identical except *F* is present in one and absent in the other (and *F*'s effect if there is any). The difference *F* makes to the situation, then, is its causal effect. Akerlof does exactly that. He models a situation with symmetric and a situation with asymmetric information, and the difference in the market result is then attributed to the difference in the information distribution. But because since the result crucially depends on the assumptions made, we can judge it to be present only in systems of which Akerlof's assumptions are true.

The point is this. If we aim to establish that a factor *X* has the capacity to Ψ, we had better make our conclusions as independent of the test conditions as possible. This is the lesson one can draw from McMullin's treatment of Galilean idealizations. Due to the particular manner in which results are determined in models such as the lemons model, however, conclusions are highly dependent on test conditions—in this case the model assumptions. Therefore, in themselves, they cannot establish a capacity claim.

One might object that the lemons model is a nonstarter as a tool for establishing capacities anyway. The model is a piece of theory after all, whereas a capacity claim is a claim about a particular kind of causal relations in the world. The reason I include theoretical models in my brief survey of methods in social science is that they are frequently taken—in themselves—to provide evidence for capacities. For instance, an argument why third-world labour markets fail might go as follows. In third-world countries labour markets often fail. In these markets quality matters. In markets where quality matters, asymmetric information can lead to market failure (we have established that with Akerlof's model). In third-world labour markets, the quality of labour cannot be observed by employers, i.e. there is an asymmetry in the information distribution. Hence, in third-world labour markets asymmetric information causes market failure.

This argument is obviously fallacious. One cannot argue from the fact that in a specific situation a causal factor is responsible for a result to the conclusion that it does so too in the envisaged situation. The least we need to do is to rule out all alternative explanations for the result. But worse, the result has been established only for a very unrealistic situation (one where there are only two types of agents, both von Neumann–Morgenstern maximisers of expected utility, distinguished only by their respective utility functions, cars have only one property, viz., "quality", which is uniformly distributed, etc.). Thus one would first need to establish that the conclusion holds in a real experimental situation too before one exports it to other situations. Nonetheless, arguments of this kind can be found, so I wanted to include theoretical models here. Exhibit II examines a case where the conclusion is established experimentally.

Exhibit II: Natural Experiments in Economics

There is a movement in contemporary econometrics which has been labelled 'natural experiments movement' (see Heckman 1999). Its basic strategy can be summarised as follows:

> *Natural Experiments*. To measure the causal effect of *C* on *E*, find a set of economic units on which one can measure *E* such that one can partition them naturally, i.e. without intervention, into treatment group (where *C* is present) and control group (where *C* is absent) in a way that resembles a controlled experiment. That is, the distribution of factors, which are causally relevant to *E*, is identical in treatment and control group and the assignment of a unit to a group is independent of any factor that may be causally relevant to *E*. Then measure the causal effect by taking the difference between the averages of the *E*-values in treatment and control groups.

An example illustrates this. Economic theory predicts that a rise in the minimum wage leads employers to cut jobs. David Card and Alan Krueger challenged this (universal) prediction with an analysis of a natural experiment that occurred in New Jersey in 1992 (Card & Krueger 1994, 1995). In that year, New Jersey's minimum wage rose from $4.25 to $5.05 per hour. In order to measure the causal effect of the minimum wage rise (*C*) on the change in employment (*E*), Card and Krueger surveyed 410 fast-food restaurants in New Jersey and eastern Pennsylvania before and after the rise. The economic units of interest (the fast-food restaurants fall naturally into two groups, the ones in New Jersey, which form the treatment group, and the ones in eastern Pennsylvania, which form the control group). Several items of background knowledge allow the authors to judge that the natural setup resembles a controlled experiment sufficiently. They argue, for example, that 'New Jersey is a relatively small state with an economy that is closely linked to nearby states' (Card & Krueger 1994: 773), and therefore, one has no reason to believe that the distribution of factors that could be relevant to employment differs between New Jersey and eastern Pennsylvania. This choice of control group is further tested by means of a second control group, viz. restaurants in New Jersey, which initially paid at least $5.00 per hour wage, and thus should not be affected by the rise. In particular, they observe that 'seasonal patterns of employment are similar in New

Jersey and eastern Pennsylvania, as well as across high- and low-wage stores within New Jersey' (Card & Krueger 1994: 773.), such that the "natural development" of employment, which could confound their result, should be controlled for. There is furthermore no reason to believe that the change in the level of employment is dependent on the assignment to groups, that is, whether the restaurant is in New Jersey or Pennsylvania.

Card and Krueger present their results as follows:

> . . . we find no evidence that the rise in New Jersey's minimum wage reduced employment at fast-food restaurants in the state. Regardless of whether we compare stores in New Jersey that were affected by the $5.05 minimum to stores in eastern Pennsylvania (where the minimum wage was constant at $4.25 per hour) or to stores in New Jersey that were initially paying $5.00 or more (and were largely unaffected by the new law), we find that the increase in the minimum wage increased employment.
>
> (Card & Krueger 1994: 792, emphasis added)

I do not want to comment on whether Card and Krueger are successful in their analysis of the natural experiment (but for a discussion, see e.g., Neumark and Wascher 2000). They surely *try* to replicate the structure of a controlled experiment. The point to draw attention to is rather that *if* their results are valid, they cannot be understood as an ascription of capacity. What I believe they can claim is that *ceteris paribus*, raising the minimum wage increases employment. But since one way of understanding statements of *ceteris paribus* law is as an ascription of capacity (Cartwright 2002), the difference I point out requires elaboration.

To ascribe a capacity to a causal factor means that one believes that certain inductive inferences are licensed. Surely the usual inference to all situations that are relevantly similar to the test situation is made. But, importantly, we know more: Even when the conditions of the test situation are not fulfilled, if the causal factor is present, it will still "try" or "tend" to produce the result. If it does not succeed, then there must be a very good reason for it, viz. a countervailing capacity. For example, saying that the Earth has the capacity to attract heavy bodies means that it will still try to do so even when gravity does not operate on its own. Now, if the Earth does not succeed in attracting a given heavy body, we ascribe this failure to the presence of another capacity, for example a strong magnetic field above the body that pulls it upwards.

Inferences that are licensed by a *ceteris paribus* law in the sense used here are narrower in scope. They allow only the usual induction to cases that more closely resemble the test situation. My point about the Card and Krueger paper is that they can make (at best) only the latter kind of inferences, not the former, broader kind. They did not find a general truth about minimum wages (nor do they believe they did). Rather, what they found (again, of course, if their results are valid) is a law that under certain conditions raising the minimum wage will increase employment. What these conditions are is difficult to say. Crucially, however, the failure of raising the minimum wage to produce more employment in a very different situation (e.g., when the minimum wage is already very high, when the rise is very large compared to its level, when economic conditions are radically different, etc.) will not induce us to seek for a countervailing tendency. Rather, we will attribute the failure to a relevant difference between the two situations we have compared.[5]

One may object that this relevant difference is exactly a countervailing tendency and that the difference is only terminological. However, I think that reading the claim as a capacity claim would be highly unnatural. Let us suppose that there is a second natural experiment involving two different states with characteristics very similar the Card and Krueger case. One difference is that the minimum wage in these states is initially much higher, say, $10.00. Let us also suppose that raising the wage to $12.00 results in a *de*crease of employment. Now it seems to me that it would be absurd to say that there is a capacity of raising the minimum wage (of the first $5.00?) to increase employment, which is offset by a countervailing, stronger capacity (of the second $5.00?) to decrease employment, such that the overall result is negative. Rather, one would say that the situation differs in crucial respects and that the law we found in the first case is not at work in the second case.

This discussion, I believe, points towards a more general feature about thinking in capacities and thinking in *ceteris paribus* laws. Thinking in capacities presupposes a method of analysis and synthesis. Situations are broken down to tractable parcels, the behaviour of these parcels is analysed severally, and finally, the bits are synthesised to let us know about the initial situation. Among other things, the method of analysis and synthesis presupposes that it makes sense to investigate what the parcels do on their own. Many cases Cartwright examines have this property. It makes perfect sense to talk about bodies subject to no other force than gravity. Even certain physically impossible scenarios, e.g., the behaviour of bodies that have charge but no mass, are relatively easily conceptualised.[6]

In the social sciences, by contrast, the method of analysis and synthesis (in this sense) seems less applicable. No factor produces anything on its own. It does not even make sense to ask, for instance, what a minimum wage does in the absence of everything else. We need a thick network of causal conditions to produce any result. Furthermore, the result that is actually produced very often depends crucially on the conditions that are present when the factor operates. It seems then, that in such situations the language of *ceteris paribus* laws is more applicable than the language of stable capacities.

Exhibit III: Singular Causal Analysis

The example in the preceding discussion was one where the causal effect of a single event (the raising of the minimum wage in New Jersey on April 1,

1992) on a property measured on a population (employment in fast-food restaurants in New Jersey) was measured. At the microlevel, i.e. the individual restaurants, employment is caused by a variety of factors, many of which probably escape subjection to causal law. The effect of increasing the minimum wage was extracted using, among other things, the assumption that the distribution of all other causes of employment was identical between the "treatment group" in New Jersey and the "control group" in eastern Pennsylvania. In many cases in social science, however, we will not be so lucky as to have such favourable circumstances. In particular, in many cases both relata of the causal statement of interest will be singular events (when the questions are, e.g., whether a certain decision or a certain battle stopped a war or whether the decision of the Fed to increase interest rates on a particular date triggered the financial crisis in Asia). So how do we establish singular causal claims? Cartwright tells us that in many cases in the physical sciences such claims can be established by bootstrapping (see for example Cartwright 1989, 2000). In general, the bootstrapping methodology allows us to infer a hypothesis deductively from the data and background knowledge (Glymour 1980). In the Stanford/NASA gyro experiment example I have alluded to above, the relevant hypothesis is whether space-time curvature causes relativistic precession of amount x, which is predicted by general relativity theory. Our background knowledge consists of a disjunction of hypotheses about the various sources of precession different from curvature coupled with the knowledge (or assumption) that all such sources have been controlled for successively. The data consists in the measurement result that precession is indeed x. Thus, we can derive the hypothesis from background knowledge and data deductively. More importantly, our background knowledge assures us that we have established a singular causal claim. Since nothing else *in this particular case* could have caused precession, space-time curvature must have been responsible for it.

In social science, unfortunately, the requirement about background knowledge seems unduly restrictive. This is for at least three sets of reasons. First, it seems impossible to find a disjunction of factors that could cause the phenomenon of interest that which exhausts all possibilities. Not only does experience tell us that such a list would be very long, it is also open ended. Nobody can predict the rise of the dot.com industry, but once that phenomenon is extant, it will serve in many causal explanations of other phenomena. Second, in social phenomena there is less room for manipulation. Very often the aim is the explanation of a historical event, where manipulation is impossible to begin with. But even disregarding that problem, experimental control is often out of reach for ethical, practical and economic reasons. The third difficulty is associated with the second one. In physics, even if we cannot literally control for a confounding factor, we can very often either calculate precisely the contribution of that factor or at least run simulations and thereby calculate upper limits. Most laws in social science, by contrast, are of a highly qualitative nature. Therefore, if the question is, say, whether a certain event has triggered a financial crisis, and we know that another factor that can contribute to financial crises was present, it is hard to tell whether the presence of that latter factor by itself would have been "enough" to trigger the crisis or whether the particular event we focus on was necessary in the circumstances.

These three sets of difficulties notwithstanding, a number of authors have attempted to tackle the issue of singular causation. To my knowledge, however, only Max Weber has developed a systematic account of causal inference in a single case. In my view, Max Weber is the only methodologist who has developed an account of singular causal inference tailored to the epistemic situation social scientists are often interested in. So let us examine whether his ideas can be exploited.

Two concepts are central to Weber's ideas about singular causal analysis: that of "objective probability" and that of "adequate causation". Objective probability is a term Weber borrowed from the German physiologist and statistician Johannes von Kries, who himself developed a tradition in the German legal philosophy (Ringer 1997: Ch. 3). Broadly speaking, an event is objectively probable[7] if the range of possibly relevant conditions under which it will occur is greater than the sum of further conditions under which it will not occur.

Weber first notes that, as I have discussed above, social phenomena are usually brought about by a vast number of factors[8], all of which are necessary in the circumstances for the result. In particular, against Mill Weber points out:

> Rather it is to be emphasized once and for all that a concrete result cannot be viewed as the product of a struggle of certain causes favouring it and other causes opposing it. The situation must be seen as follows: the totality of *all* the conditions back to which the causal chain from the "effect" leads had to "act jointly" in a certain way and in no other for the concrete effect to be realized. In other words, the appearance of the result is, for every causally working empirical science, determined not just from a certain moment but "from eternity". (Weber 1949: 187)

The first step in his causal analysis is that a number of factors of interest are isolated from the network of interacting causal factors.[9] When, for example, Eduard Meyer asks whether the battle of Marathon was significant for the subsequent development of Western civilisation, we notice that a myriad of factors is responsible for the development of our civilisation as it actually occurred, but we single out a particular event of interest *C*, the battle of Marathon, and ask whether it was significant for the phenomenon of interest *E*, viz., the development of Western civilisation.

The essential procedure to answer a causal question of the form 'Did event *C* cause event *E*?' is to ask oneself if *E* would be expected had *C* not occurred, or in Weber's words:

> in the event of the exclusion of that fact [*C*] from the complex of the factors which are taken into account as co-determinants, or in the event of its modification in a certain direction, could the course of events [*E*], in accordance with general empirical rules, have taken a direction in any way different in any features which would be *decisive* for our interest? (Weber 1949: 180)

Thus, we ask whether either subtracting *C* from the course of events or modifying it to *C′* would have made a difference to *E*. Now in order to judge whether the change in *C* would have made a difference, we ask in a second step whether the occurrence of *E* was *objectively probable* given only the conditions or factors *F* that were co-present with *C* but now without *C*. For Weber, the question is thus whether the event $E \mid F.{\sim}C$ was "to be expected". If the answer is Yes, then *C* is judged to be causally insignificant, and if it is No, then *C* is causally significant. Weber then uses the term "adequate causation" to label cases where *C* did change the objective probability of *decisive aspects* of *E*, while he reserves the term "chance" or "accidental" causation to cases where *C* may have changed aspects of *E* that were not essential or decisive from the point of view of the inquiry of interest.

The details of Weber's analysis are not relevant for my argument, so I will not discuss them here. Let me just point out a worry. Weber takes it to be a necessary condition for causality that cause-events should make a difference to the probability of effect-events. However, there may be cases where the cause-event leaves the numerical probability of the effect-event unchanged but still is causally connected with it (and in fact responsible). The standard example discussed widely in the literature on causation is that of birth control pills and thrombosis. Pills cause (directly) blood clotting, but they also prevent it by preventing pregnancies, which themselves are one of the major causes of blood clotting. Now, the probability of a particular woman's getting thrombosis may be the same whether or not she takes pills because the probability-rise due to the direct effect exactly cancels the probability-lowering due to the indirect effect. Hence, Weber's method would wrongly conclude that (in this particular case) birth control pills did not cause thrombosis.

Let us assume, however, that Weber's method is sound. The point I make about it is that it does not help us in any way to learn about social capacities. The method is tailored to suit cases historians are interested in, that is, cases of singular causation. Whether or not a particular event did indeed raise the objective probability of another event is as good as irrelevant to the question whether it does so in other circumstances. Weber's method presupposes that all factors but the one we focus on behave regularly and that knowledge gained about them in other contexts is applicable to the case at hand too. But the knowledge it yields is tied to the one context under scrutiny.

The lesson of this section is this. Very widely used and important methods of causal inference in social science fail to yield knowledge about social capacities—for a number of different reasons. We might infer from this that Cartwright's scepticism is warranted. But there is a missing link in the argument. The inference presupposes that these are the best methods indeed to find capacities. In the next section I argue that there is something wrong with the way (at least some) social scientists use these methods. If that is true, one may grant Cartwright that there is no good (positive) reason to believe in the existence of social capacities. But I add the caution that there is no good (negative) reason to believe in their nonexistence either.

HOW WELL-FOUNDED IS SCEPTICISM ABOUT SOCIAL CAPACITIES?

So far, I have tried to give meat to Cartwright's scepticism about the existence of social factors with stable capacities. As any other form of scepticism, this variety can be read in two basic ways: as a positive disbelief and as a suspension of judgement. In this section I argue that Cartwright has good reason for the latter but little evidence for the former. In other words, I think agnosticism is a sensible stance regarding the reality of social capacities, and full-blown atheism is ill-founded.

Cartwright herself seems to oscillate between the two forms. Pretty dire sounds a joint statement with Jordi Cat in a paper on the German Historical School:

> The analytic method supposes that the causes of the phenomena of interest can be conceptually separated into distinct factors each of which has its own characteristic law of action. [. . .] Physics has been able to make effective use of this method in the study of motions; but political economy does not seem to lend itself to treatment by the analytic method.

And this is because:

> [The judgement about the above claim] is based on looking at cases of what is judged within the sciences themselves to be good practice. . . . (Cartwright & Cat 1998: 2)

I offer two arguments for the weaker reading, according to which it is more sensible to just suspend judgement rather than to claim positively that 'political economy does not seem to lend itself to treatment by the analytic method': a cheap and nasty one and a more involved one.

The cheap and nasty argument is that philosophers of science often make an unfair comparison of social with natural science. I would always tend to agree with Cartwright and Cat that '[p]hysics has been able to make effective use of [the analytic] method in the study of motions' and, in fact, in the study of many other phenomena. However, the claim that economics (or social science more generally) has failed to make use of the analytic method seems inequitable.

The social and those parts of the natural world where the analytic method has been applied most successfully differ in a number of important (and well-known) respects.[10] Let me mention just a few. Social phenomena (of interest) tend to be complex while natural phenomena (of interest) tend to be simple. Social phenomena (of interest) tend to be unstable and evolve over time while natural phenomena (of interest) tend to be stable and immutable. Social kinds tend to be interactive while natural kinds tend to be inert.[11] Social systems tend to prohibit experimentation while natural systems tend to allow it.

In my view, none of these differences motivates a principled distinction between natural and social science, but they tend to make causal inference in social science harder. Coupled with the (contingent) fact that social scientists tend to be interested in relatively young phenomena (such as the capitalist economy), it seems unfair to demand from them results comparable to those of their physics colleagues, who have had thousands of years to analyse their phenomena.

This argument is cheap and nasty indeed. Let me provide a second, more involved, argument. One of Cartwright's methodological principles for finding out about the nature of a subject matter is to investigate the best methods employed in the science that studies the subject matter and make inferences on that basis. This explains the selection of theoretical modelling in economics, the natural experiments movement, and Weber's singular causal inference scheme which have all been discussed above. I think that this methodological principle, defensible or not for other sciences, fails in the case of economics. The reason is simply that economics' so-called "best" methods are still characterised by a methodological oddity natural science was able to overcome in the seventeenth century.

The argument, in short, is as follows. Most economists, and, increasingly, other social scientists as well, presuppose a lot of theory in their empirical work. This, in my view, results in a certain disability to establish the existence of or facts about social phenomena.[12] Knowledge about capacities, however, is parasitic upon knowledge about phenomena. Hence, the theoretical bias also impedes learning about social capacities.

Let us therefore examine how social scientists establish phenomena. The best place to look for a sound methodology of social observation and measurement should be the early work of the Cowles Commission. Jakob Marschak, Tjalling Koopmans, Ragnar Frisch and others here established econometrics as a proper branch of economics through a combination of mathematics, statistics, and economics. Importantly, at least in the early years they regarded measurement as central to economics and adopted Kelvin's dictum "science is measurement" as the motto for the Commission.[13]

However, as much as they were interested in empirical investigation, theory was to play a strong part. In particular, they rejected the institutionalists' attempts to base economic analysis on empirical and historical investigation without recourse to theory. By contrast, they were aiming at a combination of theory and measurement in which the most fruitful use of both could be made. A good statement of this agenda can be found in Koopman's review of Arthur Burns and Wesley Mitchell's *Measurement of Business Cycles* (Burns & Mitchell 1946). He writes,

> . . . this reviewer [Koopmans] believes that in research in economic dynamics the Kepler stage and the Newton stage of inquiry need to be more intimately combined and to be pursued simultaneously. Fuller utilization of the concepts and hypotheses of economic theory . . . as a part of the processes of observation and measurement promises to be a shorter road, perhaps even the only possible road, to the understanding of cyclical fluctuations. (Koopmans 1995/1947: 492)

Although I accept that theory sometimes can play a role in observation, measurement and experimentation, I deny that it is necessary, and in particular I reject the dogmatism with which economic theory is acknowledged as *sine qua non* of economic measurement.

To see that theory is not necessary, consider William Stanley Jevons's investigation of the phenomenon of monetary inflation (Jevons 1863). Without an essential use of theory, and surely not of economic theory in the modern sense (which he was yet to co-invent), Jevons successfully establishes that the gold discoveries of the 1840s in Australia and California led to an increase in prices of about 13 percent. True, Jevons believed in the quantity theory. But with his investigation he tested the quantity theory at best and never presupposed it or used it in the construction of the measurement procedure. Further, Jevons makes use of the fact that prices are caused by what he calls 'the conditions of supply and demand'. Again, one might think that economic theory—in some sense—is sneaking in here, but in fact all that amounts to is a conceptual divide of causal factors into two groups.[14]

Popper is famous for stressing the principle "theory before observation". Now, if we accept—pace Jevons—that in order to make sense, observation must be made in relation to *some* theory, even Popper would regard it as pure dogmatism if it was taken for granted that it must be *a particular* theory and that that theory is beyond questioning. Economic theory (in the sense of a canon of general presuppositions and methods), however, does have such a position. It is not very surprising, then, that the empirical results of the early Cowles Commission have been disappointing.

The Cowles Commission is not the only place to look for this theoretical bias. I used their example because they are the inventors of modern economic measurement and therefore should speak with some authority. To turn to a more contemporary example, reconsider the natural experiments movement. On the face of it, it seems that by following this approach one can do solid empirical work without much theory.[15] However, a main criticism that has been levelled against it is exactly that the results cannot be interpreted in the light of economic theory and are therefore very limited in their usefulness. Even James Heckman, who himself is a proponent of the natural experiments movement and an ingenious developer of its methods, writes:

> Applications of this [natural experiments] approach often run the risk of producing estimates of causal parameters that are difficult to interpret. Like the evidence produced in VAR accounting exercises, the evidence produced by this school is difficult to relate to the body of evidence about the basic behavioural elasticities of economics. The lack of a theoretical framework makes it difficult to cumulate findings across studies, or to compare the findings of one study with another. Many applications of this approach produce estimates very similar to biostatistical "treatment effects" without any clear economic interpretation. (Heckman 2000: 85)

Why do economists get so excited about "theory" and "economic interpretation"? One reason is pointed out by Margaret Morrison: Sometimes there is a link between theory and our ability to carry experimental results to other contexts. Commenting on Cartwright's analysis of the Stanford/NASA gyro experiment, she writes:

> What the experiment shows is that in space the dragging effect produces gyro precession but that tells us nothing about frame dragging in other contexts; the theory tells us that this is a global effect and since the experiment bears out what the theory predicts will happen in space we consider it confirmed.
>
> (Morrison 1995: 168)

Morrison disagrees with Cartwright about whether we need the capacities framework in order to understand such exportability of results. Their disagreement need not concern us in the present context. What is important is Morrison's claim that if we have a theory which is universal throughout the specified domain, and we have good reason to believe our theory to be confirmed by a particular experiment or series of such experiments, then we also have good reason to believe that our experimental results are exportable to other contexts within that domain.

The complaint recorded by Heckman is, then, because we cannot bring the results achieved by proponents of the natural experiments movement to bear on economic theory—which, after all, is universal in its domain—we do not have an off-the-shelf procedure that tells us how to export our claims beyond the particular experiment that established it in the first place. Metaphysically, my reaction to this obstacle was to bite the bullet and accept that in economics there may be truths that are entirely local and not at all exportable: i.e. to accept that we must understand many of the results of the natural experiments movement as *ceteris paribus* causal laws, where the *ceteris paribus* condition ties the result to the experimental population.

However, this reaction may have been too hasty. There are probably many causal factors in economics that are not as stable as the universal capacities we know from parts of physics but are more stable than a *ceteris paribus* causal law. One possible aim of future research in methodology is to find a number of "off-the-shelf" principles that are informative about how to export claims established by a natural experiment to other contexts. For example, we may ask whether it matters that Card and Krueger's study investigated fast-food restaurants, that the study was conducted on the East Coast, or that the initial minimum wage was $4.75. Geoffrey Hodgson, I believe, made some advance on this question. In his book *How Economics Forgot History* (Hodgson 2001), Hodgson attempts to answer what he calls the 'problem of historical specificity', viz. the problem of knowing how historically (and geographically) specific a claim about socioeconomic systems must be to have the potential to be valid. His response consists essentially in relegating concepts and principles to the right level of abstraction, five of which he distinguishes (see his Table 21.2: 326–327). Certain concepts and principles pertain to *all* "open, evolving and complex systems". At this level, theorising is informed by evolutionary theory, general systems theory and complexity theory. At the second level, concerning all human societies, human instincts and psychology as well as general anthropological principles govern theorising. The usual laws of supply and demand come into play at the third level, which concerns only "civilised and complex human societies", and the fourth and fifth levels differentiate between kinds of socioeconomic systems.

I understand this schema to be a schema for exporting claims beyond the experimental population. Certain properties are shared by, say, all open, evolving, and complex systems. If an experiment establishes a new result about such a property, we should be able to export it to all other open, evolving, and complex systems and similarly for the other levels.

In my view, Hodgson's schema fails for a variety of reasons.[16] But what is immensely valuable about it is that it provides a starting point for research on a topic which I believe to be of fundamental importance for methodology. In their empirical work, economists have usually attempted shortcuts that exploit economic theory in order to, for example, identify causal parameters in an econometric regression or aspects of a measurement procedure. The methodological point of view put forward in this chapter suggests that no such shortcut is possible. We need a methodology that is informative about empirical ways to determine how projectible claims established on the basis of experiments are.

Therefore, I do not believe that the current state of economics is a good place to examine what is possible in economic analysis. The empirical road has not been walked yet and we do not know what fruits it will bear. Tjalling Koopman rightly distinguishes a "Kepler stage" and a "Newton stage" of scientific inquiry—one of empirical generalisation and one of fundamental law. But he errs that one needs to pursue them simultaneously. Empirical laws do not require fundamental laws to be found. By contrast, fundamental laws are void unless established on the basis of a range of empirical laws. There is no shortcut to fundamental laws that bypasses empirical laws.

The lesson of this section is that, pace Cartwright and Cat, there is no reason to lose hope. Yes, the record of finding social factors with stable capacities is poor. But it is poor because much empirical work done in social science has presupposed a particular theory about human behaviour. In my view, this has incapacitated the ability to establish real social phenomena, which in turn makes learning about social capacities a near impossibility. Giving up reliance on economic theory, and allowing social science to be a more empirical, more Baconian science, may result in learning about real social phenomena governed by real social capacities. To be sure, no one can predict that one day we will find only a single social capacity. However, I believe that there is no reason not to try.

CONCLUSION: HOW TO FIND SOCIAL CAPACITIES

The previous section ended with a mild optimism regarding the existence of social capacities. My hope to someday find such things is rooted in the conviction that there is something wrong with the way in which much of social science achieves its results. What I think is wrong is a certain dogmatism in an area one might label "phenomenal inference". Phenomenal inference is the establishment of phenomena on the basis of observations and measurements. Phenomena, the object of scientific explanations, do not lie around to be collected by the scientist, neither natural nor social. To take a hopefully uncontroversial natural scientific example, consider Newton's method of "deduction from the phenomena". What are the things Newton took as a basis for inferences? Surely not the naked observations of dots of light in the night sky (to take an example). Rather, he would construct a phenomenon such as the trajectory of a planet on the basis of observations or measurements made. Bas van Fraassen has recently remarked:

> Patrick Suppes had long emphasized that theories do not confront the data bare and raw. The experimental report is already a selective and refined representation, a "data model" as he calls it. This is especially true today, as Fred Suppe has emphasized, now that scientists routinely process gigabytes of data. It was already true in Newton's time when he claimed to deduce laws from the phenomena—for of course he used

> as basis very smooth functions distilled from thousands of astronomical observations. But it is true even of the idealized, simple observation report discussed by the logical positivists, as they themselves came to agree after some debate. (van Fraassen 1997: section 3.2)

There is no one way to infer a phenomenon on the basis of "thousands of observations". Especially in economics, choices such as the formula used to construct an index number or the specification of an econometric regression matter. As I have described in more detail elsewhere (Reiss 2002a: Ch. 4), in my view too much use of theoretical considerations is made in these inferences. It is as if Newton had used the laws in the process of constructing (or inferring) the phenomena from which he was to deduce his laws. Moreover, relatively theory-free approaches such as the natural experiments movement are regarded as deficient exactly for the fact that they cannot readily be connected to economic theory.

A more empiricist stance in economics would attempt to make inferences to phenomena with as little explicit reliance on theory as possible. Unlike proponents of natural experiments in econometrics themselves, I see nothing wrong with, say, finding out that, under certain conditions, an increase in the minimum wage causes employment to rise—even if that does not tell us much about elasticities. In a second step, research would proceed to investigate the stability of such a law. It would ask under what conditions minimum wages have what effects on employment. Finding a range of different conditions that affect the wage-employment relation differently, research may further proceed to drawing up hypotheses about mechanisms responsible for the different relations and thus explain them.

Effectively, this is partly what David Card and Alan Krueger do in their 1995 book. After the natural experiment in New Jersey, they analyse a second one in Texas; they reanalyse previous evidence from California, statewide, as well as international evidence from Puerto Rico, Canada, and Britain. Their aim, however, is only to undermine economists' traditional belief in the universal adverse effect of minimum wages. Hence it is enough for them to present a single case where an increase actually raised employment and to cast doubts on the validity of studies that find evidence for the opposite claim. They stop short of a systematic empirical investigation into the conditions and mechanisms responsible for the wage-employment relation. Further, they do try to explain their results by means of models of the kind discussed under "Exhibit I". Even for Card and Krueger, economic theory is sacrosanct. All they do is amend the simple model that predicts the negative effect slightly such that the resulting model predicts a positive effect.

Natural experiments à la Card and Krueger (as well as other models of causal inference that make little use of theory, such as Kevin Hoover's [2001] or the Bayes' Nets approach) do, nonetheless, provide a starting point. On their basis, a range of phenomena can be established, phenomena of the kind "under conditions *xyz*, increases in the minimum wage lead to higher employment", "for population *P*, schooling increases earnings" or "in system *S*, money causes prices". Phenomena can then be classified according to similarities and dissimilarities as well as compared and analysed. Again, on the basis of such a classification and analysis, attempts can be made to explain them with reference to underlying mechanisms. If we are lucky, such mechanisms have parts that can be used in the explanation of a range of different phenomena. They may be factors with (relatively) stable capacities.

This Baconian vision of social science is not new. It is essentially what social science would have looked like had the discipline followed Gustav Schmoller's methodological principles (see Schmoller 1998/1911). Ironically, then, I ask to use Schmoller's ideas to achieve what he himself thought would be impossible. As we have seen near the beginning of this essay, Schmoller argued against Mill that social factors do not have stable capacities that can be moved from situation to situation and that, in general, the analytic method is not applicable to social systems. But Schmoller may have been overly hasty in his conclusion. There has never been a prolonged attempt to do social science the way he envisioned it. Nonetheless, if we want to find social capacities, I do not currently see any better way.

NOTES

1. There seem to be differences, however, between the nineteenth-century concept of tendencies and Cartwright's concept of capacities. For a discussion, see Schmidt-Petri (this volume).
2. It is important to notice that we cannot salvage a law-as-regularity view by claiming that the account presented here simply misdescribes the actual situation because the "true law" is the *combined* law. The reason is that there are many cases in which the intervening factor cannot be brought under a more comprehensive law. Suppose that the motion in the second direction is brought about by a sudden gust of wind. According to Cartwright, there is no law that describes the operation of this kind of intervening factor in the regularity sense but the capacity (of the first factor) still holds.
3. The choice of the lemons model as an example is mine rather than Cartwright's.
4. Principles are few in number indeed: "self-interested actors maximise their utility" being one in microeconomics; "models should be solved using expectations derived from the model itself" being one in macroeconomics. They are meagre, as very little real-world behaviour is constrained by them.
5. It is important to note that the issue is not one of regularities versus causal powers. I do not want to defend a regularity view of law against a capacities view but rather indicate that the causal powers we find in social phenomena seem to be more fragile than the causal powers we find in many physical phenomena. Social causal powers seems to interact more frequently with other powers when they bring about a result.
6. This verdict is not an artefact of the choice of examples from simple mechanics. Even in more complex systems, such as systems described by particle physics, the general method employed by physicists remains the same. Often the synthetic step is more involved than adding forces by means of vector addition.

But still, the laws of the individual parts contribute in a principled way to the solution of the complex.

7. The original German term is "möglich" (possible) rather than "wahrscheinlich" (probable); I will stick with the usual translation, however.
8. Weber in fact thinks that there is an infinite number of such factors.
9. I sidestep issues about describing events *C* and *E* here. This is done in order to focus on the causal relation between *C* and *E* and not because these issues lack importance.
10. My comparison here involves paradigmatic cases on both sides. This is not to say, of course, that a vast number of cases from the less fundamental "natural" sciences (meteorology, geology, engineering, epidemiology . . .) more closely resemble my characterisation of the "social" sciences.
11. This is Ian Hacking's terminology; see for instance his 1999 article. His claim is that entities examined by social sciences tend to be responsive to our conceptions of and theorising about them in a way natural entities are not. Atoms do not care whether we have a good or bad theory about them while Marxism has changed a lot in the world.
12. By "social phenomena" I mean the social equivalent to Duhem's experimental laws or Hacking's or Bogen and Woodward's phenomena: stable features of the world that can be predicted (with some accuracy) and/or manipulated (with some accuracy), and/or explained (with some accuracy), or simply low-level social laws. (See Duhem 1991/1914; Hacking 1983; Bogen & Woodward 1988.)
13. This was later (1952) replaced by the diluted "Theory and Measurement". The reasons for this move will be apparent momentarily.
14. I have defended this interpretation of Jevons (Reiss 2001).
15. I am not saying that one can learn about causal relations from statistics without background knowledge. But that background knowledge can come from a variety of sources, including knowledge about institutions, previous econometric studies, common knowledge, etc. There is no requirement of economic *theory* here.
16. For a detailed discussion, see Reiss (2002b).

REFERENCES

Akerlof, G. (1970) 'The market for "lemons": Quality uncertainty and the market mechanism', *Quarterly Journal of Economics*, 84: 488–500.

Bogen, J., and J. Woodward. (1988) 'Saving the phenomena', *The Philosophical Review*, 97: 303–352.

Burns, A., and W. Mitchell. (1946) *Measuring Business Cycles*, National Bureau of Economic Research, Studies in Business Cycles, no. 2, New York: National Bureau of Economic Research.

Card, D., and A. Krueger. (1994) 'Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania', *American Economic Review*, 84: 772–793.

———. (1995) *Myth and Measurement: The New Economics of the Minimum Wage*, Princeton: Princeton University Press.

Cartwright, N. (1989) *Nature's Capacities and Their Measurement*, Oxford: Clarendon.

———. (1998) 'Capacities' in J. Davis et al. (eds) (1998) *Handbook of Economic Methodology*, Cheltenham: Edward Elgar.

———. (1999) 'The vanity of rigor in economics: Theoretical models and Galilean experiments', *CPNSS Discussion Paper Series* DP 43/99, CPNSS, LSE.

———. (2000) 'An empiricist defence of singular causes', in R. Teichmann (ed.) (2000) *Logic, Cause and Action: Essays in Honour of Elisabeth Anscombe*, Cambridge: Cambridge University Press.

———. (2002) 'In favour of laws that are not *ceteris paribus* after all', *Erkenntnis*, 57: 425–439.

Cartwright N., and J. Cat. (1998) 'Abstract and concrete knowledge: Why the historical school should matter to how we do economic theory today', Leverhulme/Thyssen Conference on 19th Century Historical Political Economy, King's College, Cambridge & MS, LSE.

Dalla Chiara, M. L. et al. (eds) (1997) *Logic and Scientific Methods*, Dordrecht: Kluwer.

Davis, J. et al. (eds) (1998) *Handbook of Economic Methodology*, Cheltenham: Edward Elgar.

Duhem, P. (1914; 2nd edn 1991) *The Aim and Structure of Physical Theory*, (trans.) P. P. Wiener, Princeton: Princeton University Press.

Glymour, C. (1980) *Theory and Evidence*, Princeton: Princeton University Press.

Hacking, I. (1983) *Representing and Intervening*, Cambridge: Cambridge University Press.

———. (1999) *The Social Construction of What?*, Cambridge: Harvard University Press.

Heckman, J. (1999) 'Causal parameters and policy analysis in economics: A twentieth-century retrospective', *NBER working paper* 7333.

Hendry, D., and M. Morgan. (eds) (1995) *The Foundations of Econometric Analysis*, Cambridge: Cambridge University Press.

Hodgson, G. (2001) *How Economics Forgot History*, London: Routledge.

Hoover, K. (2001) *Causality in Macroeconomics*, Cambridge: Cambridge University Press.

Jevons, W. S. (1863) 'A fall in the value of gold ascertained, and its social effects set forth', in W. S. Jevons (1884) *Investigations in Currency and Finance*, London: Macmillan.

———. (1884) *Investigations in Currency and Finance*, London: Macmillan.

Keynes, J. M. (1957/1921) *A Treatise on Probability*, London: Macmillan.

Koopmans, T. (1947) 'Measurement without theory', in D. Hendry and M. Morgan (eds) (1995) *The Foundations of Economic Analysis*, Cambridge: Cambridge University Press.

McMullin, E. (1985) 'Galileian idealization', *Studies in History and Philosophy of Science*, 16: 247–273.

Menger, C. (1963) *Problems of Economics and Sociology*, (trans.) F. J. Nock, Urbana: University of Illinois Press.

Morrison, M. (1995) 'Capacities, tendencies and the problem of singular causes', *Philosophy and Phenomenological Research*, 55: 163–168.

Nau, H. H. (1998) *Gustav Schmoller: Historisch-ethische Nationalökonomie als Kulturwissenschaft*, Marburg: Metropolis.

Neumark, D., and W. Wascher. (2000) 'Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania: Comment', *American Economic Review*, 90: 1362–1396.

Reiss, J. (2001) 'Natural economic quantities and their measurement', *Journal of Economic Methodology*, 8: 287–312; also in DP MEAS 14/01, *Measurement in Physics and Economics Discussion Paper Series*, CPNSS, LSE.

———. (2002a) 'Epistemic virtues and concept formation in economics', unpublished thesis, University of London.

———. (2002b) 'Review of *How Economics Forgot History* by Geoffrey Hodgson', available HTTP: <http://www.eh.net/bookreviews/library/0567.shtml.

Ringer, F. (1997) *Max Weber's Methodology: The Unification of the Cultural and Social Sciences*, Cambridge: Harvard University Press.

Schmidt-Petri, C. (this volume) 'Cartwright and Mill on capacities and tendencies.

Schmoller, G. (1911) 'Volkswirtschaft, Volkswirtschaftslehre und–methode', in Nau (1998) 215–368.

Teichmann, R. (ed.) (2000) *Logic, Cause and Action: Essays in Honour of Elisabeth Anscombe*, Cambridge: Cambridge University Press.

van Fraassen, B. (1997) 'Structure and perspective: Philosophical perplexity and paradox', in M. L. Dalla Chiara et al. (eds) (1997) *Logic and Scientific Methods*, Dordrecht: Kulwer.

Weber, M. (1949) *The Methodology of the Social Sciences*, (trans and eds) E. A. Shils and H. A. Finch, New York: Free Press.

Reply to Julian Reiss

Nature's Capacities and Their Measurement does a number of things. It distinguishes capacity ascriptions from context-dependent causal laws and from singular causal claims; it argues that various methods in science, especially in physics and economics, presuppose capacities; and it makes a case for the intelligibility of the notion of capacity, in particular arguing that capacities are not occult, or unscientific, or unverifiable. Not only are they themselves measurable; facts about capacities regularly play a part in justifying measurement procedures for things in other categories.

But there is one big issue that *Nature's Capacities*, in common with most other defenders of capacities or the related analytic method, says little about: How do we know, when we measure a capacity, that it is a *capacity* that we are measuring? This is the issue that Julian Reiss takes up. He not only asks, 'Are there capacities governing social phenomena', but also 'How can we know whether there are or not?'

As a foil for the discussion Reiss provides a helpful description of a number of methods in economics for making causal inference. Some of those methods are very powerful, if only they could be carried out in the ideal way. They are *bootstrapping* methods: The background assumptions plus the results imply the hypothesis or its negation. Reiss points out that some of these methods may allow us to measure the strength of capacity if there are capacities to be measured, but none of these go any way to establishing that it is a capacity that has been measured, a capacity rather than a context-dependent causal strength. This immediately suggests the question raised by students in my University of California-San Diego based seminar on idealization. Namely, are there any methods for bootstrapping to the claim that a result is due to a capacity and not just a context-dependent cause?

Reiss points out that a number of economists maintain that theory can do the job, and Margaret Morrison argues the same for physics. I would like to endorse Reiss's lack of enthusiasm for this proposal. My reason is that very often how theory does this turns out to be just by assertion that a given behaviour is universal. So what needs to concern us is the warrant for such assertions. Here I see no alternative to Reiss's own. The warrant must be broadly inductive, and he offers promising proposals for how to begin to rethink these inductions. This in turn lends plausibility to Reiss's own predilection for social capacities. For theories do not generally come from a burst of fantasy but are a response to a large amount of disparate kinds of what we might call "mid-fare" knowledge of how society works. We need to look on a case-by-case basis. But clearly there is reason for optimism that this warrent of mid-fare knowledge will provide reasonable support for a claim that capacities of certain kinds are at work in a given domain, even though it falls far short of supporting any proper theories. So I think I must accept that Reiss is right that a big bet against social capacities is currently a bad bet.

12 Cartwright and Mill on Tendencies and Capacities

Christoph Schmidt-Petri

INTRODUCTION

In this chapter I discuss to what extent Nancy Cartwright's appeal to John Stuart Mill's use of "tendencies" to defend or motivate her central notion of "capacity" is justified. My observations are meant to shed some light on the relation between these two concepts rather than to criticize or defend either, and so I shall argue that the differences between Mill and Cartwright are more significant than Cartwright's writings suggest. This need not be seen as a fundamental problem for Cartwright, as she has a number of other, independent arguments to defend her claim that capacities should be taken to be the fundamental building blocks of the natural and social sciences; it simply shows that she should probably not appeal to Mill to support this claim. In any case, Mill's concept of "tendencies" is also problematic: It is not clear whether it squares well with his empiricist account of laws.

Cartwright refers to Mill in a large number of publications, most prominently in *Nature's Capacities and their Measurement* (see Cartwright 1989; 1994; 1999a). There she literally takes "tendencies" and "capacities" to be synonymous:

> Mill believed that the laws of political economy and the laws of mechanics alike are laws, not about what things do, but about what tendencies they have. . . . Substituting the word "capacity" for Mill's word "tendency", his claim is exactly what I aim to establish in this book . . . I suggest that the reader take my "capacity" and Mill's "tendency" to be synonymous [until later in the book]. (Cartwright 1989: 170)[1]

It might appear surprising that Cartwright appeals to the writings of Mill, for his—official—Humeanism is something she is vehemently arguing against. However, the apparent similarity of her views with those of Mill, as she reads him, gives them a historical dimension which supplements her arguments from the practice of contemporary science.

I look at Mill's use of "tendencies"; Anscombe and Geach's criticism of it, which Cartwright uses to support her reading of Mill; and then argue that Mill's use of the language of "tendencies" is much less universal than most think. In fact he only uses "tendencies" in the particularly simple and relatively rare cases of what he calls the "mechanical" composition of causes. Furthermore, he is not realist about "capacities" as he himself uses this concept. Hence, I conclude, an appeal to Mill provides little support for "capacities" as a general and fundamental concept of the natural and social sciences.

MILL ON TENDENCIES

I first want to briefly outline why Mill uses "tendencies" in natural and social science. I think it is important to note that he does so for primarily methodological, that is, entirely practical rather than metaphysical reasons.

In the natural and social sciences, particularly in economics, but also in the moral sciences, Mill sees a multiplicity of causes giving rise to whatever phenomena we observe. Just like Cartwright, Mill believes that the world is "dappled" in the sense that there are very few occurrent regularities.[2] Hence there is little scope for a systematisation of our experiences just by regrouping phenomena under phenomenological laws by induction. In any case, these empirical laws would mostly be uninteresting, as they would generally be restricted in their range of applicability to the context in which they have arisen and hence not be stable enough for useful predictions. However, Mill also believes that experience shows that the phenomena are produced by relatively few causes. In the domain of economics, for instance, man's desire for wealth is by far the most important cause. By looking at just this desire we can relatively accurately predict what will happen in the markets, provided we manage to present a good description of the circumstances in which this cause operates. Two other causes also operate constantly in the economic realm by directly counteracting this desire for wealth, and these consequently always need to be taken into account when making predictions. These are man's laziness, in Mill's words, his 'aversion to labour' (Mill 1836: 52) as well as his myopic time preference, his 'desire of the present enjoyment of costly indulgences' (Mill 1836: 52). Mill, suggesting mechanical interaction, further notes that these 'accompany it always as a *drag*, or *impediment*' (Mill 1836, 53, italics added). Although other causes operate only occasionally, there will always be some that do.

The laws of the discipline of economics are deductively derived from putting these desires of man into an "economic" context. They state, in abstract, what would happen in the economic realm if no other causes were operative. In economics, experience shows that such theorising may already enable one to predict relatively efficiently a lot of the actual phenomena. Nonetheless, other "disturbing" causes are operative. Therefore, Mill says, if one wants to predict phenomena accurately one should not overconfidently predict *actual* results, but only a "tendency" to the result: 'a power acting with certain intensity in that direction' (Mill 1836: 67).

The situation in the natural sciences is similar. Gravity always operates on every object; however, not every object actually falls to the ground as the law would seem to predict, considered just by itself. Other causes also operate on any individual object, which may offset the gravitational "pull" entirely. According to Mill, objects 'have a tendency' to fall even when, as described, they do not. He phrases the general point thus: 'All laws of causation, *in consequence of their liability to be counteracted*, require to be stated in words affirmative of tendencies only, and not of actual results' (Mill 1843: 445, italics added). Mill, then, uses the language of tendencies specifically when talking about causes that are impeded in their operation by other causes.

AN OBJECTION TO MILL

The usual interpretation, and the one Cartwright adopts, is to read Mill as making claims about tendencies of things to behave in some way. A tendency, in this sense, would be a feature of an object—a property, or a property of a property.[3]

However, this reading invites the following objection to Mill, the *locus classicus* of which is Anscombe and Geach (Anscombe & Geach 1961: 101). They argue that Mill's use of "tendencies" as delineated above is incompatible with his "official" Humeanism about laws and causation. The reasoning is simple: Officially, Mill thinks that causation is nothing but constant conjunction of cause and effect. Of causal laws, it is then nonsense to say that they are "true" or that the effect of any cause is "fully realised", as Mill does, if, actually, there is no constant conjunction. But this is exactly what happens in the case of interference. Given Humeanism, the absence of actual constant conjunction must mean that there is no law.[4] But Mill, it is observed, does not go all that far. When he says that in such cases, there is "interference" and that the laws are nevertheless true, as they are actually about tendencies of things, which just happen not to be realised (or, counterfactually, in the absence of the interference would be realised), then, Anscombe and Geach contend, he is departing from his Humeanism, contrary to what he may believe. In fact, adopting tendencies is to subscribe to a rather more Aristotelian metaphysics.

Cartwright endorses this objection and how it forces Mill into accepting tendencies (and she also sees a further problem, which I discuss later). The problem I see with this argument is the following. Mill does not in fact claim that the relevant laws are laws *about* tendencies of objects to behave in a particular way.[5] He merely says that these laws *require to be stated* in words affirmative of tendencies only. Mill's language does not have the existential import both Anscombe and Geach as well as Cartwright see—he does not say that there are such things or properties as "tendencies".[6] What Mill is after is a way of stating causal laws—that is, laws of constant conjunction—such that these laws are not falsified just because the causes do not operate one at a time but simultaneously. Mill's point is merely verbal, or about the representation of laws, whereas Cartwright and Anscombe and Geach take him to make an assertion about the metaphysics, or the object, of laws.

The standard scenario may help to illustrate.[7] Consider the case where some object is pulled in a northern direction by some force, and in an eastern direction by another force. Suppose further that these forces are of equal strength; as a result, the object moves northeast (this is a philosophically nontrivial fact of mechanics). What is uncontroversial here is that the object *actually* moves neither "just" north nor "just" east—it moves northeast. What is controversial is how to best analyse what is "really" going on.

Cartwright claims that Mill fails in his analysis of this case. According to Cartwright, Mill, because he does not want to engage in talk of tendencies in a substantial sense here (though he does so elsewhere), talks as if the body was in motion towards the east as well as towards the north (Cartwright 1989: 179). And this is, to all empirical appearances, just plain wrong, because the body moves in precisely one direction—northeast. The Millian stipulation of motion where there is none is not worthy of an empiricist, and certainly much less compatible with empiricism than the adoption of tendencies, which here just might not be realised. What would be accurate to say in this case is that the body has a tendency to move eastwards and a tendency to move northwards—but Mill does not say this. It turns out that Mill is right when, and only when, he is using tendencies. Hence, he really is giving up his Humeanism.

I think that Cartwright, as well as Anscombe and Geach overstate their case. It is quite possible to make sense of the above scenario without using "tendencies" in a deep sense.[8] Mill is discussing the composition of causes: in particular, in what sciences we can rely on a "mechanical" composition of causes, as in Newtonian vector addition—and hence can rely on the deductive a priori method—and in what sciences "chemical" combinations of elements render such a neat deduction impossible and extensive testing inevitable (Mill 1843: Bk. III, Ch. IV, §1). Although Mill believes that the mechanical composition is the rule (Mill 1843: 373), he is aware that this principle 'by no means prevails in all departments of the field of nature' (Mill 1843: 371). As mentioned, two interesting cases are economics and mechanics, in which the causes do combine mechanically.

What I consider a relevant observation is that in the contested passage, Mill is talking specifically about those cases in which causes *do* combine mechanically, and have been *established* to combine mechanically. Of these only he says that

> In this important class of cases of causation, one cause never, properly speaking, defeats or frustrates another; both have their full effect. If a

> body is propelled in two directions by two forces, one tending to drive it to the north, and the other to the east, it is caused to move in a given time exactly as far in both directions as the two forces would separately have carried it; and is left precisely where it would have arrived if it had been acted upon first by one of the two forces, and afterwards by the other.
>
> (Mill 1843: 371)

Cartwright interprets Mill's mention of both causes "having their full effect" as if both effects should thus be simultaneously realised in the sense of becoming individually visibly apparent (she changes the example slightly by talking about the more vivid concept of motion). Of course, if the causes did operate one after the other, it would have to be admitted that one *could* say that both effects were fully realised, in the strongest sense imaginable—there would first be a motion to the north, then to the east. And this is where Mill's argument starts: If and only if it is both the case that both effects are fully realised when the causes operate consecutively *and* the result of both causes acting simultaneously is exactly the same as when they do operate consecutively—and he only talks about cases where this is a testable, empirical and established matter of fact rather than a "counterfactual supposition"—*then* we are in the lucky circumstance of being able to derive this result deductively, that is to say, in these cases there is Composition of Causes. And it is only in such cases that Mill talks of "tendencies" and only when both causes actually operate simultaneously.[9]

Mill's point is, first of all, one about the most efficient method. We may use the comparatively convenient deductive method if it has been established that causes combine in a way that is amenable to such deductive reasoning.[10] In those cases we can talk of the individual causes as having "tendencies".

The question now is whether to make sense of this phenomenon we must invoke the reality of tendencies, as I think is Cartwright's claim.[11] Mill's language might suggest this, yet the question is whether this really shows that Mill is a (closet) realist about tendencies. Mill's further examples are illustrative, as in all of them he describes how a consecutive operation of causes yields the same result as a simultaneous operation. For instance, he mentions a stream running into a reservoir that at the other end has a drain that simultaneously releases exactly as much water as is entering—the result of this is that the water level in the reservoir remains unchanged. He says that 'even if the two causes which are in joint action exactly annul one another, still the laws of both are fulfilled' (Mill 1843: 372). Although Mill does not specify which "laws" he is thinking of, he refers to the stream that 'tends to fill [the reservoir] higher and higher' and the drain which 'tends to empty it'.

But does this require one to take these laws to be laws about tendencies? I do not think so, for the simple reason that when Mill talks of the consecutive operation of the causes, he does not use the "tendency" vocabulary. And surely if he does not even use the word it would be far-fetched to claim that he is nonetheless talking about tendencies. He says that the two causes, if they acted, 'would *produce* effects' (Mill 1843: 372), not that they would "tend" to produce effects. In other words, Mill only states laws using the language of tendencies when he expects the law to be operating in a causal context in which the effect will not come to full realisation (i.e. will not produce the effect it would produce were no other causes operative) and when it is also the case that the final result is exactly the same as if all causes had operated in isolation, but one after the other.

This is quite compatible with our everyday usage of "tendency", where it is typically implied that the effect did not get realised. For instance, one would expect the assertion that people who start to go running several times a week "tend" to lose weight—rather than that they do lose weight—to be continued with an explanation of how the effect of running is in fact counteracted, for instance, by increased energy intake. And it is also understood that the running and the additional energy intake do not "interact" in special ways: Both activities have the same effect that they would have in the absence of the other cause, as a sufficiently long period of running followed by a sufficiently long period of additional energy intake would confirm.[12]

A "tendency" statement, on this reading, is thus a statement not about "undercover" goings-on, but about how causes combine, namely, that the composition is of the mechanical kind; that is, that it is a case of Composition of Causes. The regularity highlighted is not one about the stability of the mechanisms or "tendencies" that conjoin to produce the result, but about a feature of the conjunction itself, namely, that the conjunction of causes yields the same effect as if the causes had operated consecutively.

The point of this somewhat lengthy demonstration is this: An argument that runs from Mill's use of "tendency" statements to the assumed stability of the "tendency" mechanism, and from then on maybe to a mechanical composition turns Mill's approach on its head. It is precisely the regularity in the composition that is represented in the "tendency" statement. First comes the observation of mechanical composition or "stability" of causes then only the use of the "tendency" language.

This approach can only be faithful to Mill if in cases of "chemical" combination of causes he does not use the language of tendencies (though this by itself would clearly not suffice to establish my interpretation). This is indeed so. In these cases Mill says that

> most of the uniformities to which the causes conformed when separate, cease altogether when they are conjoined; and we are not . . . able to foresee what result will follow from any new combination, until we have tried the specific experiment. (Mill 1843: 371)

Furthermore he says that concerning the

> combinations of elements which constitute organized bodies; . . . the phenomena of life, which result form the juxtaposition of those parts in a certain manner, bear no analogy to any of the effects which would be produced by the action of the component substances considered as merely physical agents.
>
> (Mill 1843: 371)

Again, this corresponds to ordinary language use of "tendency". It would indeed seem odd to talk of "tendencies" in cases of "chemical" composition of causes. For instance, if in one experiment chemical elements *A* and *B* combine to *X*, but in another *A* and *B*, together with *C*, combine to *Y*, it does not seem accurate to say that in the latter case, *A* and *B* had a "tendency" to form *X*, which was in some sense "offset" or counteracted by the addition of *C*.[13] The facts here seem most accurately stated without using "tendencies" altogether (even though a counterfactual of the form: "in the absence of *C*, *A* and *B* would have formed *X*" is true just as it would have been if *A* and *B* had had a tendency to form *X*, which was offset by *C*).

Mill, to conclude, uses the language of tendencies only in very specific cases. He is under no illusion that causes do not always combine mechanically, and, more importantly, that whether they do is itself to be determined empirically (not just counterfactually). The majority of the "interesting" causes in economics and physics may combine mechanically, but this itself needs to be established empirically. Hence it would be unwarranted to conclude that for Mill, "tendencies" are a fundamental and ontological building block of the sciences.

MORRISON ON CAPACITIES AND TENDENCIES

These observations need to be contrasted with an argument by Margaret Morrison, who also argues that Cartwright's "capacities" are different from Millian tendencies (Morrison 1995). Her argument is that the former do not remain constant in the face of all interfering causes while the latter do (and that hence to invoke capacities rather than context-dependent causal laws seems unwarranted). Capacities do not always produce their characteristic effect (even when there is no capacity-modifying interaction), but tendencies in the Millian sense do universally make their characteristic contribution. Indeed, Morrison observes, Mill takes the leap of referring to tendencies even when they are not measurable because counteracted.

I endorse Morrison's arguments. However, once it is realised that Mill uses the language of tendencies only when he has previously established that the causes produce the same effect when operating simultaneously as when operating consecutively, to reproach Mill for not following his empiricist teachings seems overly strong. Mill is not appealing to counterfactuals without empirical basis. But more importantly, the limited universality of capacities with respect to tendencies appears no longer as a problem for Cartwright: Mill simply never talks of tendencies when causes do not combine in the "correct" way. He therefore only artificially achieves the "universality" of tendencies which in fact is itself limited to particularly fortuitous scenarios. Hence although Morrison's arguments are valid for those cases in which Mill talks of tendencies, because he does not do so universally, her case against Cartwright is itself less universal than it seems. However, Morrison's general observations that capacity claims are of limited universality is nevertheless strengthened if my observations are correct.[14]

MILL AND THE REALISM OF CAPACITIES

"Capacity" in the sense that Cartwright uses the word is a technical term that overlaps but does not correspond one-to-one with its usage in ordinary language.[15] What exactly the metaphysical import is of saying that *X* has the capacity to φ, in Cartwright's sense, is therefore sometimes difficult to discern, and Cartwright has been criticised for being too vague about this quite crucial concept (which in ordinary language gets used quite indiscriminately).[16] In fact, Cartwright even says that 'I . . . have no metaphysical views about dispositions versus capacities versus powers. I choose the word "capacity" since it is less often used by others; hence it carries fewer presuppositions with it' (Cartwright 2002: 3).[17]

What matters for my purposes is that Mill clearly did not endorse the realism of capacities in one of the senses it is used by Cartwright. Hence, to the extent that Cartwright's and Mill's conceptions of capacities coincide, there is a clear case that Mill cannot be adduced to support Cartwright's realism about capacities. Though Cartwright never says that what Mill calls "capacities" corresponds to what she calls thus—she restricts her arguments to Millian "tendencies"—and clearly Mill thinks these are very different concepts, capacities à la Mill bear enough resemblance to capacities à la Cartwright to justify the following quotations.

Mill says: '[A] capacity is not a real thing existing in the objects, it is but a name for our conviction that [these objects] will act in a particular manner when certain new circumstances arise' (Mill 1843: 337). He presents an example:

> Putting a coat of white paint upon a wall does not merely produce in those who see it done, the sensation of white, it confers on the wall the permanent property of giving that kind of sensation. . . .
>
> (Mill 1843: 337)

He therefore agrees that the wall has acquired a permanent property. But he continues: 'no one now supposes the property to be a substantive entity "inherent" in the object' (Mill 1843: 337). Another example Mill uses is gunpowder. Gunpowder is in a "state of preparation" which conjoined with its lighting will result in an explosion. But this property of gunpowder is reducible to a purely physical description as it "consists in a certain collocation of its particles relatively to each other" (Mill 1843: 337).

Although in these cases it is arguable whether Mill's sense of "capacity" corresponds to Cartwright's, the following will suffice to drive the point home. Mill talks about the interaction of gravitational and magnetic forces, an example Cartwright repeatedly uses.

> The earth causes the fall of heavy bodies, and it also, in its capacity of a great magnet, causes the phenomena of the magnetic needle . . . The purpose to which the phraseology of Properties and Powers is specially adapted, is the expression of this sort of cases . . . it is usual to say that each different sort of effect is produced by a different property of the cause. Thus we distinguish the attractive or gravitative property of the earth, and its magnetic property: the gravitative, luminiferous, and calorific properties of the sun . . . (Mill 1843: 345)

However, Mill continues by saying that

> These are mere phrases, which explain nothing, and add nothing to our knowledge of the subject; but considered as abstract names denoting the connexion between the different effects produced and the object which produces them, they are very powerful instruments of abridgment. (Mill 1843: 345)

For Mill, then, talk of "capacities" may be pragmatically useful, but in his opinion such "capacities" will always be reducible to more primitive physical facts.

CONCLUSION

I have argued that Cartwright's appeal to Mill's writings provides relatively little support for her conception of capacities. This, of course, must not to be taken to constitute an argument against her own views, which I did not discuss in any detail. My observations are not likely to damage her approach as such; if right, they simply show that it has less support from the ultra-empiricist Mill than might have otherwise have been supposed.

ACKNOWLEDGMENTS

I thank Luc Bovens, Nancy Cartwright, Daniel Hausman, Marco Nievo, Stathis Psillos, Julian Reiss, and James Woodward for helpful comments.

NOTES

1. Later the synonymy is relativised: '"Capacity" is reserved for a special subset of these [tendencies]—those tendencies which are tendencies to cause or to bring about something' (Cartwright 1989: 226; see also Cartwright 1998: 45, 48; 1999b: 4).
2. This, of course, is not intended to be a complete description of Cartwright's position. For a discussion of the sense in which she sees the world as "dappled" (see Lipton [2002] and Cartwright's reply [Cartwright 2002]).
3. Cartwright says explicitly that she discusses claims *about* tendencies (Cartwright 1989: 178), in order to distinguish such from "tendency laws"—these being laws of irregular correlation only.

   In interpreting Mill's language of tendencies in his moral philosophy, the second of these interpretations is prominent: Urmson (1953) has claimed that only *types* of actions have tendencies to *P*, these being a "more often than not" correlation between its tokens and the effect *P*; this is also endorsed by Quinton (1973). Champlin & Walker (1973) instead argue that a *token* action has a tendency to *P* if, among its many effects, most of them *P*. In all of these, a "tendency" is nothing beyond some type of correlation. These readings must, however, be wrong: Mill clearly uses "tendencies" to avoid having to talk about exceptions altogether, rather than to model them, as Cartwright realizes.
4. But see Mackie (1980: Ch.3, 75 in particular).
5. At the very least it is accurate to claim that these formulations do not show that he is committed to such tendencies. But there is no better evidence in Mill for the claim that he is so committed.
6. My objection might also apply to Hausman's reading of Mill: 'Tendencies are the causal powers underlying the genuine regularities . . .' (Hausman 1992: 127).
7. See Creary (1981); Cartwright (1980, 1983); Gibson (1983); Psillos (this volume).
8. But this is not to rescue Mill from all problems with his notion of tendency . In particular it does not show that Mill's use of "tendency" is in the end compatible with a constant conjunction view of causation.
9. The common criticism that Mill does not provide sufficiently detailed rules of composition for tendencies or even neglects this problem is therefore somewhat besides the point. See most prominently Hausman (2002: §§ 4, 5) who objects that 'To speak, as Mill does, of a deductive method, is misleading because the law governing the conjoint operation of causes cannot be deduced from the laws governing the component causes separately' (Hausman 2002: 302). Such objections ignore the fact that Mill only ever talks about tendencies when these "laws" *are* well known. Mill does not need, as Hausman claims, *assumptions* of "additivity", "compositionality", or "some sort of persistence or non-interaction" (Hausman 2002: 303), nor is it true that 'Mill has no answer to those who doubt whether causal laws of complex phenomena such as economies can be deduced from the laws of the separate causes' (Hausman 2002: 304). But it certainly is true that Mill generally provides not enough detail about how he conceives of the operation of "tendencies".

10. To my knowledge Mill does not explain how stable across different contexts he thinks this observation will be.
11. 'What makes capacity claims true are facts about capacities' (Cartwright 1999a: 72).
12. The claim is not that this is in fact so (it is not) but that this is what the typical utterer of such a "tendency" statement would want to express. The testing of such claims may in practice also be done differently.
13. It may seem accurate if *X* and *C* combined to *Y*. But then the composition would not be "chemical" in the relevant sense but "mechanical". Note that actual chemical reactions extremely rarely take place in this "chemical" way; in fact, on some level of analysis, they might never.
14. As Cartwright does not agree with Morrison that capacity claims are less than universal, her reply to the present objection might similarly be to give up Mill's even further reduced endorsement (Cartwright 1995).
15. This is sometimes not realised (e.g., Glennan 1997).
16. Cartwright does give a precise definition in her 1998 encyclopaedia entry on "capacities", but this is restricted to its use in economic methodology (see, e.g., Psillos this volume: §6.2).
17. Though in various places Cartwright contrasts capacities with dispositions (e.g., Cartwright 1999a: §3.4).

REFERENCES

Cartwright, N. (1980) 'Do the laws of physics state the facts?', *Pacific Philosophical Quarterly*, 61: 75–84.

———. (1983) *How the Laws of Physics Lie*, Oxford: Clarendon Press.

———. (1989) *Nature's Capacities and Their Measurement*, Oxford: Clarendon Press.

———. (1994) 'Mill and Menger: Ideal elements and stable tendencies', *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 38: 171–188.

———. (1995) 'Reply to Eells, Humphreys and Morrison', *Philosophy and Phenomenological Research*, 55: 177–187

———. (1998) 'Capacities', in *The Handbook of Economic Methodology*, U. Mäki (ed.) Cheltenham: Edward Elgar.

———. (1999a) *The Dappled World*, Cambridge: Cambridge University Press.

———. (1999b) 'The vanity of rigour in economics: Theoretical models and Galilean experiments', *CPNSS Discussion Paper Series* DP 43/99, London School of Economics.

———. (2002) 'Reply', in Book Symposium on *The Dappled World*, *Philosophical Books*, 43: 271–278.

Anscombe, E., and P. Geach. (1961) *Three Philosophers*, Oxford: Blackwell.

Champlin, T. S., and A. D. M. Walker. (1973) 'Tendencies, frequencies, and classical utilitarianism', *Analysis*, 34: 8–12.

Gibson, Q. (1983) 'Tendencies', *Philosophy of Science*, 50: 296–308.

Glennan, S. S. (1997) 'Capacities, universality, and singularity', *Philosophy of Science*, 64: 605–626.

Hausman, D. (1992) *The Inexact and Separate Science of Economics*, Cambridge: Cambridge University Press.

Hausman, D. (2002) 'Tendencies, laws, and the composition of economic causes', in U. Mäki (ed.) *The Economic World View*, Cambridge: Cambridge University Press.

Lipton, P. (2002) 'The reach of the law', *Philosophical Books*, 43: 254–260.

Mackie, J. L. (1980) *The Cement of the Universe*, Oxford: Clarendon Press.

Mill, J. S. (1836) 'On the definition and method of political economy', as reprinted in excerpts in Hausman (1994) *The Philosophy of Economics*, 2nd edn, Cambridge: Cambridge University Press.

———. (1843; 1974) *A System of Logic*, Collected Works, vols. VII & VIII, London: Routledge.

Morrison, M. (1995) 'Capacities, tendencies and the problem of singular causes', *Philosophy and Phenomenological Research*, 55: 163–168.

Quinton, A. (1973) *Utilitarian Ethics*, La Salle: Open Court.

Urmson, J. O. (1953) 'The interpretation of the moral philosophy of J. S. Mill', *Philosophical Quarterly*, 3: 33–39.

Psillos, S. (this volume) 'Cartwright's realist toil: From entities to capacities'.

Reply to Christoph Schmidt-Petri

Christoph Schmidt-Petri's defense of J. S. Mill's empiricism in the face of his talk of tendency laws blends historical and philosophical analysis neatly into one, and I find it convincing when he says that I am quite possibly wrong to suppose Mill at one with me in endorsing tendencies. Rather, Schmidt-Petri argues that, for Mill, to talk about tendency laws is not to endorse the existence of tendencies but rather to point to the fact that regularities exhibit a certain pattern: What regularly follows when a number of factors co-occur is the "sum", in some sense, of what would happen were they to occur consecutively.

I suspect Schmidt-Petri is right in rejecting tendencies on Mill's behalf. The position he defends is consistent with Mill's empiricism, and it fits the texts. But as Schmidt-Petri points out, I do defend tendencies. That's because I do not believe that there are regularities of the kind Mill needs for his account, because what regularities there are in physics and political economy do not involve only factors that can be admitted in an empiricist ontology.

Suppose though that I am wrong. We do not need to refer to interferences, triggers, shields, nomological machines, or the like to state the relevant regularities that will, as Schmidt-Petri argues, save Mill's empiricism. Still reference will not, I should like to point out, save what is called the "Mill–Ramsey–Lewis" view of laws. In this view, laws are those true regularities that best balance breadth of coverage and simplicity. I think it is worth here recalling an old point from *How The Laws of Physics Lie*: Because we want "true" regularities, to secure any predictive or explanatory power, we shall have to sacrifice simplicity entirely—and once we have done that, the empiricist sense of laws as the simplest regularities will turn out to be a sham.

Consider Mill's case of vector addition of forces, as we might do it in elementary mechanics today. I suppose, for the sake of argument, that there is a true regular association expressed by "$f_t = ma$": The total force on an object is always equal to its mass times its acceleration. I take it that we want to be able to use this to explain or predict—on a scientific basis—actual accelerations. That requires, for a given object, a law (a regularity law!) that links that object's circumstances to the total force on it. Now even if only gravity were at stake, that law would be incredibly complex, as it must have a term for every tiny piece of every material object in the universe. Of course we do not do this in a real scientific explanation or prediction. Instead we idealize to some short description that will yield accurate enough predictions for the purposes at hand. But, as Craig Callender has pointed out in describing my view, that would leave us with an unpleasant trade-off: Either we never explain what really happens or we admit an indefinite number of excruciatingly complex laws.[1] This latter is a problem for the Mill–Ramsey–Lewis view; however, as none of these incredibly complex descriptions are apt to recur even once, let alone often. So, with the exception of $f_t = ma$, the laws we need to explain or predict what really happens will not be regularities after all.

Of course we know what these very complex laws will look like. As Schmidt-Petri reminds us, Mill says they will be a sum of terms describing what would have happened if each cause '*had* operated in isolation' (Schmidt-Petri this volume: 296). To say that is to describe a common pattern among these very complicated "laws" that we need for explanation and prediction, not to reduce them to a handful of simple ones. Also note what a strange tactic we must take to identify these terms: Via subjunctive conditionals, and worse, conditionals that are never instantiated even once, let alone regularly. So we don't find any empiricist regularities here either—not that it would have helped with the original problem anyway. The lesson I want to draw is that we must not take Schmidt-Petri's probably successful defense of Mill's empiricism to double as a defense of the contemporary view of laws named in part after Mill.

NOTES

1. Personal correspondence, University of California-San Diego Seminar, January 2005.

Part III

Antifundamentalism and the Disunity of Science

13 For Fundamentalism

Carl Hoefer

INTRODUCTION

Recent philosophy of science has been marked by a strong wave of support for heterodox views of the nature and ambitions of the natural sciences and the relationships among the various sciences. The themes of this new wave are disunity of science, autonomy (of each of the several sciences), antireductionism, anti-imperialism (of physics), and, most recently, antifundamentalism. Nancy Cartwright has been an important leader of this new wave, and unlike most earlier philosophers of science she has a political agenda—a very progressive one—that accompanies her views on science. She calls for society to support science that demonstrably works to help people live better and not to give undue eminence (nor financial support) to so-called fundamental physics, with its ever-larger and more expensive particle accelerators.[1]

But progressive goals are never, in the end, well served by flawed arguments. And the arguments given by Cartwright against fundamentalism—i.e. against the traditional view that there are true fundamental laws of nature that govern the behaviour of matter at all places and times—are, I believe, flawed. The goal of this chapter is to mount a counterattack in defence of fundamental laws. But to defend fundamental laws is not to challenge the overall accuracy and utility of Cartwright's evolving picture of how science works. Nor is it to defend gross imbalances in society's approach to the funding of scientific research. Even if there are ultimate fundamental laws out there, waiting for us to discover them, it hardly follows that the best way to spend the next $10 billion on science is to add an order of magnitude to the energy of some underground proton-proton collisions.

But, along with gargantuan particle accelerators, Cartwright's arguments put in a bad light a different, much less expensive endeavour: (much of) the current philosophy of physics. This chapter aims to help justify the practice, common among philosophers of physics, of taking for granted that there are fundamental physical laws.

WHAT IS FUNDAMENTALISM?

The first thing to note is that Cartwright's baptism of her philosophical opponent is a real linguistic coup. Who wants to call him- or herself a fundamentalist? Despite this, the term is so apt that I will continue to use it. A fundamentalist believes in something rather ultimate and mysterious; not God, of course, but something that nevertheless "governs" the whole universe, from top to bottom. What she believes in is the *fundamental* law(s) of nature. These are what physics has been seeking, and getting closer and closer to actually grasping, since the time of Descartes. They are truths, expressable in mathematical language, that accurately describe the behavior of all things in the physical world, at all times and places. This view has been standard among physicists, and most philosophers of science, for at least a hundred years.

There are a number of questions about fundamental laws that do not matter for this essay. For example: do they have some kind of physical necessity, or are they rather mere Humean regular associations? Would fundamental laws (if they existed) explain everything—or nothing? Do the laws need to be explained, themselves, to have explanatory power? Do all causal or other nonfundamental laws need to be derivable somehow from fundamental laws, in order to be real? None of these issues is pertinent to Cartwright's attack nor to my defence.

WHY I AM A FUNDAMENTALIST

Fundamentalism only makes sense in the context of certain other philosophical assumptions—widespread ones, to be sure, but not universal. One has to believe in an external physical world and that we have at least some nontrivial epistemic access to it. One has to believe that it would be nice to have explanations for the widespread and reliable regularities that we observe in the world; and that true universal laws, if there were any, could play at least some part in providing such explanations.

Given these minimal starting assumptions, a fundamentalist believes that the recent history and current state of knowledge in physics provides strong and variegated evidence that there are indeed universal fundamental laws with which all physical phenomena are in accord. Later we will look at, and try to answer, Cartwright's arguments for the weakness of this evidence and implausibility of the fundamentalist's picture. But here it will be helpful to note that Cartwright is ready to offer an explanation of some of the remarkably precise (and often useful) regularities in nature that physics has been able to disclose. She favours an explanation that invokes stable causal capacities in nature and a "patchwork of laws" neither universal nor fundamental. By contrast, some philosophers—perhaps e.g., van Fraassen—would deny that we can or should seek *any* explanations of nature's regularity. I will not try to defend fundamentalism against these more seriously sceptical views here.

Why, then, do I think that physics today gives strong evidence for the existence of true universal and fundamental laws? Before getting down to cases, let me note that by "laws", throughout, I will mean usually mathematical equations, and never so-called "causal laws". Sometimes a fundamental law may take the form of a prohibition or nonexistence claim (e.g., the Pauli exclusion principle), but most of the time they take the form of mathematical equations relating one or more functions to each other or to a constant such as zero. The equations, as Russell (Russell 1912) pointed out, are often such as to suffer no easy reading in terms of causation. Whether or not Russell was right to claim that causation had been banished from fundamental physics, we can at least assume that the laws we are discussing are not usually best read as mathematized ways of saying "*X*s cause *Y*s". Nor are they intended to be read with a tacit "*ceteris paribus*" at the end or beginning. They are universal, exceptionless, precise regularities.

My reasons for thinking there must be such things are probably no different from those of most other fundamentalists. We have already found such mathematical regularities that are true or very close to true wherever we are able to check. And their nature is such that we can imagine them being replaced one day by other mathematical laws still more accurate or universal (as has happened before in the history of physics), but not their being superseded by nonmathematical statements of some kind, or given up without any replacement at all. To go any further, we need to start looking at some examples. For reasons of brevity, I will look at just two: the Schrödinger equation and atomic structures, and free fall phenomena.

Atoms

With the help of the Schrödinger equation, physicists have been able to calculate quite a lot about the structures of atoms and how atoms combine to form simple molecules. A lot of this achievement has, on close examination, the look of Cartwright's image of physics: a motley assortment of models involving idealizations and abstractions of varying degrees of incredibility, chosen in opportunistic ways and often constrained and guided by independent bits of causal knowledge. But not all atomic models have this patchwork character and, in particular, the simplest atom—hydrogen—reveals a quite different picture to us.

Working through the exact solution of Schrödinger's equation for the hydrogen atom was an important milestone in my formation as a fundamentalist. I had never before been, and still was not, happy with quantum mechanics (QM) overall as a candidate fundamental theory; at a minimum, such theories should allow a coherent interpretation, and QM falls down badly on that front. Nevertheless, it offers us a well-defined differential equation and at least clearly says: 'This mathematical law governs the structure of matter.' When you work through the exact solution of the hydrogen atom, you see that in some very important sense, at least, this claim has to be right. The existence of a stable state, in which the proton and electron are bound to each other spatially yet never collapse as one would classically expect (and as one would also expect based on the ascription of their capacities *qua* oppositely charged things), falls out beautifully from the solutions of the equation. More impressively still, perhaps, the energy eigenvalues of the permitted orbitals fall out also, and their differences precisely match the measured emission spectra of hydrogen. And unlike just about every other application of QM and the Schrödinger equation, these calculations can lay claim to being exact rather than approximate, realistic rather than idealized.

What is particularly salient about the hydrogen solution is that its achievements transparently flow from the solution of an equation and from nothing else. You do not arrive at the Leguerre polynomials describing the electron's orbitals by happenstance or by crafting a model using a mix of intuition, antecedent causal knowledge, and so forth. So even though QM is a shambles in many ways and should be replaced as soon as possible by a better theory, if that theory is going to retain QM's ability to account for the atomic structure of hydrogen, it is going to have to give us a mathematical equation structurally isomorphic to the Schrödinger equation as to what governs that structure. From the 1930s onward, our understanding of hydrogen has been and will continue to be based on a mathematical equation. I can see how we might come to view that equation as nonfundamental but rather derivable from some *other* mathematical law or laws. But I can't see how we might come to view the equation as a mere codification of the result of the actions of *capacities* under highly constrained circumstances.

The reason is this: To maintain this stance, we would need to be able to specify what the relevant capacities are, independently, and then show how under such-and-so circumstances their operation makes a certain equation true. For electrons and protons, we can't do the former (other than in a trivial sense) and hence can't get anywhere near doing the latter. What are the capacities carried by electrons and protons? I guess we could say they have (because of the charges they carry) the capacity to attract and repel positively and negatively charged things. We can even quantify this capacity, via Coulomb's law. But this doesn't help explain the stable hydrogen atom; on the contrary, it leads us to expect that electrons and protons should in general collide, not form a stable "orbiting" type situation. We could add that electrons and protons have the capacity to form (relatively) stable neutral combinations, called "atoms" and that sometimes this capacity overrides the attraction/repulsion relationship. But we can only get beyond this triviality and say more by writing down the Schrödinger equation and calculating its solutions. The explanatory primacy of the law over the capacities-talk here is evident.

Returning to the claim of (at least approximate) truth for the Schrödinger equation, what about other atoms, where we can't solve the Schrödinger equation exactly? A fundamentalist will view their complexity as placing a veil between us and the exact operations of the very same mathematical law at work in the hydrogen atom. That is, she will view the patchwork nature of our treatments of more complex atomic and molecular structures as a mere artefact of our cognitive/epistemic limitations and not as evidence that no fundamental laws are really "at work" in the real atoms and molecules. We will have to come back to the question of whether this is cheating later.

Free Fall

The dramatic successes of Newtonian mechanics and gravity theory were, of course, the early font of much fundamentalist belief. In this century Einstein's relativity theories took over most of the domain of phenomena where Newton's physics worked well and added not a few new domains of applicability. Gravitational phenomena are clearly some of the best grounds from which to argue for fundamentalism, because the claim of universality is clearest and most plausible here. Everything that has mass or energy produces gravitation (i.e. affects the curvature of space time), and there is no way to shield any process from gravity. Because gravity is (classically speaking) a relatively weak force, it is hard to test in certain ways "in the lab"—hard, but not impossible. The famous Eötvös experiments and their twentieth century counterparts by Dicke, and the Pound-Rebka experiment, can be considered lab tests of the theory. But by and large the better tests of Newton's or Einstein's theories are carried out by planets and stars. I am referring here to the kinds of tests that seek to distinguish Einstein's from Newton's theory, or Einstein's from Nordtrøm's, and so on. But these are all tests that seek to home in on which candidate fundamental law framework is correct; they are not tests of whether some such laws apply universally—that is just taken for granted.

Nor is it hard to see why it should be so. Terrestrially, much of the everyday phenomena in our lives gives us evidence of the universality of some gravitational law. The fact that everything falls when unsupported, and at the same rate (modulo factors such as air resistance, which we can easily uncover and model if we care to); the fact that the apparent weights of things do not change other than, again, by easily comprehended disturbances such as eating and drinking; these are rather good tests of gravity's universality, at least for all phenomena in our neck of the woods. Of course, it is possible to wonder whether gravity perhaps works quite differently in a different solar system or galaxy. Perhaps the gravitational constant G is actually variable across time or space, though not rapidly enough for us to have detected?[2] No matter—these speculations are just about whether Newton's or Einstein's laws are closer or less close to the *true* universal law or laws. That some law or laws *are* true and govern the gravitational phenomena universally is not called into question.

But do gravity phenomena give us reasons to believe in fundamental laws in the sense we are after, i.e. mathematical equations, rather than (say) some universally carried causal capacities, perhaps the capacity *qua* mass-bearing object to attract other mass-bearing objects? I think it is pretty clear that they do. In the first place, even Newtonian gravity fits awkwardly, at best, into the conceptual framework of cause-effect. The forces acted instantaneously and at a distance, violating most philosophers' intuitions about what cause-effect relationships could be. Then in the twentieth century it was discovered that Newtonian gravity could be translated into a curved-space formalism, analogous to general relativity. This then gives a new ontological picture in which bodies never *do* anything to each other (by gravity) at all! Rather they curve space, or an "affine field". Yet it is unclear what sort of status this affine field should have, whether it should be considered part of space, or as a mere mathematical artifice. What remains clear and unchanged, in a structural sense, are the mathematical laws being encoded and interpreted now one way, later another.

General relativity adds some new twists that complicate a causal reading further. Consider gravitational red-shift: Light travelling up out of a "gravity well" is shifted toward the red end of the spectrum. The effect is much like the more familiar Doppler shift but now linked to gravity or curvature rather than relative motion; it is what the Pound-Rebka experiment verified in the 1950s. We understand the equations giving rise to red shift; a causal interpretation is, in my view, hopeless. Shall we say that space time itself "drags" the photons passing through a region with curvature, slowing them down?[3] Or should we attribute this capacity to the matter that "caused" the gravity well in the first place, even though it is not in contact with the light? Aside from not knowing where the capacity should reside, it is still a misdescription either way. What the theory says is not that anything happens *to* the photons, but rather that they are just moving from a region of space time where time passes more slowly into one where it passes faster. Or rather—this too being a misdescription, as time does not "pass" anywhere—the theory simply gives us mathematical rules for calculating path lengths, time intervals, frequencies, and so on. When we strive for an accurate portrayal of these kind of phenomena, we are forced out of easy, causal metaphors and back onto the equations, the only real account we have of what is going on.

All in all I find gravity theory to be the area of physics where fundamentalism looks most clearly plausible. Cartwright readily concedes this much, though she remains sceptical of the reality of fundamental laws even here.[4] But suppose there is one genuinely true and universal fundamental law of nature—the True Law of Gravity. Can we still happily suppose that most or all of the rest of nature is governed by no universal laws, on a patchwork of laws and causal capacities? Fundamentalist philosophers, at least, will find such a mixed-bag view highly implausible. But we already knew this, I suppose: fundamentalists like their world view tidy and well-ordered.

Above I have discussed just two areas of modern physics that incline me toward belief in true, universal fundamental laws. But similar examples could, I believe, be developed from other successes of quantum theories and general relativity (GR). Everywhere I look, I seem to see such laws in action, producing the wonderful variegation of the blooming and buzzing confusion in which we live on the basis of a few underlying, perfect regularities. But this should sound suspiciously reminiscent. Fundamentalists of the *other* sort, i.e. believers in a certain kind of God, often claimed to see evidence of God's perfection and goodness everywhere they looked. To put it mildly, many of us now incline to a different view on that issue. Perhaps I am deluding myself in just this way about laws of nature. The best way to address this is to now look at Cartwright's arguments against them.

AGAINST FUNDAMENTALISM

Cartwright's arguments against fundamental laws are many sided and have evolved in several ways over the course of the nineteen years since *How the Laws of Physics Lie*. It is not possible to do justice to them in a brief sketch, because their full strength depends on the overall plausibility of the competing metaphysics and methodology of science that she develops to replace the fundamentalist's picture. So the present description will inevitably be somewhat unfair. Hopefully most readers are already familiar with the main arguments and the following remarks can serve more as reminders than as a fair summary.

The main elements of her antifundamentalist arguments can be found in Cartwright (1999: Ch. 2; 2000). Cartwright claims that all the laws in physics ought to be read as *ceteris paribus* laws: They tell us what happens, as long as nothing from outside the domain of the given law interferes. When factors from the outside do occur, they can mess things up quite easily, and the regularity stated in the physical law fails.

> My conclusion from looking at a large number of cases of how theories in physics are used to treat real situations in the world, both in testing the theories and in their impressive technological applications, is that it is always *ceteris paribus* regularities that come into play. All the cases I have looked at have just the characteristic I point to: they are either especially engineered or especially chosen *to include only those causes that occur in the preferred set* of the theory. They are, moreover, always arranged in a very special way: a way that the theory knows how to describe and to predict from. That is not surprising where *ceteris paribus* laws are involved, since we can neither test laws of this kind nor apply them until we are sure the *ceteris paribus* conditions are satisfied. The

> point is that these are the kinds of cases that give us our most powerful reasons for accepting our theories in physics. And the laws they give us reason to accept are all *ceteris paribus* laws. (Cartwright 2000: 210.)

When one gets down to specific examples, I see Cartwright's arguments as falling into two groups. The first I will call the no-forces group; the second, the no-models group. Let's first look at an example from the former group.

Cartwright uses Neurath's example of a thousand-mark banknote falling in a public square as an example of the failure of Newton's second law ($F = ma$). Unlike a compact sphere dropped in a vacuum, whose motions *will* obey the second law (with the law of gravity supplying the force), the banknote will flutter and fly around quite a bit, eventually coming to rest far from where it was dropped. Does this falsify the second law? Of course not, says the fundamentalist: The bill's deviation from a free-fall trajectory is explained by *other forces* on it (the wind and air resistance). But where, asks Cartwright, *in physics* does one get the wind forces from? The answer is: nowhere, because physics tells us practically nothing about wind or how it affects floppy paper objects. To hold that the second law is true in this case, you have to assume on faith that if one back-calculates the forces necessary to produce the motions of the bill correctly, assuming the second law and subtracting the force of gravity, then (a) the forces you calculate really *did* exist, on the bill, as it fluttered around; and (b) those forces are in principle derivable from other fundamental physical laws (QM, perhaps). This is an awfully big thing to take on faith, Cartwright thinks. It's much better to simply allow that the banknote's fall doesn't fall under the second law, because that law's *ceteris paribus* clause is clearly not satisfied. In order to justifiably assert that the second law does apply here, we need more than fundamentalist faith; we need a *good model*, derived in a non-ad hoc manner from the relevant other areas of physics. For the banknote, we don't have one, nor much reason to think we ever can have one.

The no-forces sort of objection thus naturally brings us to the no-models objections. Cartwright doesn't exactly demand that a defender of fundamentalism should be able to come up with a good physical model of something like our banknote fluttering or a cheesecake baking. But if we are to have faith in fundamental laws, at least the theories presenting those laws ought to tell us, in a principled way, how one goes about constructing such a model. But this is what our fundamental theories fail to do. Instead, they typically give us a set of interpretive models that demonstrably obey the relevant laws. Wherever we can force nature to fit the mold of one of these interpretive models, there we can say that the theory applies. But the range of the interpretive models, for our actual fundamental theories, is quite poor.

This is a claim Cartwright has been able to argue with particular force in the realm of quantum mechanics. The fundamental law, Schrödinger's equation, can only be applied to something if one knows the right Hamiltonian function to use. But the theory itself does not give rules for how to construct a Hamiltonian for any given system. The theory does say how to translate the classical Hamiltonian for a (presumably) analogous system into a quantum Hamiltonian; but this rule is by no means enough to cover all intended applications of the theory. So what ends up being the case is that a handful of Hamiltonian functions are known, for a handful of well-defined types of physical situation. Where we have reason to think that a system is structurally *like* one of these models, there we can apply Schrödinger's equation and hence QM. Where none of the handful of models fits, there—in Cartwright's view—QM is silent.

Something similar might, I think, be said for the case of GR. To apply the theory one needs a stress-energy distribution *T* faithful to the system being modelled. But there are really only a handful of such distributions that are mathematically tractable and demonstrably *faithful enough* to the systems—usually stellar or larger in scale—that we wish to model. We have no stress-energy functions that model the wind, much less a wrinkled old banknote fluttering in same. However, the case is perhaps better than that of QM, for two reasons. First, there is a better fit between GR and classical fluid mechanics; generally speaking, we have better guidance about how to move from a classical treatment to a GR treatment. Second and more importantly, GR is not now intended to be viewed as a fundamental theory, by most fundamentalists. It is acknowledged to hold only for large-scale processes and low-enough energies; wherever phenomena seem as though they should fall into the camp of QM, there GR is not expected to hold fully.[5]

The upshot of these observations about the limitations of what we can successfully model with our current theories, for Cartwright, is a strong limitation on what we have a right to induce from their successes.

> This raises one of the most central questions we face in philosophy of science: what should be the bounds on our inductions? . . . I should like to appeal to a crude intuitive principle: when we can recognize a clear boundary within which all our successful cases have been located and, moreover, we can offer good reasons why that boundary might well be relevant, then failing compelling reason to the contrary, we should not extend our inductions beyond that boundary.
>
> For a large number of theories in physics that I have looked at, I think we have such a clear boundary: the empirical successes of the theory are all for cases that fit the theory's interpretive models, or better, that fit some arrangement licensed by the theory of its interpretive models. (Cartwright 2000: 215)

This takes us to one of the central theses of *The Dappled World*: We have reason to think that laws are true where reality matches one of the models in which we know the laws hold; but not elsewhere. Laws are true in bits of reality that match our interpretive models—nomological machines—but not outside of those bounds.

ANSWERING THE ARGUMENTS

Cartwright (Cartwright 2000) sets the core of the dispute out very clearly: What may we induce, from the successes of our physical theories, including those I described earlier? Her answer seems to boil down to this: You can induce that the theories truly describe those systems that have been shown to fit the core interpretive models of the theories, and nothing more.

Notice how dangerously close her answer is to the following: We have reason to think that the laws of a physical theory hold only in those cases where we can show that they hold. But this is not so much a principled restriction on induction, as a flat unwillingness to induce anything at all! Much depends, obviously, on how reasonable and principled the dividing line Cartwright offers really is. A fundamentalist thinks that the range of (approximate) truth of the Schrödinger equation goes quite a bit further than the list of cases where it can be explicitly demonstrated and that this is a reasonable inductive conclusion to draw from the successes of QM. Clearly, we are faced with competing burden-of-proof arguments. What I want to suggest here is that Cartwright's arguments saddle the fundamentalist with unreasonable reductionist demands.

At this point we need to look at a distinction, introduced by Cartwright, between two types of physical reduction: crosswise vs. downward reduction (Cartwright 2000: 207–208). Downward reduction is the familiar reduction of macroscopic processes to the microscopic particles/events composing them. Cartwright claims not to be saddling the fundamentalist with the burden of providing downward reductions. Instead, she asks for successful demonstration of crosswise reductions, meaning: demonstration that the laws holding inside the laboratory also hold outside of it.

The fundamentalist thinks that all of physical nature is governed by some fundamental mathematical law or laws. They are true everywhere and at all times. But obviously, the phenomena these laws allow, which we see all around us, can be of enormous complexity and variety. A fundamentalist thinks that the phenomena studied in chemistry, biology, meteorology, etc. all are composed of the doings of atoms, molecules, photons, fields, and so on, and that these constituents are perfectly governed by the fundamental laws. But she need not believe *any* sort of thesis of the reducibility of biology, chemistry, or meteorology to physics. The lessons we have learned in the past half-century from the failure of various reductionist programmes are many, but they do not include a lesson to the effect that there are no fundamental laws of nature.

Yet it seems that in order to answer Cartwright's objections in the way she desires, the fundamentalist would have to deliver a successful reduction of all the sciences (and much that is not overtly covered by any science) to fundamental physical theories. Suppose we discuss the gasoline-oxygen explosions in my engine's cylinders. In line with what she says about the banknote, I suppose Cartwright would not want to admit that the Schrödinger equation holds inside the cylinder, without being given an appropriate Hamiltonian for this kind of system, and the calculations to show that an adequate model within the theory is available. But this is to demand either theory-theory or type-type reductionism of a very strong sort—downward reduction. I suspect most fundamentalists have no wish to argue that such a reduction is possible, for us at least.

This means, then, that given the way Cartwright draws her principled boundary on inductions, we can never say we have good grounds for believing fundamental laws to hold everywhere unless we can provide the explicit reductions to prove it. We may call these reductions crosswise if we wish, but they will in general have to be downward also. This is, I submit, an unreasonably strong requirement. Cartwright's principled boundary on inductions does make sense if we start by assuming the correctness of her patchwork ontology of capacities without fundamental laws. But equally, the fundamentalist's induction of the holding of laws such as the Schrödinger equation outside the laboratory setting makes sense, if we start with the assumption that nature is fundamentally governed by mathematical regularities, with causality being a mere imperfect, anthropomorphic (though often very useful) conceptual tool.

A WORLD OF SIMPLE BUILDING BLOCKS

To end, I want to discuss two final issues: the simplicity argument for believing that laws hold outside our models as well as inside, and the vexed problem that all the fundamental-type laws we have been able to conceive to date are known to be false, perhaps even badly false (for the kinds of reasons fundamentalists themselves give, not the kind highlighted by Cartwright).

The primary argument for fundamentalism, not yet mentioned, is this: we all believe, with very good reason, that things in the physical world are all composed of a few basic types of particles: electrons, protons, neutrons, and photons, mostly, along with a tiny amount of more esoteric particle kinds.[6] We know that these tiny things are puzzling in various ways, and they cannot be thought of as Newtonian-style billiard balls moving on smooth trajectories under the influence of purely local force fields. Nevertheless, they are here to stay. Whatever radical changes future physics may bring, it is not really conceivable that, á la phlogiston, these entities will vanish without a trace and come to be seen as embarrassing errors with no correlate or counterparts in the True Physics. Moreover we know a good bit about how these things behave in certain settings. A big part of this knowledge is given by QM and is connected with the Schrödinger equation. Where we are clever enough to be *able* to test this theory and this equation, they seem to be correct. But—aside from this question of what we are clever enough to be able to model and treat with a theory—there seems to be no very relevant difference between matter inside the labs and matter outside the labs. A hydrogen atom in a spectrometer is, plausibly, much the same as a hydrogen atom floating in your living room. The simplest hypothesis would seem to be that if there are mathematical laws governing these things in one setting, then the same laws govern them everywhere.

The sentence above is precisely where Cartwright would say I have gone astray. (Or she might agree with the simplicity claim but deny that that has any epistemic force.) Her view is that these successful tests show only that certain kinds of systems, which can be modelled in such a way as to let us deploy our well-understood models, obey mathematical laws. They may be outside the laboratory as well as inside, but most of what goes on outside cannot be so modelled. Instead, she proposes, an equally good hypothesis is this: the mathematical laws manage to capture the effects of the operation of real capacities in nature under certain restricted conditions; we may induce the existence of the same capacities outside the laboratory, but not the truth of the mathematical laws.

We are back almost to square one: How can the fundamentalist argue that the tests and successes show more, especially when she (in all likelihood) accepts that QM is not even a fully interpretable theory, much less a part of the True Final Physics? For it has to be acknowledged that the failure of QM to be demonstrably valid everywhere is not merely a matter of calculational complexity and a lack of cleverness on our part. It is also a product of two further factors. The first, stressed by Cartwright, is that QM provides only incomplete model-building prescriptions—in particular, it has no rules for constructing the right Hamiltonian for any *arbitrary* system.[7] The second, related, reason is equally important: QM, wonderful though it is, is not The Truth, not a part of the True Final Physics, but only an approximation of some kind to the latter for certain domains. And the same goes for other theories such as GR or QFT, even though they may be nicer in some ways than QM.

I suspect that the Schrödinger equation does "govern", in the appropriate sense, quite a lot of what goes on outside of labs and superconductors and other well-regimented environments. But as a fundamentalist who is convinced that QM is ultimately a *false* theory that merely gets close to the truth in certain ways and certain domains, I do not have to argue at length over this question. And here we have arrived at perhaps the most important reason why fundamentalists feel they can resist Cartwright's patchwork of laws. The ultimate set of mathematical laws that a fundamentalist believes in is meant to be unified, consistent, coherent, and of clear applicability to any real situation. Unlike GR, it should *not* say patently false things about matter (GR says it is a continuous fluid); unlike QM, it should not use an unprincipled mix of concepts from earlier theories and uninterpretable new mathematical objects. The ultimate laws will be true in supernovae and in teacups, lasers, and banknotes. We won't be able to prove this case-by-case nor reduce molecular biology or chemistry to fundamental physics: I repeat, no reductionism need be possible. But we should have much better grounds for thinking that our inductions can proceed beyond the bounds of our nomological machines than we presently do for incomplete and false theories such as GR and QM. It is *these* laws that the fundamentalist believes in, not the half-way houses we have managed to construct to date.

It may seem as though my defence of fundamentalism has in the end collapsed back into an expression of blind faith, as I argue not for the literal truth of anything we currently call "fundamental physical laws" but, rather, for their ideal future replacements.[8] Not so. For even though we don't have this physics in hand, or even on the horizon, we may still have evidence that such a thing exists. Let me recount the components of the answer to Cartwright's antifundamentalist arguments.

- The simplicity argument is surely onto something relevant and important. A hydrogen atom is a hydrogen atom, whether in an interferometer or a dirigible; if its behaviour is governed by mathematical laws in one setting, there is *prima facie* reason to expect it is so governed in the other.
- One cannot simply insist that inductions should stop at the boundaries of what has already been successfully modelled, for this is tantamount to claiming that fundamentalism can only be vindicated by the demonstrated achievement of a very strong reductionism, much stronger than what any fundamentalist should (or, I suspect, does) currently believe possible.
- If we accept our starting point above, namely that there is a need to explain such widespread and reliable regularities in nature as we have been able to uncover, both in daily life and in science, then we seem to face a choice between the fundamentalist's picture or Cartwright's patchwork, capacity-based picture. This brings us back to what I tried to stress in connection with the calculation of the hydrogen atom's structure. Many of the scientific and technological successes of physics can be adequately described in the language of stable capacities and Aristotelian natures. But quite a lot of it cannot or can only be done very awkwardly. This speaks in favour of the idea, widely accepted since the eighteenth century at least, that the ultimate explanations of nature's many regularities will be couched in mathematical language, not the language of cause and effect, tendencies and propensities, strivings and so forth. We know from many examples how phenomena at first describable only imprecisely using causal talk can be given a deeper account by bringing them under mathematical laws (examples: reflection and refraction of visible light; attraction and repulsion

between charged macrobodies). We don't, I think, have good examples that go in the opposite direction.

To summarize: we have reasons to believe in the truth—in certain settings—of highly precise and *mathematical* (but not *causal*) laws. Because nature is mostly composed, everywhere, of the same kinds of things, we have reason to induce that these laws hold in much of the world outside our test situations. To demand an explicit demonstration, for settings such as the cylinder of an auto engine, is to demand unfairly a strong reductionism. We have of course many reasons for thinking the laws we have concocted to date are not perfectly true nor genuinely fundamental. We understand from many examples in the history of physics how it could be that these laws get supplanted later by more universal and fundamental laws—if they are *mathematical* laws of the kind the fundamentalist seeks. But we have little reason for confidence that our understanding of things can be deepened by moving away from fundamental mathematical laws and to a patchwork ontology of false-but-useful laws approximating a reality of capacities having no true *general* description (in mathematical or ordinary language).

Cartwright's patchwork of laws and capacities offers us a picture of science and its possibilities that is very faithful to the current state of theory and practice. That is its weakness: It holds out no reason to think that our deepest explanations can get significantly better (though at least our engineering can). The fundamentalists' view does however aim at significantly deeper and better explanations at a fundamental level—even though they may not help us with our engineering. To engineers and experimentalists, I commend Cartwright's philosophy of science wholeheartedly. But I hope to have made space for theoreticians and philosophers of physics to keep their faith in a world with fundamental physical laws.

NOTES

1. It takes courage for a philosopher to challenge directly the entrenched power structure, dominated by physicists, in this way. Andersen (2001) illustrates nicely that this is so.
2. It bears conceding here that the models of the universe on a large-scale that most astrophysicists now believe in, though still fundamentalist through and through, are wildly speculative rather than well-confirmed, and if taken seriously involve a modification of Einstein's equations—though such a possibility has been included in standard textbook treatments for decades.
3. The slowing down referred to here is in terms of the frequency of the waves, not the locally measurable velocity *c*.
4. In conversation.
5. This discussion raises the important question of the overlaps or intersections of candidate fundamental laws. This is remarked on briefly in the final section. Here let me note that Cartwright's banknote is an excellent example of such intersection: The actions of air molecules on the bill really ought to be in the domain of a quantum theory (as well as the internal structure of the bill itself),

and GR really should provide either the force of gravity (on each part??), or perhaps the space-time setting in which the fall process occurs.

[6] From this list of reasonably well-established characters I would leave out a number of much more dubious types such as: the space-time manifold, Higgs bosons, quintessence, virtual particles . . . even though they play quite important roles in some current theories. Some fundamentalists, more credulous than I, might point out that (a) neutrinos far outnumber the particles I've mentioned, and (b) so-called "dark matter" allegedly outmasses them as well. My point is just that the p, n°, e^{-}s and photons make up most of the world of our everyday experience.

[7] What is actually reasonable to demand is not a set of rules to take you from a macrosystem to a microspecific model but, rather, complete rules for how to "build up" from atom-sized systems, gradually adding more and more particles, until macrolevel phenomena are achieved. Having such rules wouldn't necessarily tell you much at all about how to construct a complete model of an engine's cylinder.

[8] In a recent paper, Sklar (2003) defends fundamentalism against Cartwright too. But he tries to do so not merely for the much-desired future theories that replace and unify GTR and quantum theory, but rather defending the near-truth in all domains of current QFT and QCD. By so doing, he opens himself up to some important objections from Teller (2004). Teller defends an ontologically dappled world but explicitly exempts his argument from applying against a hypothesized future, "perfect" fundamental physics. I don't believe one can or should discuss fundamentalism in isolation from its ideal goals. If theoretical physics were really (somehow) finished already, with nothing better to come in the future than what we already have, then I would have to concede that Cartwright's view is the more accurate.

REFERENCES

Anderson, P. W. (2001) 'Review of *The Dappled World*', *Studies in History and Philosophy of Modern Physics*, 32: 487–494.

Cartwright, N. (1999) *The Dappled World: A Study of the Boundaries of Science*, Cambridge: Cambridge University Press.

———. (2000) 'Against the completability of science', in M. W. F. Stone and J. Wolff (eds) *The Proper Ambition of Science*, London: Routledge.

Russell, B. (1912) 'On the notion of cause', in *Mysticism and Logic*, London: Allen & Unwin.

Sklar, L. (2003) 'Dappled theories in a uniform world', *Philosophy of Science*, 70: 424–441.

Teller, P. (2004) 'How we dapple the world', *Philosophy of Science*, 71: 425–447.

Reply to Carl Hoefer

My differences with Carl Hoefer are about how to understand the "fundamental" equations of modern physics. He says they are not to be read as statements of capacity or with a *ceteris paribus* restriction. I claim they must if they are to be plausibly taken as true (or at least true for the nonce). Much of the dispute depends on the scope of inductions. I take it as a very good rule of thumb that smaller inductions are better warranted than larger and, for reasons I rehearse in my discussion of Suárez in this volume, I am especially suspicious of inductions to the fundamental equations of contemporary physics when they are read as Hoefer desires.

Hoefer suggests that my reluctance to induce farther than necessary reduces us to the position that 'We have reason to think that the laws of physical theory hold only in the cases where we can show that they hold'. Yes and no. I do maintain that claims have the most warrant when they have been shown to hold and less elsewhere. But, to adopt Hoefer's language, there is also a "principled" stopping point well beyond this, but far short of his own universal scope. Roughly, so long as all relevant features can be correctly described by the concepts in the theory, then the theory holds. Notice that this is not at all "inside the laboratory" versus "outside" as Hoefer often says, though it is of course inside the laboratory where we have our best shot at ensuring that all the relevant features are described by the theory.

Again Hoefer and I would disagree about how constricting this is. He tells us that hydrogen atoms behave the same within the laboratory and without. But he does not really mean this because they are subject to different influences in different places. Hoefer is right to say that my arguments suppose some kind of reductionism: If Schrödinger's equation is to be true of a hydrogen atom in any setting, the relevant features of that setting must be represented by terms in the quantum Hamiltonian. I am very sceptical that they can be. Hoefer offers a "simple hypothesis": If a law governs things in one setting it will do so in all. But this is not a simple hypothesis. It supposes that because some relevant features can be represented within the concepts of a theory—concepts that rightly have very strong strictures on their rules for application—all can be.

In considering Hoefer's hypothesis it is important to remember two things. First, physics has had trouble not only in finding causes that can be represented as theory demands, but also in finding effects. Each theory deals with a highly selective set of very unnatural effects—the second derivative of distance with respect to time, the quantum state function, the electromagnetic field strength, and so on—and each introduces a very special set of concepts that matter to these effects. The trick has been to adjust simultaneously the effects studied and the concepts used till a kind of closure is achieved. We can find equations that predict what happens when all the features relevant to the selected effects can be represented by the designated concepts. This is a considerable achievement, and the strategy has been much envied by a variety of social scientists who, rightly or wrongly, do not feel at liberty to pick and choose their effects.

The second is that the concepts of physics have very strict rules of application, which is what provides physics its impressive predictive powers compared with more ad hoc mathematical sciences such as economics. This does not mean just that across a wide range of subdisciplines physics concepts have highly precise measurement procedures; to the contrary, many do not. But they *are* applied through specific interpretational principles—bridge principles. Hoefer admits that quantum mechanics is applied 'via classical mechanics'. What is so important about that is not that it provides meaning where there was none but, rather, that it ties these concepts to a vast network of knowledge of what must be the case if they apply. We do not casually apply the label "harmonic oscillator"; there are by now volumes of details about what that representation implies.

But this rich interlocking network of detailed constraints is a two-edged sword.

It provides physics with great predictive strength, but it can also constrain its range. Concepts that have strict rules for their application may well not apply very widely. Of course, as Pythagoreans think, Nature may be made through and through for just concepts such as these. But maybe it isn't, and we are lucky that physics can work—and work wonderfully—where it does work.

14 Cartwright on Wholism

Michael Esfeld

WHOLISM AS A METAPHYSICAL BACKGROUND FOR ANTIFUNDAMENTALISM

Nancy Cartwright is famous for rejecting fundamentalism in the sense of the view that the laws of nature reduce in principle to the laws of one fundamental physical theory or supervene on a set of fundamental laws. Instead, there is a patchwork of laws, that is, several groups of laws that are not related to each other in a systematic or uniform way (Cartwright 1999). By a law of nature, Cartwright means a necessary regular association between properties (Cartwright 1999: 4, 49). Consequently, she claims that our description of nature cannot even in principle be reduced to one fundamental theory. There is an irreducible plurality of different theories, each of which has its own limited area of application. There is no systematic relation between these theories.

Note that what Cartwright calls fundamentalism is a more general position than what is known as fundamentalism in the epistemological debate about fundamentalism versus wholism. A fundamentalist in Cartwright's sense need not subscribe to the claim that there is one foundation of knowledge—such as the sense data of the classical empiricists, the Cartesian *cogito*, or the Kantian transcendental unity of apperception—which is unshakeable in the sense that it is the point where justification comes to an end, justifying itself. A fundamentalist in Cartwright's sense is anyone who holds that there is a basic description of the world to which all other true descriptions can in principle be reduced. What is more, according to Cartwright, to be a fundamentalist, it is sufficient to endorse global supervenience—the view that there is a basic level of the world on which everything else supervenes (Cartwright 1999: 32–33).

Cartwright does not regard her antifundamentalism as a version of antirealism *tout court*. She takes herself to be a local realist, that is, a realist about a wide variety of phenomenological laws that each have a limited domain of application (Cartwright 1999: 23). To avoid being classified as an antirealist, however, Cartwright has to do more. On the one hand, she has to say something about the relationship between these domains. On the other hand, in order to steer clear of fundamentalism, what holds these domains together cannot be anything like universal laws. As an antidote to both antirealism and fundamentalism, she contemplates a metaphysics of wholism in the following sense: Nature is one interacting whole. Different theories carve out different aspects of this underlying whole. The whole cannot be adequately described by any one of these theories or any combination of these theories.[1] This metaphysics of wholism is attractive in the context of Cartwright's position, because it seems to explain both why there is a patchwork of laws and how the different domains that these laws describe hang together.

The aim of this chapter is to inquire whether Cartwright's antifundamentalism might be supported by a metaphysics of wholism, independently of whether or not Cartwright is in fact prepared to endorse such a metaphysics. In this section, I consider Cartwright's wholism and its relationship to her antifundamentalism as well as her metaphysics of capacities. The following two sections compare the wholism that Cartwright contemplates with those areas in contemporary philosophy in which some sort of wholism is widespread: the interpretation of quantum physics on the one hand and semantics on the other. The aim is to establish whether or not a metaphysics of wholism can lend support to antifundamentalism.

In her paper, 'Can wholism reconcile the inaccuracy of theory with the accuracy of prediction?' (1991), Cartwright sets out her aim as follows:

> I begin with the observation that the laws of physics are true only of what we make. . . . We do not measure the success of modern physics by its ability to explain the material world around us as it naturally comes but, rather, by its ability to create very particular environments in which nature behaves in highly regular and precisely predictable ways, as in a laboratory, where we put our theories to the test, or in new pieces of technology, where we put them to use, as with the hydrogen bomb or the laser.
>
> This observation bears on recent debates about scientific realism. A kind of anti-realism is an easy next step. If the laws of physics are true just of what we make, then there is a sense in which we make the laws of physics true. In this paper I want to give a model about how this might be the case. The model begins with the assumption that nature is far more wholistic than we imagine; using the model I then try to explain why our atomistic descriptions can be deployed to produce such precisely accurate results. (Cartwright 1991: 3)

Cartwright thus starts from her patchwork view: What we take to be laws of nature has only a very limited application, that is, an application confined to situations that are like those ones which we create in a laboratory. Nevertheless, she has to explain the success of science, which constitutes the main argument for a fully fledged scientific realism. According to what is known as the no miracles argument for scientific realism, the success of our scientific theories in the prediction of novel phenomena would be a miracle if our theories were not approximately true, that is, if they were not accepted as giving an approximately true description of what the world is really like (Putnam 1975: 73).

According to Cartwright, modern science proceeds by induction: on the basis of descriptions that prove to be applicable to the situations that we ourselves create, we make an induction to general laws of nature. Cartwright maintains that this induction leads to a wrong result. She offers wholism as a model of how it can be possible that, although our theories are wrong from a perspective larger than the one of the situations we ourselves create, they have worked so far with amazing success. She says:

> The most immediate way to parley a wholistic intuition into a model for anti-realism is to imagine cases in which the variations that we find salient are determined within a far larger context; yet, like the chicken [Russell's example of a chicken whose induction works for a time, but is fatally wrong], we encounter them only during an epoch in which the relevant background remains relatively stable.
>
> (Cartwright 1991: 7)

To elaborate on this wholistic model, Cartwright takes an example from economics:

> Haavelmo . . . imagined the economy to be governed by a set of linear, simultaneous equations, containing random shock terms. . . . They do describe separate and stable mechanisms which can be manipulated and deployed to produce predictable economic consequences. (That's the hope at least.) But the existence and stability of these mechanisms is an epiphenomenon of the entire economic and social context. They can, so to speak, be "carved out" of this whole; but what we carve out need not be there to begin with. . . . The punch line is of course that the fundamental laws of physics may not be so fundamental either. Just as Haavelmo hoped conceptually to carve out separate mechanisms from an underlying interacting whole, physics carves them out physically. By choice and arrangement of materials and either by intensive shielding or heavy over-determination, we create special environments which hold fixed the principle effective parts. We may in this way arrive at very precise and reliable regularities without in any way grasping the true form of what is going on.
>
> (Cartwright 1991: 8–9)

Hence the idea is that nature as a whole is far more complex than we might imagine. Our theories carve out certain aspects of nature, and we create the environments that are necessary for our predictions to work, that is, environments that keep certain factors stable and eliminate other disturbing factors. However, we cannot develop a true fundamental theory of nature as a whole. This wholistic metaphysics has a Kantian ring: We know only the way in which nature appears to us—that is, the various aspects which our theories describe and which we create in our scientific and technological activities; but we cannot know what nature is like in itself—that is, we are ignorant of nature as a whole. This is a principled ignorance, as Cartwright's point is that there can be no true or approximately true fundamental theory that applies to nature as a whole.

How does this wholistic metaphysics relate to Cartwright's theory of capacities? According to Cartwright, capacities are more basic than laws. A capacity is more general than a disposition: It is not tied to any single kind of manifestation. In other words, capacities are determinable, whereas dispositions are determinate (Cartwright 1999: 64). She says:

> It is capacities that are basic, and laws of nature obtain—to the extent that they do obtain—on account of the capacities; or more explicitly, on account of the repeated operation of a system with stable capacities in particularly fortunate circumstances. Sometimes the arrangement of the components and the setting are appropriate for a law to occur naturally, as in the planetary system; more often they are engineered by us, as in a laboratory experiment. But in any case, it takes what I call a *nomological machine* to get a law of nature. (Cartwright 1999: 49)

Capacities 'can be assembled and reassembled in different nomological machines, unending in their variety, to give rise to different laws' (Cartwright 1999: 52. See 1999: Ch. 3; 1989: Ch. 5).

According to a position that is widespread in philosophy of science, capacities, powers, or dispositions require something that has the capacities, powers, or dispositions in question. That something cannot consist solely of capacities and the like; over and above that, it has to have some intrinsic properties or other that are in some sense a basis for its capacities (although it is not necessary that the capacities supervene on the intrinsic properties). This reasoning also leads to a sort of Kantian metaphysics of nature: Scientific inquiry can only reveal the capacities of the things in nature but not their intrinsic properties. We can thus only know the way in which things appear to other things, including ourselves, by manifesting certain capacities but not what they are like in themselves. We have no access to their intrinsic properties (Foster 1982: Ch. 4, appendix; Jackson 1998: 23–24; Langton 1998).

This is also a metaphysics that one might contemplate employing as a basis for anti-fundamentalism (although none of the authors just referred to intends to receive this metaphysics in that way): We know only the manifestations of capacities; we do not know what it is in the things themselves in virtue of which they have various capacities. Cartwright, however, rejects such a metaphysics. According to her, it makes no sense to draw a distinction between—intrinsic—properties and powers:

> As we represent the world, objects have properties, and by virtue of having properties they are empowered to do things, in particular to change facts about properties in other objects, including facts about what we perceive and what we experience. . . . Thus the question of whether every dispositional property is grounded in an occurrent property makes no sense. There just are properties and all properties bring powers with them.
>
> (Cartwright 1997: 74; 1999: 73)

Cartwright is prepared to endorse the view of (Shoemaker 1984: Ch. 10), according to which properties consist in their causal powers (Cartwright 1999: 70). Consequently, she is not committed to intrinsic properties: All properties may in the last resort turn out to be relational.

The position that dispositions or capacities are grounded on intrinsic properties gives rise to an atomistic metaphysics: The essence of things is their intrinsic properties. Intrinsic are all and only those qualitative properties that a thing has irrespective of whether or not there are other contingent things. That is to say, having or lacking an intrinsic property is independent of accompaniment by other things or loneliness (Langton & Lewis 1998; Lewis 2001). Consequently, things are held together not in virtue of what they are in themselves, but in virtue of the relations they enter into.

Cartwright's reservations about such a position link up with her wholism: If there is no distinction between properties and powers, then things are connected by their very nature; manifesting their capacities in causal interaction is their essence. Cartwright's metaphysics of capacities can thus be combined with a view of nature as being one interacting whole instead of there being unknowable intrinsic properties of individual things on which their dispositions are grounded. In other words, a metaphysics of wholism can be employed in order to counter the argument that capacities presuppose intrinsic properties. Our ignorance concerns the fact that we cannot know nature as a whole, but only various aspects of this whole.

A WHOLISTIC MODEL FROM QUANTUM PHYSICS

One would like to know more about this wholism than just learning that nature is one interacting whole and that our theories carve out different aspects of this whole, thereby simplifying its real complexity. When it comes to wholism with respect to the domain of physics, there is one physical theory that is often received as revealing some sort of a wholistic feature of nature, namely quantum theory. The purpose of this section therefore is to sketch a model for a wholistic metaphysics of nature based on quantum theory. The aim is to spell out wholism as precisely as possible on the basis of our current knowledge in order to see whether a wholistic metaphysics really supports Cartwright's antifundamentalism.

Quantum systems often have to be described as being in what is known as entangled states: There is no description available that attributes to each of the quantum systems in question a well-defined state each. Instead, only the whole of these quantum systems taken together is represented as being in a well-defined state (that is, a pure state). That state of the whole includes correlations between the conditional probability distributions of properties of its parts. These correlations are known as EPR-correlations following a famous paper by (Einstein et al. 1935). These correlations are independent of any spatiotemporal distance between the parts of the whole in question. They are well confirmed by experiments, notably experiments that carry out measurements on two quantum systems with entangled states at a space-like distance: The setting of the parameter to be measured and the measurement on the one side of the arrangement are separated by a space-like distance from the setting of the parameter to be measured and the measurement on the other side.[2]

The dynamics of quantum systems is described by the Schrödinger equation (or a relativistic generalization of this equation). According to the Schrödinger dynamics, interaction leads to ever more entanglement. Consequently, in the end, if we assume that there is direct or indirect interaction between any two quantum systems, we get to a view of ubiquitous entanglement. Even if we do not take interaction into account, starting from the formalism of quantum theory, it is to be expected that whenever we consider a whole that has two or more quantum systems as its parts, the states of these systems are entangled (Scheibe 1991: 228). Hence, if we imagine the state of all quantum systems taken together, this will be an entangled state. On the basis of considerations such as the mentioned ones, a number of philosophers of science interpret quantum theory in terms of wholism.[3] One can thus build a model of nature being one wholistic system on quantum theory.

How does this wholism relate to experience? If we employ the Schrödinger dynamics to describe a situation of measurement, we have to conclude that the state of the quantum system becomes entangled with the state of the measuring apparatus (instead of system and apparatus being in separate states). This is the source of the notorious measurement problem in the interpretation of quantum theory. Fortunately, ubiquitous entanglement is not the end of the story. Decoherence shows how the appearance of classical properties and states can arise within a world of quantum entanglement.[4] Note that as long as only decoherence is in the play, there is no question of nonlocality in the sense of one event being causally relevant to another event at a space-like distance, as no reduction of entanglement to separate states occurs.

However, decoherence can at most account for why there appear to be classical properties and states, because decoherence does not include the notion of a state reduction, that is, a dissolution of entanglement. Decoherence does not enable us to understand the existence of classical properties and states (if they exist). To put the matter in more technical terms, decoherence refers to an improper mixture (entangled states) that cannot operationally be distinguished from a proper mixture (systems with separate states each).[5] It is therefore in dispute whether the reference to decoherence is sufficient to cope with the measurement problem or whether, in addition to admitting decoherence, a change to the Schrödinger dynamics—such as the one proposed by (Ghirardi et al. 1986)—is called for. Furthermore, it is in dispute whether, if one commits oneself to a metaphysics of quantum entanglement without countenancing a change to the Schrödinger dynamics, decoherence is sufficient to account for our impression that there is a classical world or whether controversial additional ontological commitments have to be endorsed (such as, e.g., the commitment to many superposed experiences as in the many minds interpretation; Albert & Loewer 1988; Lockwood 1989: Ch. 12–13). These issues are not relevant here. The point is that, owing to decoherence, there is a basis for understanding how the appearance of a classical world to observers can in principle be integrated into a model of a quantum domain of ubiquitous entanglement.

There are a number of differences between this model of a wholistic quantum world and Cartwright's wholism in connection with a metaphysics of capacities:

1. The reason the quantum whole is a wholistic system is not that it is an interacting whole. Although interaction leads to entanglement, entanglement is not a sort of interaction; it is not a causal relation. Insofar as there is a causal relation between two or more systems, it is presupposed that these systems each have a well-defined state. If this were not the case, a causal dependence between changes in state-dependent properties of each of the systems in question could not be formulated. Insofar as quantum systems are subject to entanglement, by contrast, they do not each have well-defined states.
2. It can be argued that causal relations, powers, and dispositions or capacities in general require intrinsic properties on which they are in some sense grounded, although Cartwright, for one, does not accept such an argument. In any case, it seems that this type of argument cannot be applied to the quantum relations of entanglement, as they are not causal relations. Referring to the nonseparability of quantum systems and the issue of whether or not quantum systems are individuals in particular, one can maintain that all there is to quantum systems

insofar as they are subject to entanglement are the relations in which they stand, intrinsic properties that could in some sense be a basis for these relations being excluded.[6]

3. The sketched position of quantum wholism is committed to the concept of a quantum state of the world (or of nature as a whole). Of course, no one will ever be able to write down that state. Nonetheless, there is nothing here whose nature is in principle unknowable. This metaphysics of quantum wholism does not admit of the Kantian distinction between the world as it appears to us and as it is in itself, insofar as the latter is in principle inaccessible. We know what the world is in itself: namely an unimaginably complex network of quantum relations of entanglement. And it seems that we can know in principle how this way the world is in itself is connected with the way in which the world appears to us, decoherence being the clue.
4. Quantum wholism is in no sense a basis on which a claim to the effect that the world is dappled can be built, because quantum wholism encompasses the world as a whole at the quantum level. Furthermore, if there is a path from the quantum domain to the classical domain via decoherence, then it is shown that the metaphysical position of quantum wholism leads to the epistemological position that there is a systematic relation between the various theories of the natural sciences and a fundamental physical theory of the quantum realm. It seems that one can say at least that the phenomena described by other mature scientific theories supervene on the quantum domain taken as a whole. It can therefore be claimed that quantum theory, interpreted in terms of wholism, is a fundamental theory and perhaps even a universal physical theory. It may not be possible to reduce other theories of the natural sciences to quantum theory; but reduction is not necessary in order to show that there is a systematic connection between the theories of the natural sciences. This is not to say that quantum theory, interpreted along the lines of the sketched model of quantum wholism, is the final truth of the matter. There is no question of metaphysical realism here. Quantum theory may tomorrow be superseded by another physical theory and the entire case for wholism break down. On the other hand, there is as yet no experimental evidence that disconfirms quantum theory.

Cartwright is, of course, aware of the discussion on quantum wholism. Given her position, she warns us not to succumb to what she calls the quantum takeover. In her view, quantum theory is no universal theory. There are both quantum and classical states; one and the same system can be in both at the same time without contradiction. There is no general formula describing how quantum properties relate to classical properties (Cartwright 1999: Ch. 9). For Cartwright, thus, quantum theory is a theory like all the other theories of the natural sciences: It has a limited domain of application, and there is no systematic relation to other theories; the laws of quantum physics are nothing but one piece in the patchwork of natural laws. As regards the EPR-correlations, Cartwright favours a particular causal account of the correlations that are measured in the experiments. She conceives that causal account as an alternative to quantum wholism (Cartwright 1989: Ch. 6; Chang & Cartwright 1993; Cartwright & Suárez: forthcoming).[7]

This is not the place to dwell on the interpretation of quantum theory and to consider the merits and demerits of quantum wholism versus a causal account of the EPR-experiments that avoids a commitment to quantum wholism. The conclusion of this section can be summed up as follows: (1) if one contemplates a metaphysics of nature in terms of wholism, one should be able to spell out that metaphysics. (2) The only physical basis for a wide-ranging and substantial wholism of nature stems from quantum physics. (3) If one works out quantum wholism, one realizes that a metaphysics of nature built on quantum wholism can in no way serve as a basis for an epistemology of a patchwork of laws of nature. One may envisage generalizing this point and tentatively claim the following: As soon as the idea of a wholistic metaphysics of nature is spelled out, that spelling out results in a fundamental theory that (a) describes nature as one wholistic system at a certain level and that (b) thereby seems to describe something which can serve at least as a supervenience basis for the claims made by other theories instead of lending support to a patchwork view of scientific theories. To establish such a claim, more case studies would be necessary. Nonetheless, wholism does not seem to be the appropriate candidate for a metaphysics of nature on which a patchwork view of science can be grounded. To put it in a nutshell, if the world is wholistic, it is unitary rather than dappled.

CONFIRMATION WHOLISM AND SEMANTIC WHOLISM

Apart from quantum physics and the metaphysics of nature, there is another area where a sort of wholism is widespread: the theory of the meaning, confirmation, and justification of our beliefs. Perhaps the most prominent source of this wholism is Quine's seminal paper 'Two dogmas of empiricism' (Quine 1953). Quine claims that only a whole theory and in the last resort, only the whole of our knowledge can be confirmed or disconfirmed by experience. This is confirmation wholism. Furthermore, single statements have meaning only insofar as they are integrated into a whole theory and in the last resort into the whole of our knowledge. This is semantic wholism.

In later papers, Quine qualifies the claims made in 'Two dogmas': It is not the whole of our knowledge at once that is confronted with experience, but only a cluster of statements. A cluster of statements is also sufficient for meaning. Nevertheless, Quine maintains that such a cluster finally encompasses the whole of our knowledge: There is no strict partition within our knowledge. For any two parts of our knowledge, there are circumstances imaginable in which these parts may become relevant to each other as regards confirmation and/or meaning (Quine 1975: 313–315; 1986: 619; 1991: 268–269). One of Quine's central examples is that it may turn out to be reasonable to abrogate the logical law of the excluded middle consequent upon the results of experiments in quantum physics (so that this law is in effect not a logical but an empirical one; Quine 1953: 43; 1986: 620; 1991: 268–269).

Quine's confirmation wholism is widely accepted in contemporary epistemology and philosophy of science. Some sort of semantic wholism also is widespread in the theory of meaning. Today's most popular version of semantic wholism is inferential role semantics: The meaning of a predicate consists in the inferences that a statement in which the predicate in question is employed licenses. Nonetheless, semantic wholism and conformation wholism are two distinct positions. In particular, one can be a confirmation wholist without being a semantic wholist (Fodor & Lepore 1992: Ch. 2).

Semantic wholism and confirmation wholism are both opposed to a patchwork view of the natural sciences. As regards confirmation wholism, the relations between the different domains may not always be evident in normal science, but they become manifest in a situation that calls for changes, as Quine's example of quantum logics shows. As regards semantic wholism, when confronted with the objection that science is split up in many compartments, Quine replies by referring to the logical and mathematical components that are common to all scientific theories (Quine 1975: 314; 1986: 620). Quine sees logic and mathematics as being able to guarantee a minimal unity of science, because he does not regard logical and mathematical statements as having a meaning in separation from empirical statements (Quine 1991: 269). Whatever the exact status of logical and mathematical statements may be, if the idea of a strict separation between analytic and synthetic statements is rejected and an inferential semantics accepted, there seems to be no principled limit to the inferential relations that contribute to the meaning of a given statement. In other words, there are no isolated patches of knowledge.

The widespread acceptance of some sort of wholism in confirmation theory as well as in semantics illustrates that a theory of confirmation and a theory of meaning is indispensable in order to make the claim of a patchwork of natural laws even intelligible. We need a theory of confirmation that shows how confirmation can be limited to a particular theory (or a particular patch for that matter). And we need a theory of meaning that is an alternative to an inferential role semantics for scientific concepts. If this theory is to be atomistic, it has to tell a story as to how scientific concepts can get their meaning one by one. If this theory is to be a localism in the sense that the concepts of each theory are interdependent, but there are various and many theories, then we need a principle that is capable of keeping the inferences that are constitutive of the meaning of a concept within the boundaries of one theory. In any case, given that logical and mathematical principles are pervasive in scientific theories, it seems that we need a strict distinction between analytic and synthetic predicates or statements in order to keep the meaning of logical and mathematical predicates apart from the meaning of empirical predicates. Cartwright would have to provide us with such a theory of meaning and confirmation in order to support her antifundamentalism.

Furthermore, Cartwright's metaphor of different theories carving out different aspects of an underlying whole needs clarification: How are the different theories related to one another? Can they be translated into one another? If so, why does the translatability of the concepts not contribute to their meaning? If not, there seems to be no communication possible across different theories; we are then on the well-known route from meaning wholism to social wholism and from there to social relativism and social constructivism. However, Cartwright would then face all the well-known objections to social relativism. In particular, Davidson argues in his famous essay 'On the very idea of a conceptual scheme' on the basis of meaning wholism and social wholism that the transition from this wholism to social relativism is incoherent: The idea of different conceptual schemes by means of which we approach the world is unintelligible (Davidson 1974). Hence, Cartwright either has to rebut the argument against different conceptual schemes (or different incommensurable perspectives on the world, etc.) or to elaborate on the metaphor of different theories carving out different aspects of an underlying whole in such a way that she is not committed to conceptual schemes or the like. If, however, there are no conceptual schemes in the sense of Davidson's attack on this notion, then, again, a unity of science and our knowledge as a whole seems to be, in principle, possible.

The metaphysical wholism that Cartwright contemplates (nature as an underlying whole) does of course not imply a commitment to semantic wholism, which Cartwright has to reject (and vice versa). However, in any case—the semantic case as well as the metaphysical one—wholism does not lend support to the thesis of a patchwork of laws. Thus, as far as semantics is concerned, Cartwright needs a theory of meaning as well as a theory of confirmation that is a credible alternative to the mainstream wholism in order to make her thesis of a patchwork of laws intelligible; as far as metaphysics is concerned, she cannot simply rely on the wholistic conception of nature as an interacting whole in order to be able to give an account of the relationship between the different domains of scientific theories in a dappled world.

NOTES

1. See in particular Cartwright (1991) and compare Cartwright (1999: 29–31).
2. For an introduction to the philosophy of quantum theory including entanglement and the relevant experiments see, for instance, Albert (1992).
3. As to a conceptual analysis of what this wholism amounts to, see Teller (1986); Howard (1989); Healey (1991); Esfeld (2001: Ch. 8).

4. As to decoherence, see the papers in Giulini et al. (1996) and in Blanchard et al. (2000).
5. See d'Espagnat (1971: Ch. 6.3) as to the differentiation between proper and improper mixtures.
6. See Ladyman (1998) on what he calls ontic structural realism, French & Ladyman (2003), and Esfeld (2004).
7. For a criticism, see Cachro & Placek (2002).

REFERENCES

Albert, D. Z. (1992) *Quantum Mechanics and Experience*, Cambridge: Harvard University Press.

Albert, D. Z., and B. Loewer. (1988) 'Interpreting the many worlds interpretation', *Synthese*, 77: 195–213.

Blanchard, P. et al. (eds) (2000) 'Decoherence: Theoretical, experimental, and conceptual problems', *Proceedings of a Workshop Held at Bielefeld, Germany*, Heidelberg: Springer.

Cachro, J., and T. Placek. (2002) 'On Cartwright's models for EPR', *Studies in History and Philosophy of Modern Physics*, 33B: 413–433.

Cartwright, N. (1989) *Nature's Capacities and their Measurement*, Oxford: Oxford University Press.

———. (1991) 'Can wholism reconcile the inaccuracy of theory with the accuracy of prediction?', *Synthese*, 89: 3–13.

———. (1997) 'Where do laws of nature come from?', *Dialectica*, 51: 65–78.

———. (1999) *The Dappled World. A Study of the Boundaries of Science*, Cambridge: Cambridge University Press.

Cartwright, N., and M. Suárez. (forthcoming) 'A causal model for EPR', *Reverberations of the Shaky Game: Essays in Honor of Arthur Fine*, Chicago: University of Chicago Press.

Chang, H., and N. Cartwright. (1993) 'Causality and realism in the EPR experiment', *Erkenntnis*, 38: 169–190.

d'Espagnat, B. (1971) *Conceptual Foundations of Quantum Mechanics*, Menlo Park: Benjamin.

Davidson, D. (1974) 'On the very idea of a conceptual scheme', *Proceedings and Addresses of the American Philosophical Association*, 47; reprinted in D. Davidson (1984) *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press.

Einstein, A. et al. (1935) 'Can quantum-mechanical description of physical reality be considered complete?', *Physical Review*, 47: 777–780.

Esfeld, M. (2001) *Holism in Philosophy of Mind and Philosophy of Physics*, Dordrecht: Kluwer.

———. (2004) 'Quantum entanglement and a metaphysics of relations', *Studies in History and Philosophy of Modern Science*, 35: 601–617.

Fodor, J. A. and E. Lepore. (1992) *Holism. A Shopper's Guide*, Oxford: Blackwell.

Foster, J. (1982) *The Case for Idealism*, London: Routledge.

French, S. & Ladyman, J. 'Remodelling structural realism: Quantum physics and the metaphysics of structure', *Synthese*, 136: 31–56.

Ghirardi, G. et al. (1986) 'Unified dynamics for microscopic and macroscopic systems', *Physical Review*, D34: 470–491.

Giulini, D. et al. (1996) *Decoherence and the Appearance of a Classical World in Quantum Theory*, Berlin: Springer.

Healey, R. A. (1991) 'Holism and nonseparability', *Journal of Philosophy*, 88: 393–421.

Howard, D. (1989) 'Holism, separability, and the metaphysical implications of the bell experiments', in J. T. Cushing and E. McMullin (eds) *Philosophical Consequences of Quantum Theory. Reflections on Bell's Theorem*, Notre Dame: University of Notre Dame Press.

Jackson, F. (1998) *From Metaphysics to Ethics. A Defence of Conceptual Analysis*, Oxford: Oxford University Press.

Ladyman, J. (1998) 'What is structural realism?', *Studies in History and Philosophy of Modern Science*, 29: 409–424.

Langton, R. (1998) *Kantian Humility. Our Ignorance of Things in Themselves*, Oxford: Oxford University Press.

Langton, R. and D. Lewis. (1998) 'Defining "intrinsic"', *Philosophy and Phenomenological Research*, 58: 333–345.

Lewis, D. (2001) 'Redefining "intrinsic"', *Philosophy and Phenomenological Research*, 63: 381–398.

Lockwood, M. (1989) *Mind, Brain and the Quantum. The Compound 'I'*, Oxford: Blackwell.

Putnam, H. (1975) 'What is mathematical truth?', in *Mathematics, Matter and Method. Philosophical Papers volume 1*, Cambridge: Cambridge University Press.

Quine, W. V. O. (1953) 'Two dogmas of empiricism', in *From a Logical Point of View*, Cambridge: Harvard University Press.

———. (1975) 'On empirically equivalent systems of the world', *Erkenntnis*, 9: 313–328.

———. (1986) 'Reply to Jules Vuillemin', in L. E. Hahn and P. A. Schilpp (eds) *The Philosophy of W. O. Quine*, La Salle: Open Court.

———. (1991) 'Two dogmas in retrospect', *Canadian Journal of Philosophy*, 21: 265–274.

Scheibe, E. (1991) 'Substances, physical systems, and quantum mechanics', in G. Schurz and G. J. W. Dorn (eds) *Advances in Scientific Philosophy. Essays in Honour of Paul Weingartner*, Amsterdam: Rodopi, 215–229.

Shoemaker, S. (1984) *Identity, Cause, and Mind. Philosophical Essays*, Cambridge: Cambridge University Press.

Teller, P. (1986) 'Relational holism and quantum mechanics', *British Journal for the Philosophy of Science*, 37: 71–81.

Reply to Michael Esfeld

One of the central aims of *The Dappled World* is to offer a metaphysical account of the patchwork way in which successful science operates as opposed to an epistemological account that relies on our ignorance and cognitive limitations. The challenge then is to account for how there can be the kinds of regularity and precise predictability that we see in a world that is not ordered through and through by some fundamental and precise, regularity-type laws.

I originally thought that wholism and nomological machines offered alternative answers to this challenge. Michael Esfeld, I think correctly, points out that my nomological machines story is itself a wholistic story. Nomological machines are imbedded in and interact with the rest of the hugely diverse and less systematically interacting world. He also points out, again I think correctly, that it is hard to tell the pure wholistic story—the one without nomological machines—without advertising to a systematic theory underneath. Indeed my own examples of how we might have highly successful theories that are nevertheless totally "wrong" all seem to depend on there being a "right" theory underneath. Nor would I wish to find myself having to maintain that this underlying theory must somehow remain inaccessible to us. I still believe that there is a proper wholistic story to be told without universal laws at all. But my own best efforts I think have instead been with nomological machines, which do presuppose capacity laws—as opposed to regularity laws—and that are in many cases very wide in scope, if not universal, e.g., the capacity law that masses attract other masses.

There is, however, a central issue in the second section of Esfeld's paper with which I continue to disagree: The power of quantum theory to serve as a model for an underlying theory. In *The Dappled World* I argue that quantum theory is extremely limited in its domain. Clearly Esfeld has not had the space here to take on these arguments, so the debate on this issue will have to take place elsewhere.

With regard to the third section of Esfeld's chapter, I think I am not wedded to either semantic wholism or the wholism of confirmation, so I would like to challenge his suggestion that my views are inconsistent with them.

CONFIRMATION WHOLISM

I reckon that, if it is true that almost any true claim bears evidentially on almost any other, then that is precisely because of the interconnected net of interactions in the world that my story of nomological machines, and the need for shielding, presupposes.

SEMANTIC WHOLISM

Semantic wholism, as Esfeld describes it, and, as I believe it is most plausibly constructed, depends on the very facts that support the wholism of confirmation: The meaning of a term depends on all the inferences it participates in and these are more or less the same inferences that connect with distant facts that bear evidentially on claims involving it. But these connections include the same connections that generate the need for shielding in the nomological machine story and thereby generate the account of how pockets of predictability and precision can do without a total cover of underlying universal regularity.

I wonder if Esfeld supposes the opposite because of the use of the term "aspect"—which I try to avoid. My story is not one of different perspectives, perhaps complementary in Bohr's sense. I present instead a story of one very complicated "God's eye" perspective in which there are a huge number of interacting qualities and quantities, many of which have fairly stable capacities that can be regimented to produce systematic and precisely predictable order if only properly shielded. So, on the nomological machines story, different theories do not carve out different aspects, where "aspects" are false, but perhaps useful representations of the world. Instead, different theories study different sets of features, all of which are supposed to be genuine, often interacting, features of one and the same reality.

15 How Classical and Quantum States Relate

Cartwright's Views of Quantum Theory[1]

Brigitte Falkenburg

INTRODUCTION

Over the years, Cartwright's views of quantum theory have changed twice. At first, she defended a version of quantum fundamentalism which was associated with a strong realistic interpretation of the wave functions of quantum mechanics. In 1975, she believed that quantum states and their superpositions belong to the ontological commitments of quantum theory, whereas mixtures do not (Cartwright 1975; 1983: 165, 169). The superpositions she took for real referred to states after measurements. Her argument was based on the theory of Daneri, Loinger, and Prosperi, according to which after a measurement a superposition dresses up as a mixture, in the sense that it makes the same statistical predictions (Daneri et al. 1962). According to this theory, measurement is an amplification process which ends up in thermodynamic equilibrium states with ergodic properties. The theory predicts quantum states that *look* classical even though they still *are* superpositions.[2]

In 1975 Cartwright believed that these superpositions exist because quantum theory says that they exist forever. In developing the views of her book *How the Laws of Physics Lie*, she came to think that such a metaphysical realism about superpositions is due to an unwarranted belief into the simplicity and uniformity of nature. The first step away from quantum fundamentalism was to criticize the realistic view of quantum theory. Now Cartwright defended a more modest realism of quantum processes. Her 1983 essay on the measurement problem, however, still expressed a certain fundamentalist hope; and she still hoped for a quantum statistical mechanics that might explain both the internal dynamics of a quantum system and the reduction of the wave packet by measurement. In a series of papers written in the 1990s, she took a step further away from quantum fundamentalism. Now she defends a disunified view of quantum physics. The last essay in *The Dappled World* explains how to have the quantum cake and to eat it too by ascribing to a system at once a quantum state and a classical state.

In the following, I investigate in more detail how Cartwright's positions of 1983 and 1999 relate to each other and to physics.

IS THE MEASUREMENT PROBLEM A PSEUDOPROBLEM?

Cartwright's 1983 essay on quantum theory has the provocative title 'How the measurement problem is an artefact of the mathematics'. The title seems to express the view that the notorious measurement problem of quantum theory is created by von Neumann's mathematical axiomatization of quantum mechanics. According to von Neumann, quantum mechanics distinguishes between two kinds of evolution:

1. Quantum dynamics: the internal evolution of a quantum system is governed by the Schrödinger equation. The internal quantum dynamics is deterministic, linear, and reversible.
2. Reduction of the wave function: measurements are governed by von Neumann's projection postulate; the measurement of a quantum system results in the projection of the wave function to an eigenstate of the operator that belongs to the measured observable. The reduction of the wave function is indeterministic, nonlinear and irreversible.

The distinction belongs to the quantum theory of individual systems. The reduction of the wave function describes the outcome of a single measurement. A mixture, on the other hand, results from the measurement of an ensemble, i.e. from many measurements on systems in the same quantum state, systems which have been identically prepared under well-defined experimental conditions. According to von Neumann's theory, the measurement problem is twofold. At the level of individual quantum systems, it is the question of how the wave function can reduce from a superposition to a single component. At the level of a quantum ensemble, it is the question of how a superposition can evolve into a mixture.

In 1975, Cartwright's answer was realistic in favour of the quantum dynamics and the resulting superpositions but antirealistic regarding the reduction of the wave function. In 1983, she suggests taking the reduction of the wave packet for real and mixtures too. Her early quantum fundamentalism came together with a realism of only the quantum dynamics (1). In 1983, it is replaced by a realism of quantum processes that embrace both the deterministic and reversible quantum dynamics (1) and the indeterministic and irreversible reduction of a superposition to the eigenstate of an observable (2). Such a realism of quantum processes, however, is at odds with the fact that quantum mechanics does not give a causal account of the temporal evolution of quantum processes over measurements. Thus quantum mechanics does not tell us what is really going on in a quantum process. This is Cartwright's 1983 argument in a nutshell. The central claims of the

1983 paper are: Measurement is as real as the quantum dynamics; both are real quantum processes, and, physically, both are on a par. They are just two kinds of physical interactions. But quantum mechanics in Hilbert space makes a mathematical distinction between them; here is the Schrödinger equation (1) and there the projection postulate (2). If both kinds of physical processes are on a par, they should not be split by the mathematics of Hilbert space into a quantum dynamics here and a reduction of the wave function there. Such splitting is an artefact of the mathematics.

Let us now have a closer look at the argument. Cartwright gives three distinct arguments in favour of her realism of quantum processes. I agree completely with the first one but not at all with the second and third. They run as follows.

1. Superpositions do not account for the temporal evolution of individual quantum processes.

This is an objection against her former arguments in favour of the Daneri–Loinger–Prosperi approach. First she sums up Bub's and Putnam's objections, stating that 'a superposition remains a superposition, even if it dresses up as a mixture' (Cartwright 1983: 170). In 1975, she disagreed with this conclusion. Now she supports it on the basis of the observation that dressing up as a *statistical* mixture does not explain the behaviour of *individual* quantum systems. Cartwright points at some of the oddities the Daneri–Loinger–Prosperi model predicts for the temporal evolution of the individual quantum-system-plus-apparatus behaviour. Her arguments hold also for more recent interpretations of quantum theory without reduction of the wave function, i.e. the consistent histories approach or decoherence. Because rigorous proofs exist that within quantum theory the outcome of an individual measurement in general does not have a definite result or cannot be objectified, it is here where the measurement problem has to be hunted, if anywhere (Mittelstaedt 1995, 1997).

2. There is nothing special about measurement. It is the same kind of physical process as preparation, and therefore it is like other physical interactions.

By comparing measurement and preparation, Cartwright wants to show that in a measurement the reduction of the wave function happens with or without an observer. She argues that the interaction of a quantum system with a measuring device is like the preparation of a quantum state under well-defined experimental conditions, both kinds of process result in a reduction of the wave function. A quantum state that is prepared for performing an experiment is *not* measured, but we take it for granted that it is in a well-defined physical state. If we do not, we must assume that, e.g., the particle beam in a scattering experiment of high-energy physics is in a very complicated superposition with the absorbing screen that selects particles in a well-defined dynamical state (Cartwright 1983: 172).[3]

Indeed in many regards the preparation of a quantum state is like a measurement. As recent experiments with so-called quantum erasers show, however, there are also important regards in which it is not. In the next section, I attempt to clarify *what* is special about measurement in contradistinction to preparation. It turns out that none of both processes is like other physical interactions, and that it is hard to say which one is more peculiar. For the moment, let me just state that Cartwright's comparison of measurement and preparation does not show per se that measurement is like any other interaction. Her conclusion is based on the claim that nothing is special about *preparation*. Her argument runs as follows: (i) Measurement is like preparation, (ii) preparation is an objective physical interaction, (iii) therefore measurement is like other physical interactions. But to regard preparation as an objective physical interaction is grounded in a nonoperational, realistic picture of what goes on in a subatomic process, say particle propagation. To argue then that nothing is special about measurement because the preparation of a well-defined quantum state is a measurement means begging the question.

3. Subatomic decays and scattering processes end up in well-defined particle states.

A closer look at Cartwright's paper reveals indeed realistic intuitions about particle trajectories. Such intuitions are repeatedly expressed in the paper. In contrasting old quantum theory and later quantum mechanics, she emphasizes that old quantum theory has the advantage of assuming that subatomic processes *really* happen. Old quantum theory is just agnostic about the when and why of radiative decays, whereas according to the Schrödinger equation *nothing* happens (Cartwright 1983: 192). In a similar line of reasoning, she insists that after a scattering process the particles must be *really* 'travelling one way or another far away from the target', even if there is no detector (Cartwright 1983: 192, 194). Both claims give the impression of an ignorance interpretation of the transition probabilities of particle physics. Cartwright seems to believe that there are real particles, and particle reactions whereas the transition amplitudes of quantum mechanics express that we do not know them. Indeed in the appendix of her 1983 paper she suggests, 'following Bohm', that after scattering, the quantum state is reduced to a momentum eigenstate, which corresponds to a particle 'travelling in a specific direction, and with a specific energy' (Cartwright 1983: 210). She tries to reconcile this idea with the optical theorem which is based on the unitarity of the S-matrix, whereas the reduction of the wave function she favours gives rise to a nonunitary evolution.

However, her defence of particle trajectories does not work.[4] The optical theorem relates the total scattering cross section at a given energy to the imaginary part of the scattering amplitude $f(\theta)$ in forward direction $(\theta = 0)$:[5]

$$\sigma_{\text{tot}} = 4\pi/k\ Im\ f\,(0)$$

The unitarity assumption on which the optical theorem is based expresses the conservation of probability. In nonrelativistic quantum mechanics, this means that the scattering conserves the particle number. No particles get lost. Here, the term "particle" has an operational meaning. It is related to the measurement of the particle flux of a given beam by means of a particle detector. The optical theorem means that the flux of the incoming particles must be equal to the flux of the outgoing particles. Correspondingly, it tells us that due to the scattering the wave function Ψ_0 of the incoming particles is partially cancelled, taking into account that particles are scattered off at angles $\theta \neq 0$. This scattering off gives rise to a shadowing effect which is expressed by the optical theorem. After scattering, the incoming wave Ψ_0 and the scattered wave Ψ_S are in a superposition:

$$\Psi = \Psi_0 + \Psi_S,$$

where the scattered wave Ψ_S obeys the optical theorem. It is hard to see how the optical theorem might have any meaning beyond the usual operational interpretation of the quantum mechanics of scattering and be related to a nonunitary evolution of the wave function without measurement.[6] It expresses particle number conservation in case of particle detection before respectively after the scattering of a given beam at a given target.

Cartwright argues correctly that the above superposition is merely formal, whereas the physical state is Ψ. In addition, she argues that the physical state Ψ as well as the optical theorem are compatible with the reduction of the resulting wave function to momentum eigenstates (Cartwright 1983: 210). This picture holds in the specific semiclassical case she discusses; however, it breaks completely down whenever typical quantum phenomena come into play. It holds for low-energy scattering when there are no contributions of angular momentum $l \neq 0$ to the scattering amplitude, and the scattered wave Ψ_S is isotropic. But in general, the scattering amplitude $f(\theta)$ contains additional contributions corresponding to angular momentum $l > 0$. The resulting scattered wave Ψ_S is anisotropic.[7] Now assume scattering of sufficiently high energy such that the $l = 1$ contribution can no longer be neglected. In this case (which still permits semiclassical considerations, according to Mott and Massey's classical textbook), the scattered wave is a superposition of two angular momentum eigenstates. If only a single particle is scattered at a given time, the unitarity condition underlying the optical theorem tells us that the particle must be in a superposition of these angular momentum states, both having an amplitude smaller than one. In semiclassical approximation, these states correspond to distinct trajectories with different impact parameters and different scattering angles θ (Mott & Massey 1965: 102, 356).

In the full quantum mechanical treatment, according to the Schrödinger evolution the scattering results in a 2-particle quantum state. Before any decoherence or measurement, the states of the scattered particle and the scattering center are entangled in a similar way as the 2-particle wave function of an EPR pair. Indeed the scattered particle and the scattering center can be entangled in such a way that the outgoing wave function is a superposition of two distinct momentum states. Such a superposition may give rise to an observable interference pattern. See, e.g., the interference effects that have been observed in proton-hydrogen scattering with charge exchange ($H^+ + H \rightarrow H + H^+$) and are due to the production of an entangled H_2^+ system (Mott & Massey 1965: 655).[8]

In criticizing Cartwright's second and third arguments, I do not want to support an antirealistic view of quantum theory at all. From a physical point of view, realistic assumptions about preparation, measurement, and particle propagation are good. They are good for doing particle physics. They are also good for extending the scales of the familiar physical magnitudes such as length, time, mass, or energy into the subatomic domain. In doing so, they provide particle physics with a powerful heuristics for the design and the data analysis of scattering experiments. From a philosophical point of view, however, such realistic assumptions about particles need clarification. We simply do not know what kind of entity exactly these "particles" are, except that they are collections of physical properties such as mass, spin, parity, and the (generalized) charges. According to these properties the irreducible representations of the Poincaré group or other (internal) symmetry groups of particle physics are classified. But these properties are type instead of token marks. Quantum theory does not specify the specific marks of individual subatomic particles. What is more, it is a theory telling us that there are none. According to quantum mechanics, observable particle tracks, e.g., the tracks measured in a bubble chamber, are *not* caused by individual particles. Quantum mechanics tells us that particle tracks are nothing but repeated position measurements. Without measurements there are no particle tracks (Mott 1929; Bethe 1930; Falkenburg 1996; 2007). There are only the conservation laws for mass, charge, spin, and the other dynamic properties of these so-called particles. The quantum story of particle tracks is exactly like the case of radiative decays. According to the Schrödinger equation alone, *nothing* happens.

At this point we are thrown back to Cartwright's first argument, the one which I accept. Quantum mechanics does not account for the history of individual systems. Indeed this is the only substantial argument against the view that the Schrödinger evolution might tell us the complete story about quantum processes. We do not know what exactly happens in the preparation or measurement of an individual quantum system. Quantum mechanics is an abstract and symbolic description of quantum processes which remains tacit on this most interesting subject. In her 1983 paper, Cartwright wants to fill the explanatory gap with probabilistic explanations only. On the basis of her convictions about measurement as a real physical interaction, she takes two further steps to argue that the measurement problem is a mathematical artefact. As these arguments are quite convincing, I ask myself why she has not taken them up again in her later papers on quantum theory.

4. The only real probabilities of quantum theory are transition probabilities.

(Now I say "quantum theory" because this claim also holds for quantum mechanics, quantum electrodynamics, quantum optics, etc. Transition probabilities and the calculation of the corresponding scattering matrix elements are common, and crucial, to all kinds of perturbative quantum dynamics.) Cartwright rejects the widespread quasi-classical interpretation of the squared amplitude of the wave function as a probability density. In the sense of a classical probability density, this interpretation obviously does not hold. In the famous double-slit experiment, $|\Psi(\boldsymbol{r})|^2\, d^3r$ does not mean the probability of the real particle location in a classical sense (Cartwright 1983: 175). For the wave function that enters the calculation of subatomic magnitudes such as the dipole moment, the same expression has even no direct operational meaning. The usual interpretation of $e|\Psi(\boldsymbol{r})|^2$ as the effective charge density of a hydrogen atom is not justified by the probabilistic interpretation of quantum mechanics (Cartwright 1983: 186). Against such a strict operationism, one might object that if we would repeatedly measure the position of the electron within the hydrogen atom, then we would get the probability distribution that determines the effective charge density. On the basis of the usual probabilistic interpretation of quantum mechanics, this counterfactual claim does not make much sense in defence of a statement about subatomic structure as such, i.e. without being measured.[9] Cartwright emphasizes that $e|\Psi(\boldsymbol{r})|^2$ and a classical dipole moment or charge distribution are formally analogous, but 'the analogy is purely formal' (Cartwright 1983: 191).[10] She argues that there is no real event space to which the probability $|\Psi(\boldsymbol{r})|^2\, d^3r$ is related to (Cartwright 1983: 176). Therefore the only real probabilities of quantum mechanics are the probabilities of transitions between (pure) quantum states (Cartwright 1983: 179). The probabilistic interpretation of quantum mechanics is only based on transition probabilities, as it was in Born's original papers (Born 1926 a, b).[11] Convincing as this claim is, however, in one regard it has to be qualified. For the theoretical description of complicated experiments with several subsequent preparations and measurements we need in addition *conditional* probabilities. They are causally relevant factors in explanations of the event structure of quantum mechanics. In this sense they are real probabilities of quantum theory as well.

5. Regarding the transition probabilities, measurement is like scattering or radiative decay.

On Cartwright's line of reasoning, the following conclusions can be drawn (Cartwright 1983: 195). If transition probabilities are the only real probabilities of quantum theory, then the expectation values of observables are on a par with scattering amplitudes, and at the probabilistic level measurement is indeed like scattering or radiative decay. The adequate formal tool to treat measurement and scattering in the same way is the density matrix of a state, the basic expression of quantum statistical mechanics. Thus the quantum theory of measurement here and the quantum theory of scattering there should turn out to be two related cases of quantum statistical mechanics. In contradistinction to the internal Schrödinger evolution of a quantum system, the reduction of the wave function is irreversible and indeterministic, like radiative decays. Therefore Cartwright compares measurement with exponential decay. She reminds of her analysis of the Wigner–Weisskopf derivation of the exponential law according to which a radiating atom is coupled to a quasi continuum of electromagnetic field modes. Analogy teaches that the indeterministic and irreversible reduction of the wave function should occur whenever a quantum system is coupled to a system with a very large number of degrees of freedom (Cartwright 1983: 196). She concludes that in quantum statistical mechanics of open systems it should be possible to derive the reduction of the wave function, and there should no longer be a measurement problem.

From this point of view, the measurement problem seems a pseudoproblem. It seems to be generated by quantum mechanics in Hilbert space as a mathematical theory of closed systems with finite degrees of freedoms, and it should better be made to disappear in a quantum theory of open systems. In the last analysis, the reversible and deterministic Schrödinger evolution of a quantum system and the reduction of the wave packet should turn out to be distinct cases of one-and-the-same quantum statistical mechanics. Cartwright accepts the objection that this is nothing more than a research program, but she expresses the 'hope that this mundane, though difficult, job of physics is all that there is to the measurement problem' (Cartwright 1983: 206).

WHY PREPARATION AND MEASUREMENT ARE DISTINCT

According to quantum mechanics in Hilbert space, the measurement problem dwells in the distinction between reversible and irreversible processes. The Schrödinger evolution of the wave function is reversible and deterministic. Measurements are not. A similar distinction makes the difference for preparation and measurement. Quantum mechanics gives completely different descriptions for both ways of handling quantum states. The preparation of a quantum system aims at a pure quantum state described by a wave function. Whether the wave function is in a superposition depends on the choice of the basis in Hilbert space. The choice of a basis where the wave function is in an eigenstate of a Hermitean operator corresponds to the choice of an observable. Correspondingly, by means of an appropriate experimental device the eigenstate of an observable can be changed into a superposition and vice versa. Exactly this happens when a quantum state is prepared. The wave function tells us in addition how to undo the preparation. Measurements are completely different. After the reduction of the wave function by a measurement it is impossible to re-establish the nonreduced quantum state before measurement nor to recover its full former information.

To a certain extent, the distinction between preparation and measurement is an artefact of quantum theory in Hilbert space. Let me illustrate this point by discussing a typical experiment with electromagnetic waves. It may be performed either with classical light (white light or a laser beam) or with a low-intensity light beam from a short-pulsed laser. In the first, the experiment is done with classical electromagnetic waves. In the second, single photons stemming from very short pulses of laser light are used.[12] They have wave-like properties even though they can be localized as particle-like energy quanta at a screen or by means of a photon counter, as in the famous double-slit experiment.

Let white light from an ordinary light source pass through three subsequent polarizers $P_{|}$, $P_{/}$ and $P_{\_}$ crossed against each other with angles of 45° respectively 90°. Behind the polarizer $P_{|}$ the light is polarized in vertical direction, behind $P_{/}$ it is polarized with 45° relative to $P_{|}$, and behind $P_{\_}$ it is polarized horizontally. If you remove $P_{/}$ no light passes $P_{\_}$, as the remaining polarizers are perpendicular to each other. Now look at a screen behind $P_{\_}$. Even in the classical case it is amazing to see the light on the screen appear and disappear when you put the second polarizer in and out.

The corresponding quantum phenomenon was first discussed in Dirac's famous textbook on quantum mechanics (Dirac 1958: 4).[13] The experiment with low-intensity light from very short laser pulses and a photon counter in place of the screen is performed. The beam intensity respective to the pulse time should be so low that only one photon at a time is in the field. According to quantum theory, in correspondence to the classical wave picture, single photons are detected behind $P_{\_}$ if and only if the second polarizer $P_{/}$ is put in. In the quantum case, however, this is very striking. The physical effect of the polarizers is obviously to select photons of a given polarization, respectively, to absorb all photons with perpendicular polarization. How can the single photons pass three absorbers given that they cannot pass two absorbers? My answer is that the polarizers prepare distinct wave-like quantum states with or without the second polarizer $P_{/}$, whereas the photon detector measures single photons.[14]

At this point, the experiment requires further analysis. The quantum light consists of photons in well-defined polarization states $|\Psi_{\varepsilon}>$ prepared by means of the polarizer $P_{|}$, $P_{/}$ or $P_{\_}$. Quantum field theory describes them in terms of field operators for the annihilation and creation of field quanta. The photon states $|\Psi_{\varepsilon}>$ represent field modes of given frequency and polarization. A low-intensity field with single photons is in a well-defined number state. In a number state, the phase is totally unsharp. The highly nonclassical properties of such light do not affect its description in terms of wave-like field modes with well-defined polarization. Behind the first polarizator $P_{|}$ the light is polarized in vertical direction, let us say it is in the quantum state $|\Psi_{|}>$. $P_{|}$ reduces the wave function $|\Psi >$ of the short pulsed laser beam (which must already be a single field mode) to this state of well-defined, vertical polarization. In terms of the second polarizer $P_{/}$, this state is a superposition of a photon wave or field mode $|\Psi_{/} >$ that can pass $P_{/}$ (polarization of –45°, relative to $P_{|}$) and a photon wave or field mode $|\Psi_{\backslash} >$ that cannot pass it (polarization orthogonal to $P_{/}$):

1. effect of $P_{|}$: $|\Psi_{|}> =$ ½√2 $(|\Psi_{/}> + |\Psi_{\backslash}>)$
Behind the second polarizer $P_{/}$ the photon field is in the state ½√2 $|\Psi_{/}>$, due to reduction of the wave function. In terms of waves passing the polarizers $P_{|}$ and $P_{/}$, however, this state is a superposition ½ $(|\Psi_{|}> + |\Psi\_>)$, with a state $|\Psi\_>$ of horizontal polarization that is orthogonal to $|\Psi_{|}>$. The quantum state $\Psi_{/}$ corresponds to a superposition of photons that *could* pass the first polarizator $P_{|}$ and photons that could *not* pass $P_{|}$ but only the perpendicular polarizer $P\_$:
2. effect of $P_{/}$: ½√2 $|\Psi_{/}> =$ ½ $(|\Psi_{|}> + |\Psi\_>)$
Finally, let the photon wave pass through the third polarizer $P\_$ which is crossed with 90° to the first one. Behind $P\_$ you detect single photons if and only if $P_{/}$ is between $P_{|}$ and $P\_$. Your observations exhibit exact correspondence to the classical case already described: With $P_{|}$ and $P\_$ alone, without the second polarizer $P_{/}$, *no* light passes. Not a single photon is detected. But if you put $P_{/}$ in between them, some "surviving" photons are detected at the screen:[15]

3a. effect of $P\_$ on $|\Psi_{|}>$: $0 |\Psi\_>$
3b. effect of $P\_$ on $|\Psi_{/}>$: ½√2 $|\Psi\_>$
What is going on here? Obviously each of the polarizers reduces the wave functions to a well-defined polarization state. The reduction of the wave function is measurement-like. It gives rise to the absorption of *some* photons in the polarizers. Each polarizer damps the amplitude of the wave function or field mode by a factor ½√2; i.e. each act of preparation by polarization results in damping the photon intensity by a factor ½. The resulting photon intensity of the effect (3b) is ¼ of the effect (1). In terms of transition probabilities or counting rates, this means that finally in the photon counter 3 out of 4 photons got lost.[16] But what has happened to the single photons? Any photon in the quantum field has been either absorbed at one or the other of the polarizers or detected behind $P\_$. In stating this we should be aware that the concept of photon absorption has no operational meaning as long as no photon measurement is performed at $P_{|}$ or $P\_$.

The last point is crucial. It makes the difference for preparation and measurement. The absorption of the "lost" photons is not observed before the single photons have been detected and counted behind $P_{\_}$. It is impossible for any polarizer to damp the intensity of a single photon in a well-defined number state by a factor ½. As long as only one photon at a time is in the quantum field, no photon has been absorbed at any of the polarizers whenever a photon is finally detected. Thus in the single photon case, the preparation of the photon state by any of the polarizers before the screen is obviously not a measurement. Otherwise, we could not undo it by inserting the second polarizer. The preparation of the photon state is either made by means of the polarizer $P_{|}$ and $P_{\_}$ alone. In this case no photons are detected at all in the counter. Or it is made by $P_{|}$, $P_{/}$ and $P_{\_}$ together. In this case some photons survive. Putting in $P_{/}$ erases the well-defined polarization of state $|\Psi_{|}>$ and re-establishes a nonzero probability of measuring a photon in state $|\Psi_{\_}>$. Whenever a photon is detected behind $P_{\_}$, its polarization state was $|\Psi_{\_}>$. Otherwise, it would not have passed the polarizer $P_{\_}$.

To what extent is this distinction between preparation and measurement an artefact of quantum theory in Hilbert space? In quantum field theory, the transformation of a pure polarization state into another pure polarization state looks like a reversible process, whereas the final photon detection is irreversible. Each polarizer prepares the photons in a pure quantum state that can be transformed in a reversible way into another pure quantum state, i.e. the pure polarization states $|\Psi_{|}>$, $|\Psi_{/}>$ or $|\Psi_{\_}>$. The change can be undone as long as the wave remains in one or the other pure polarization state (and as long as some photons are left in the quantum field, i.e. the expectation value of the photon number remains > 0). Operationally, and in contradistinction to measurement, preparation means not reading out the information contained in the wave function.[17]

However, undoing the preparation of a pure state is obviously not the same kind of thing as the reversible evolution of a wave function underlying a unitary dynamics, e.g., the Schrödinger equation. The transition probabilities of quantum field theory tell us that 3 out of 4 photons are absorbed at one of the polarizers. If the probability of photon detection is 1/2 given that $P_{/}$ is there, it is 1/4 given that in addition $P_{\_}$ is there. The way in which the photon polarization is prepared has causal relevance for the transition probability, and this causal relevance is expressed in terms of conditional probabilities. The preparation of the quantum state by means of all three polarizers gives the measurement result "some photons pass" (with mean relative frequency close to 1/4), whereas taking $P_{/}$ out results in "no photon passes". And the preparation that lets *no* photon pass can be changed into a preparation that lets some photons pass by again putting the second polarizer in between the first and the third. Thus the preparation of the quantum state is causally relevant for the final photon-or-no-photon decision. The decision can be changed forth and back by changing the preparation. In this operational sense the photon preparation is a reversible act (it can be undone) whereas the photon measurement is not.

At this point the following objection could be made. In a realistic picture of photon propagation one imagines that single photons are travelling one after the other through the system of polarizers until they are absorbed somewhere, be it at one of the polarizers or be it at the final photon counter. This picture does not fit in with a strictly operational view of quantum field theory, but it might nevertheless be tenable. From a realistic point of view it seems justified to regard preparation and measurement as the same kind of physical process. But to use this comparison for giving support to a realistic picture of measurement (as Cartwright does) obviously means begging the question.

In addition, recent experiments with quantum erasers prove that preparation and measurement are indeed very distinct. They show that the distinction is much more than a mathematical artefact. The typical quantum eraser is built into a *which way* experiment with a Mach–Zehnder interferometer or a double slit.[18] In the Mach–Zehnder interferometer experiment, low-intensity light is sent through a system of two beam splitters and two mirrors in such a way that it can take two possible paths until single photons are detected by means of the two final photon counters. Both paths are provided with such a phase shift that when they are united by means of the second beam splitter an interference pattern is obtained. Then in one of the paths a *which way* detector is installed, say a nonlinear down conversion crystal (which makes a pair of two correlated lower frequency photons out of one) with a photon counter. Putting in the *which way* detector makes the interference pattern disappear. However, if two completely symmetric *which way* detectors are installed in both possible paths, it is possible to make them erase the *which way* information mutually. To make them do so they have to be installed together with an additional beam splitter in such a way that none of the photons detected in the experiment has an unambiguous path. The idea is very simple, but to perform the experiment is very tricky because it requires optical devices of high precision.[19]

I cannot go into any detail here, but in principle such a quantum eraser experiment is very similar to the seemingly trivial polarization experiment discussed above. The single photons are prepared in two distinct ways by means of two or three beam splitters, just as the polarization states are prepared distinctly by means of the two respectively three polarizers. With two beam splitters only, the *which way* detectors localize single photons taking a definite path. If the third beam splitter is put in, the *which way* information is erased (i.e. the preparation is undone), and the photons make up an interference pattern as if there was *no which way detector* at all.

There is only one crucial difference between both experiments. In the *which way* experiment with or without the quantum eraser, no photons are absorbed by the beam splitters that make the preparation. No substantial photon intensity is lost (pace some minor losses, which are due to the nonideal properties of *any* experimental device). Here, the quantum field states with or without *which way* information are transformed into each other in a completely reversible way. Now the analogy between the reversible Schrödinger evolution of the wave function and the possibility of undoing the preparation holds. Undoing the preparation respects the Schrödinger evolution, in contradistinction to measurement. No reduction of the wave function takes place before any photon is detected.[20] Thus in a *which way* experiment with a quantum eraser, the photon states that are prepared are definitely not measured before the single photons are counted.

HOW TO HUNT THE MEASUREMENT PROBLEM

The preparation of a quantum state remains *gespenstisch* or ghost-like even though it has causal relevance for the event structure of any quantum theory, and even though this causal relevance can be explained in operational terms. Photon detection seems to be much more real. Its crucial mark is its irreversibility, and it gives rise to an observable event. All transition probabilities of quantum mechanics or quantum field theory are given in terms of observable events. The very concept of a transition probability, which is crucial for the quantum mechanics of scattering as well as for the density matrix of quantum statistical mechanics, presupposes the existence of real events. They make up the event space on which probabilities are defined. But with every event that happens an irreversible measurement has taken place. In this regard transition probabilities, scattering processes, and measurements are all alike, as Cartwright claims in her 1983 paper. Like measurement results, radiative decays are irreversible. They are due to spontaneous emission, and this seems to be the inverse process of the irreversible absorption of field quanta in a measuring device, i.e. particle detection. In being irreversible, radiative decay and photon absorption are measurement-like. They can indeed all be described within quantum statistical mechanics. The preparation of a quantum state, on the other hand, is taken into account in quantum statistical mechanics in terms of conditional probabilities. The only thing that cannot be integrated into the statistical framework is the temporal evolution of individual quantum processes.

At this point, we should compare Cartwright's philosophical views on quantum mechanics with recent developments in quantum physics.[21] Twenty years after she considered the measurement problem to be a pseudoproblem generated by Hilbert space mathematics, the situation of quantum physics appears as paradoxical as it did two decades ago. Her hope to develop a common framework for the irreversible reduction of the wave packet here and the reversible quantum state evolution there has failed until today. The measurement problem is still the most painstaking obstacle for a unified physics and predominantly for the attempts at finding a quantum theory of gravity. But the decoherence program has made substantial progress. Due to this progress, Cartwright's program to resolve the measurement problem within quantum statistical mechanics has turned from philosophy into physics. In the last analysis, however, decoherence does not resolve the measurement problem, and this successful research program is one of the approaches she explicitly rejects in her later papers on quantum theory.

Today, the quantum theory of decoherence gives very detailed predictions of how superpositions may turn into quasi mixtures. Decoherence is a dissipative process. It is due to the coupling of a quantum system to the degrees of freedom of its environment. If it takes place, the interference terms coming from superpositions are damped away very fast close to zero, even if they contribute substantially to the transition probability immediately after a measurement-like interaction (Guilini et al. 1996). The theory of decoherence is based on the very quantum statistical mechanics of open systems into which Cartwright put her hope in 1983. The calculations give statistical predictions for the time scale in which the interference terms of the transition probabilities are damped away. The timescale of this damping away is very short. However, it is not too short to be measured. In recent experiments, the decoherence of the transition amplitudes has found empirical support (Brune et al. 1996). The experiments give evidence that decoherence is a physical process that should no longer be neglected in the philosophy of physics. In recent experiments, even the timescale of the disappearance of the interference terms has been measured (Myatt et al. 2000). The decoherence program has achieved enormous successes at the probabilistic level but, unfortunately, only there. Superpositions evolve indeed into quasi mixtures according to theory, whereas in the experimental mixtures are measured.

Thus at the level of the individual quantum processes that happen in measurements, nothing has been resolved. Decoherence tells us neither that a superposition really turns into a mixture nor does it give a causal mechanism for the reduction of the wave function. No one makes this claim. In this regard, the promising decoherence program has failed like all the other attempts at a quantum theory of measurement. The old objection raised against the Daneri–Loinger–Prosper approach still holds: A superposition remains a superposition even if it dresses up as a mixture. Decoherence predicts quasi mixtures of very probable events, but it does not predict mixtures of actual events. Quantum statistical mechanics remains a probabilistic theory. It does not explain why and in which way the wave function is reduced during an individual measurement process. Therefore it can not explain why individual events and processes happen. It simply presupposes *that* they happen, and it gives probabilistic predictions for them, like ordinary quantum mechanics. Quantum statistical mechanics presupposes the occurrence of the very individual events that ordinary quantum mechanics cannot explain.

Where do we have to hunt the measurement problem, then? Obviously in the transition from probabilistic predictions to the description of individual processes here and in the distinction between reversible and irreversible processes there. The whole measurement problem lies in the following question. How can we get a reversible and linear dynamics and an irreversible and nonlinear dynamics from one-and-the-same theory of individual systems? This very question touches deep issues in the foundations of physics. It is related to the problems of the arrow of time, the increase of entropy, the interpretation of probability, the validity and scope of idealizations in physics, etc. Thus the measurement problem is a serious problem in the foundations of physics, and only there.

If one rejects fundamentalism in physics, as Cartwright does, there is no measurement problem. Physics proceeds without resolving it, in a most successful patchy view of the physical world. I interpret the shift in Cartwright's views on quantum mechanics as follows. In 1983, she thought that the measurement problem was a pseudoproblem generated by of Hilbert space mathematics and that it might well be resolved by means of another mathematics in a quantum statistical mechanics of open systems. In the two past decades, this hope has not been fulfilled. Quite on the contrary, decoherence and all known stochastic collapse models have failed in this regard.[22] In addition, no other new approach to the interpretation of quantum mechanics, e.g., consistent histories, has succeeded in resolving the measurement problem. In view of these developments, Cartwright's belief that physical theories are simply not capable of giving uniform theoretical explanations became overwhelming. Thus her view of the measurement problem as a pseudoproblem shifted towards a completely disunified view of quantum and classical states. If all attempts to solve it prove fruitless, just neglect it, and see how quantum physics works.

This pragmatic way of handling the measurement problem is convincing as far as it agrees with the practice of physics. From an epistemological point of view, however, it is not. Epistemology demands that we investigate the limitations of the foundational program of physics in more detail. Indeed it is possible to get a bit further in localizing the measurement problem as a painstaking problem that lies at the heart of quantum physics. There are very nice theoretical proofs in the line of von Neumann's analysis of measurement. They indicate that at the probabilistic level, quantum mechanics is not at odds with the existence of classical events, whereas at the level of the individual events, it is. Peter Mittelstaedt has shown recently that quantum mechanics is semantically consistent with regard to the generalized Born interpretation but semantically inconsistent with regard to the existence of objective measurement results. If you treat many identical systems as uncorrelated parts of a many-body system, the quantum theory of measurement without the projection postulate gives you the right expressions for transition probabilities. Nevertheless it does not give you any objective event (Mittelstaedt 1995: 2000). If *this* result is not a serious semantical paradox I have no idea of what a paradox is. It arises in a quantum theory of individual systems, by replacing an ensemble of many independent systems by one individual system with many independent parts. In quantum statistical mechanics, however, it does *not* arise. At the ensemble level, quantum theory is semantically consistent. But its probabilistic interpretation presupposes an event space that quantum theory can by no means explain, as it does not give us individual events and measurement results.

Thus the situation we envisage is even worse than Cartwright claimed in 1983. There are rigorous proofs that the explanatory gap between quantum theory and the history of individual systems over measurements can by no means be closed. The quantum theory of measurement is semantically incompatible with the need to tell the causal stories of concrete individual systems. (Consistent possible histories are not enough, we need real histories.) Something is seriously wrong in the relation between quantum physics and the existence of a classical world. Superpositions can by no means evolve into mixtures. They are completely at odds with the actual existence of physical events.

HOW TO HAVE YOUR QUANTUM CAKE AND EAT IT TOO

Let me now turn to Cartwright's actual view of quantum theory. It is summed up in the last essay of her book *The Dappled World*. It fits in well with the general view of a disunified physics defended there. Now she no longer defends a realistic interpretation of quantum theory, neither of quantum states, nor of quantum processes, nor of the reduction of the wave function. Instead she defends an instrumentalist view of classical physics here and quantum theory there. She proposes to ascribe two states to a system, a quantum and a classical state, whenever needed. Given the theoretical results sketched above, her disunified approach to quantum physics is justified to a certain extent, as I show now.

The 1999 essay 'How quantum and classical theories relate' starts by saying what is wrong with quantum theory. It is the aforementioned superposition problem: According to quantum theory, a superposition remains a superposition even if it dresses up as a mixture. Cartwright discusses two strategies of dealing with it. The first is reduction. It embraces all efforts to derive the reduction of the wave function, for example by means of a quantum theory of open systems. After pointing out some of the enormous difficulties of this strategy, which she had adopted in her 1983 essay, she rejects it. The second strategy is the disunified approach to quantum physics that she adopts now. She calls it the "have-your-cake-and-eat-it-too strategy".

How can a quantum state evolve from a superposition to a mixed state that describes the possible outcomes of measurements? According to the theoretical results sketched above, it cannot. The reduction strategy of resolving the superposition problem does not work by any means.[23] If we insist on a quantum theory of measurement, we are left with a serious dilemma. (1) One horn is insisting on quantum fundamentalism without admitting the existence of an actual world. If quantum mechanics is fundamental, we have to conclude that there are no actual events. (2) The other horn is insisting in the existence of an actual world without admitting quantum fundamentalism. If we take it for granted that there are actual events then quantum mechanics cannot be fundamental. Given that superpositions can by no means evolve into mixtures of measured states, only these two logical options remain. Closer examination shows that the second can be understood in several ways.

1. Believe in quantum fundamentalism and forget about an actual world. This happens if we follow Cartwright 1975 in taking superpositions for real but mixtures not, and if we insist that quantum theory applies to invidual systems (and not only to ensembles). If we add decoherence, we end up in well-split possible worlds where the individual measurement outcomes *might* happen. Now, if we think that quantum theory is fundamental, we have to replace the usual ontology of the actual by an ontology of mere possibilities, given that quantum theory does not give us individual events. Then we end up with the conclusion that according to quantum theory the actual world with its (quasi-) classical spatiotemporal structure is only possible, i.e. it does not *really* exist, it is at best highly probable.
2. Believe in an actual world because we measure actual events. In view of Mittelstaedt's results, this means the following. We presuppose that there is an actual, classical world, but we have to accept that quantum mechanics can by no means explain the very events of which the event space of quantum probabilities consist. But now we have at least two further options:
   2.1 Try to maintain a modest quantum fundamentalism. Accept that quantum theory is a probabilistic theory, and be content with it. You may consider quantum statistical mechanics to be a fundamental theory which might unify the four known physical interactions. But dispense with two foundational ambitions:
      (i) Do not attempt to explain the outcome of individual measurement results.
      (ii) Do not write down the wave function of the universe. (If you want to quantize gravity, *please* do it locally.)

   2.2 Reject any quantum fundamentalism and believe in a patchy world.

Option 1 is heroic. To take it means to have the quantum cake without getting something to eat. In the last analysis, this results in idealistic or platonist disbelief in the existence of the actual world. Julian Barbour defends a fascinating version of this kind of unwarranted speculative metaphysics (Barbour 1999).[24] However, I suspect that this option is based on a serious conceptual confusion. As stated here, it confuses our modal distinction of the possible and the actual. Do not forget that a probabilistic theory such as quantum theory presupposes the existence of an event space. Any concrete application of probability presupposes actual, individual events. (Passing over from option 1 to metaphysical realism gives the many-worlds interpretation of quantum mechanics. This means to have your quantum cake and to have to eat it many times at once.)

Option 2 is pragmatic. It means that quantum theory has to be complemented in one or another way by classical descriptions of the systems under investigation. To take it is to accept a peaceful coexistence of quantum and classical descriptions of one-and-the-same system. This is a step back to Niels Bohr's complementarity view of quantum physics. According to Bohr, abstract quantum theory needs to be interpreted in terms of complementary classical concepts.

As usual in a dilemma, rejection of option 1 compels us to adopt option 2. Then we have to dispense with quantum fundamentalism and adopt one or another version of a disunified physics. So far Cartwright's 1999 position is logically justified. This is what she calls having your cake and eating it too. In choosing the second horn of the dilemma, however, she rejects the modest quantum fundamentalism (2.1) and adopts a radical version (2.2). No one is compelled to do so. Indeed today most physicists adopt position 2.1. Some belief in the possibility of unique foundations of physics is needed as a justification for the most interesting research programs of today's quantum physics. Modest quantum fundamentalism enables the physicists to work on quantum computers, to construct particle detectors for doing particle astrophysics with cosmic rays, or to play with models of super strings, quantum gravity loops, etc.

Cartwright's position (2.2) blocks such foundational research programs of physics, even if they are based on a modest quantum fundamentalism. Her antifundamentalism is close to Bohr's philosophy of complementarity, but it is not identical with Bohr's view of quantum physics, if I understand her correctly. She suggests dealing with the superposition problem by following W. Lamb in terms of a semiclassical approach. Her semiclassical solution consists in identifying the state before measurement with a quantum state and the state after measurement with a classical state. The solution consists in the simple rule: Whenever before measurement there was a superposition, just replace it by a classical state after measurement. In this way, the measurement problem is obviously not resolved but eliminated in favour of an instrumentalist view of modelling in quantum physics. Let me give a few comments on the central features of this nonsolution of the superposition problem (Cartwright 1999: 214).

1. Quantum and classical descriptions can be true at once and of the same system. This is well known in quantum physics, as found in every introductory textbook. Both descriptions can be true whenever there is exact correspondence between the classical and the quantum case or model. Here, correspondence means two things. (i) Both the quantum

and the classical model of a system are interpreted in terms of the same kinds of physical magnitudes with classical or quasi-classical operational meaning (mass, charge, momentum, energy, etc.). Formally, classical and quantum magnitudes are entirely different, but related by the latter giving probabilities for the former. In addition, their measured values are embedded into one-and-the same scales of length, time, mass, energy, momentum, etc. In this sense, physical magnitudes are always expressed in the language of classical physics even though their measurement may be based on a quantum theory of the subatomic structure of matter. (ii) Both models make approximately or exactly the same quantitative predictions. In this case, we deal with so-called quasi-classical phenomena which give rise to a variety of semiclassical models in quantum physics. For example, Rutherford's famous formula of the cross section of Coulomb scattering is identical in the classical and the quantum case. Here, the classical formula corresponds exactly to the quantum formula, and thus both are true of the scattering of an electron beam at a point-like charge, e.g., a hydrogen atom. (Both formulas are true even though the very distinct atomic models on which they are based are obviously not both true. Especially, quantum mechanics eliminates Rutherford's electron trajectories with good reasons.)

2. The formal relation between classical and quantum observables is established through a correspondence rule. The best-known rule is a generalized version of Bohr's principle of correspondence. In old quantum theory, the correspondence principle said that for large quantum numbers and low frequencies the radiation energy of atomic quantum jumps is close to the classical prediction. In the early days of quantum mechanics, a generalized version of the principle helped to establish relations between classical and quantum concepts. In 1930, Heisenberg made it explicit as a general rule of using the complementary classical particle respectively wave pictures for the description of quantum phenomena here and the interpretation of quantum theory there (Heisenberg 1930: 128). In many concrete applications of quantum physics the generalized correspondence principle is tacitly still used, especially when the squared amplitude of the wave function cannot be provided with the usual operational meaning (see below). Another example is Ehrenfest's theorem, which derives from quantum mechanics by means of interpreting the expectation values of quantum observables in terms of the temporal change of the corresponding classical magnitudes.
3. The generalized Born interpretation is not warranted. This is one of the claims already made in Cartwright's 1983 paper on the measurement problem. I agree (see section 2). Expressions such as $e\,|\Psi(\boldsymbol{r})|^2$ that enter a dipole moment or the electromagnetic form factor of a complex atom have no obvious operational meaning. They are not interpreted according to the Born interpretation but rather according

to Bohr's generalized correspondence principle. Such expressions enter many semiclassical models of quantum systems. In the semiclassical model of the laser, one of Cartwright's favourite models, the radiation field is treated classically. It induces a quasi-dipole moment in the atoms of the lasing medium. The dipole moment is described as above. The semiclassical laser model is justified due to two legitimate applications of Bohr's correspondence principle. One of them assigns a classical state to the electromagnetic field, the other one assigns a classical state to the atom. In quantum optics, however, these semiclassical models are no longer valid (see section 3). In addition, I do not understand how Cartwright maintains her claim about the generalized Born interpretation in 1999, when she explicitly rejected in 1998 her former claim that all quantum probabilities are transition probabilities (Cartwright 1998: 106). In my view of quantum theory, both claims are closely related.

4. Quantum and classical states are not automatically incompatible. It is true that they are not. From a formal point of view, they are only incompatible if the quantum model of a phenomenon has no classical counterpart. In this case, the quantum description of a system or process cannot, by any means, be approximately reduced to a corresponding classical description (and vice versa), hence the quantum state of the modelled system does not correspond to a classical state, in the sense of correspondence explained in option 2. Whenever a quantum model has classical correspondence, quantum states admit at least approximately compatible mathematical descriptions at the probabilistic level. (Such classical descriptions that correspond to a quantum state give rise to many semiclassical models which are used in subatomic physics.) The interference pattern in the double-slit experiment is described by a wave function with classical correspondence. The same is true of the photon field polarization that I discussed in section 3. The correspondence breaks down, however, when you look at a low-intensity beam and ask for the passage of single particles respectively field quanta through the double slit or the crossed polarizers. The classical and the quantum state of a single electron or photon are very distinct. The propagation of the field quanta does not correspond to a classical trajectory. Bohr's complementarity philosophy disregards the typical, striking quantum phenomena, and so does Cartwright's semiclassical solution of the measurement problem. This is obvious for entangled quantum systems, e.g., the two-photon wave function of the EPR paradox or the entangled H_2^+ system mentioned in section 2, which gives rise to interference in proton-hydrogen scattering.[25]
5. There is no general rule of how classical and quantum states relate. Their relation differs from case to case. This claim is half right and half wrong. The assignment of classical states to quantum states is possible

whenever a bridge principle links them in a well-defined way. Indeed there are general bridge principles like the Ehrenfest theorem that can be derived within quantum theory. The Ehrenfest theorem enables us to relate quantum and classical states and their temporal evolutions. However, such well-explained bridge principles do not cover all cases where you find correspondence. The quasi-classical domain is larger than quantum theory tells us. The borderline between quasi-classical phenomena and genuine quantum phenomena is fuzzy; it needs more investigation in physics and more attention in the philosophy of physics. Take for example the quasi-classical statistical behaviour of scattering events in the high-energy domain of quantum electrodynamics where the Ehrenfest theorem obviously fails. Some years ago I was puzzled about it in the case of the energy loss along particle tracks due to *bremsstrahlung* (Falkenburg 1996). In between I learnt that decoherence may explain the quasi-classical statistics of quantum scattering. Modest quantum fundamentalism leads at least to some explanatory progress. The present state of the quantum art is as follows. There seem to be some general rules that govern the relations between quantum and classical states, but we do not know them completely.

6. Quantum mechanics and classical mechanics are both necessary, and neither is sufficient (Cartwright 1999: 228). This is perhaps the most Bohrian claim in Cartwright's 1999 view of quantum mechanics. It is supported as follows. (i) Quantum mechanics is necessary due to the matter of fact that we do not live in a thoroughly classical world. There are irreducible quantum phenomena such as quantum jumps, EPR correlations, and single photons interfering with themselves when sent through a double slit or a system of beam splitters. (ii) Classical mechanics is necessary due to the matter of fact that quantum theory is at odds with the actual occurrence of classical events, that is, due to the superposition and measurement problem. As both classical and quantum mechanics are necessary, and neither reduces to the other, neither is sufficient. This Bohrian claim, however, is typical of any of the specific choices within option 2. Once you have chosen to escape idealism and take the second horn of the dilemma sketched above, you have committed yourself to make this crucial step back to Bohr's views of quantum theory.

SOME QUESTIONS

Let me conclude with some questions addressed to Nancy. My "zero" question is in which regards she abandoned her 1983 view of quantum theory and what were the reasons for doing so. My additional questions concern her actual metaphysical epistemological views about measurement and quantum fundamentalism.

1. What do you think *really* goes on in a measurement? What kind of interaction is the interaction between a quantum system and the measuring device? Is it an individual physical process without individual cause? Or should we philosophers simply be agnostic about the measurement problem as far as it cannot be resolved within physical theory?
2. Why should we not adhere to methodological fundamentalism? For three centuries the quest for a fundamental theory of physics has been the most stimulating guideline of physical research. The goal of theoretical unification functions indeed as a regulative principle in a Kantian sense. It opens completely new fields of research even though the hope for a fundamental unified theory of physics may remain forever a speculative idea that will never be achieved. In addition, without foundational research many important technologies would never have been found. To give only one actual example: Today‘s research on decoherence is indispensable for the development of future quantum computers. To investigate decoherence helps to understand why it is so difficult to construct a quantum computer. Perhaps some day it will help to overcome the notorious quantum computer crash into a classical state.
3. Do you not recommend fighting the battle against fundamentalist metaphysics on its own metaphysical grounds? Bad metaphysics is best criticized by revealing its internal incoherence and missing adequacy. Your own work gives a lot of inspiration about how this might be done. Overall, it throws much light on the role of idealizations in physics. To go on in this direction, however, requires doing more epistemological work in Bohr’s or Kant’s lines of reasoning. Such a philosophical program seems old-fashioned, but pursuing it might help us better understand the limitations of physical knowledge at which your work points.

NOTES

1. Revised version, June 30, 2003. I would like to thank Maurizio Donato, Hernàn Pringe, Michael Redhead, and Paul Teller for helpful comments on an earlier version of this chapter. In between, my views of the topics of this paper have evolved substantially; see Falkenburg 2007, Chapters 5 and 7.
2. The theory received a lot of criticism; for a modern review see Guilini et al. (1996: 274). This review concerns criticism from physicists. The decoherence approach explained in Guilini et al. (1996) has in common with the Daneri, Loinger, and Prosperi theory that decoherence also predicts quasi mixtures, that is, superpositions dressed up as mixtures.
3. In 1998, Cartwright still defends the view that preparation, like measurement, leads to reduction of the wave function (105).
4. I am indebted to Michael Redhead for drawing my attention to this point.
5. See Park (1992: 286). Here, *k* is the wave number of the momentum state of the incoming particles.

6. Cartwright follows P. H. Eberhard, 'Should unitarity be tested experimentally?', CERN 72: 1, an unpublished CERN report. However, I suspect that a violation of the optical theorem would not only violate the unitary evolution of the wave function but, in addition, the fundamental dynamical conservation laws of particle physics (charge conservation, etc.).
7. According to quantum mechanics, scattering of the plane wave $\Psi_0 = e^{ikr}$ at a central potential $V(r)$ results in a scattered wave $\Psi_S = f(\theta)e^{ikr}/r$, with the scattering amplitude $f(\theta) = \Sigma_l(2l+1)\, f_l\, P_l(\cos\theta)$ where the partial scattering amplitudes $f_l = 1/(2ik)(e^{2i\delta l} - 1)$ express a phase shift δ_l of the contributions of angular momentum l of the outgoing wave relative to the incoming wave. M. Redhead suspects that in case of higher angular momentum contributions to the scattering, a realistic account of particle tracks might give rise to problems with angular momentum conservation. As the scattering center might carry away angular momentum which might dissipate into the environment, giving rise to decoherence, I cannot see this problem. However, there are clear cases of interfering momentum states which have been observed for decades in proton-hydrogen scattering (see note 17 below).
8. The cross section of proton-hydrogen scattering with charge capture measured at a fixed scattering angle $\theta = 3°$ oscillates strongly in dependence of the momentum of the proton beam. The oscillation length depends on an energy transfer ΔE from the proton to the hydrogen atom which corresponds exactly to the energy difference of the $1s\sigma$ and $2p\sigma$ states of the bound system H_2^+. The oscillation of the scattering with charge transfer is explained as follows. The charge transfer is associated with a momentum transfer, which gives rise to an excitation of the 2pσ state of the bound system H_2^+. With varying momentum of the proton beam, the excitation is turned on and off, giving rise to interference of scattering with and without charge transfer. That is, two kinds of scattering interfere, namely the processes $H^+ + H \to H + H^+$ (charge-plus-momentum transfer) and $H^+ + H \to H^+ + H$ (no charge transfer, no momentum transfer). Whenever the momentum of the proton beam is tuned off the $1s\sigma$ or $2p\sigma$ states of the H_2^+ system, the scattered wave is obviously a superposition of two distinct momentum states.
9. In nuclear and particle physics, subatomic magnitudes like the dipole moment enter the calculation of the scattering matrix. They are relational quantities. They do not describe subatomic structures per se; rather, they describe the dynamic charge structure of matter measured at a given energy scale of the probe particles in scattering experiments (Cahn & Goldhaber 1989: 217). In the domain of relativistic quantum theory, i.e. at high energies, it becomes highly problematic to attribute to a scattering center a quasi-classical charge structure in terms of an effective charge density $e|\Psi(r)|^2$; (Falkenburg 1993).
10. I have shown elsewhere that the analogy between $e|\Psi(r)|^2$ and a classical charge distribution $\rho(\boldsymbol{r})$ is based on a generalized version of Bohr's correspondence principle, which serves as a bridge principle in interpreting several quantum concepts without direct operational meaning (Falkenburg 2002). See also section 3.
11. However, Cartwright seems to have abandoned this claim (see Cartwright 1998: 106).
12. For a discussion of different single photon light sources, see Greenstein & Zajonc (1997: 32).
13. My attention was drawn to this experiment by Thomas Lohse. In the 2001 Rovinj summer school on wave-particle dualism, he demonstrated the classical experiment to let us be puzzled about the quantum case.
14. Recent *which way* experiments indicate that storage of information (which may be read out later) is sufficient for making the difference between measurement

and preparation (see Dürr & Rempe 2000: 55). However, this point needs further conceptual analysis.

15. In case (3a), you end up with the vacuum state of the quantum field. In all the other cases, inserting the polarizer results in a single photon mode in superposition with the vacuum state.
16. The photons are detected with irregular counting rate, due to the ubiquitous vacuum state which gives rise to quantum field fluctuations. For a single mode field and unit quantum measurement efficiency the probability of the photon counts is a binomial distribution (see Walls & Milburn 1995: 51; Busch et al. 1995: 9, 177).
17. See the remarks in Peres (1993: 12), plus my remark in note 31.
18. The Mach–Zehnder interferometer quantum eraser sketched here has been realized by Ou et al. (1990) and Greenstein & Zajonc (1997: 206). The double-slit quantum eraser has been proposed by Scully et al. (1991). An optical analogue has been recently confirmed in the experiment of Walburn et al. (2001).
19. The experiment is explained in Greenstein & Zajonc (1997: 206).
20. This is not only against Cartwright (1983) but also runs counter to the 'big point about reduction being needed for state preparation' (Cartwright 1998: 106).
21. Here and in the following, I assume that quantum physics is more than quantum theory. Quantum physics includes not only all existing quantum theories (quantum mechanics, quantum electrodynamics, quantum optics, quantum field theory, etc.). In addition, it embraces the practice of making quantum models and applying them in concrete experiments. It contains also many classical concepts, semiclassical models, quasi-classical approximations, etc. They are not only needed for giving concrete content to the key models of quantum theory, but also for measurement and data analysis (Falkenburg 2002).
22. See Stamatescu's contribution to (Giulini et al. 1996: 249).
23. Here I neglect the option of reduction via consciousness mentioned in (Cartwright 1999: 212).
24. See also the related concepts of approximately real properties, or unsharp properties, which is explained in (Busch et al. 1995), and used in (Mittelstaedt 1998) for explaining the limitations of object constitution in quantum mechanics.
25. See note 17.

REFERENCES

Barbour, J. (1999) *The End of Time*, Great Britain: Weidenfels & Nicholson.

Bethe, H. (1930) 'Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie', *Annalen der Physik*, 5: 325–400.

Bohr, N. (1928) 'The quantum postulate and the recent development of atomic theory', Como lecture 1927; modified version in *Nature*, 121: 580–590; both versions in *Collected Works*, 6: 109–158.

———. (1948) 'On the notions of causality and complementarity', *Dialectica*, 2: 312–318; reprinted in J. Kalckar (ed.) (1996) *Collected Works*, vol 7, Amsterdam: North Holland.

———. (1949) 'Discussion with Einstein on epistemological problems in atomic physics', in P. A. Schilpp (ed.) *Albert Einstein: Philosopher–Scientist*, Evanston, Illinois: Library of Living Philosophers; reprinted in *Niels Bohr, Atomic Physics and Human Knowledge* (1958), New York: J. Wiley & Sons; and in J. Kalckar (ed.) (1996) *Collected Works*, vol 7, Amsterdam: North Holland.

Born, M. (1926a) 'Zur Quantenmechanik der Stoßvorgänge', *Z.Physik*, 37: 863–867.

———. (1926b)'Quantenmechanik der Stoßvorgänge', *Z.Physik*, 38: 803–827.

Brune, M. et al. (1996) 'Observing the progressive decoherence of the "meter" in a quantum measurement', *Phys. Rev. Lett.*, 77, Nr. 24: 4887–4890.

Bub, J. (1968) 'The Daneri–Loinger–Prosperi Quantum Theory of Measurement', *Il Nuovo Cimento*, 57 B: 503–520.

Busch, P. et al. (1995) *Operational Quantum Physics*, Berlin Heidelberg: Springer.

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Clarendon Press.

———. (1998) *Laws, Capacities and Science Vortrag und Kolloquium in Münster 1998*, Münster: Lit Verlag.

———. (1999) *The Dappled World. A Study of the Boundaries of Science*, Cambridge: Cambridge University Press.

Daneri, A. et al. (1962) 'Quantum Theory of Measurement', *Nuclear Physics*, 33: 297–313.

Dirac, P. (1958) *The Principles of Quantum Mechanics*, 4th edn, Oxford: Clarendon Press.

Dürr, S., and R. Gerhard. (2000) 'Wave-particle duality in an atom interferometer', *Advances in Atomic, Molecular and Optical Physics*, 42: 29–71.

Falkenburg, B. (1993) 'The concept of spatial structure in microphysics', *Philosophia naturalis*, 30: 208–228.

———. (1996) 'The analysis of particle tracks. A case for trust in the unity of physics', *Stud. Hist. Phil. Phys.*, 27: 337–371.

———. (1998) 'Bohr's principles of unifying quantum disunities', *Philosophia naturalis*, 35: 95–120.

———. (2002) 'Correspondence and the non-reductive unity of physics', in C. Mataix and A. Rivadulla (eds) *Física Cuántica y Realidad—Quantum Physics and Reality*, Madrid: Editorial Complutense.

———. (2007) *Particle Metaphysics. A Critical Account of Subatomic Reality*. Berlin Heidelberg New York: Springer.

Guilini, D. et al. (1996) *Decoherence and the Appearance of a Classical World in Quantum Theory*, Berlin Heidelberg: Springer.

Greenstein, G., and A. G. Zajonc. (1997) *The Quantum Challenge. Modern Research on the Foundations of Quantum Mechanics*, Boston: Jones & Bartlett.

Heisenberg, W. (1958) *Die physikalischen Prinzipien der Quantentheorie*, Stuttgart: S. Hirzel.

Mittelstaedt, P. (1995) 'Die wechselseitigen Beziehungen zwischen der Quantentheorie und ihrer Interpretation' in L. Krüger and B. Falkenburg (eds) *Physik, Philosophie und die Einheit der Wissenschaften*, Spektrum Akademischer Verlag: Heidelberg.

———. (1997) *The Interpretation of Quantum Mechanics and the Measurement Process*, Cambridge: Cambridge University Press.

———. (1998) 'The constitution of objects in Kant's philosophy and in modern physics', in E. Castellani (ed.) *Interpreting Bodies. Classical and Quantum Objects in Modern Physics*, Princeton: Princeton University Press.

———. (2001) 'What if quantum mechanics is universally valid?', in E. Agazzi and J. Faye (eds) *The Problem of the Unity of Science*, Singapore: World Scientific.

Mott, N. F. (1929) 'The wave mechanics of α-rays', in *Proc. Roy. Soc.*, A 126: 79–84; reprinted in J. A.Wheeler and W. H. Zurek (eds) *Quantum Theory and Measurement*, Princeton: Princeton University Press.

Mott, N. F., and H. S. W. Massey. (1965) *The Theory of Atomic Collisions*, 3rd edn, Oxford: Clarendon Press.

Myatt, C. J. et al. (2000) 'Decoherence of quantum superpositions through coupling to engineered reservoirs', *Nature*, 403: 269–273.

Ou, Z. Y. et al. (1990) 'Evidence for phase memory in two-photon down conversion through entanglement with the vacuum', *Phys. Rev*, A 41: 556–558.

Peres, A. (1993) *Quantum Theory: Concepts and Methods*, Dordrecht: Kluwer.

Putnam, H. (1965) 'A philosopher looks at quantum theory', in R. G. Colodny (ed.) *Beyond the Edge of Certainy*, Englewood Cliffs: Prentice-Hall; reprinted in (1979) *Mathematics, Matter and Method, Philosophical Papers Volume 1*, Cambridge: Cambridge University Press.

Scully, M. O. et al. (1991) 'Quantum optical tests of complementarity', *Nature*, 351: 111–116.

Walburn, S.-P. et al. (2001) 'A double-slit quantum eraser', *quant-ph/0106078*, (June 13, 2001).

Reply to Brigitte Falkenburg

My paper 'How the measurement problem is an artefact of the mathematics' begins, as Brigitte Falkenburg reports, from the assumption that sometimes superpositions turn into mixtures.[1] If superpositions do turn into mixtures then some quantum evolutions must be describable by nonunitary Hamiltonians. It was typical at the time I wrote that paper to suppose that evolution described by a unitary Hamiltonian was the norm and evolution described by a nonunitary Hamiltonian was the exception that happens, if at all, only on measurement.

But then, what features does a measurement have that Nature can recognise in order to decide to evolve a system in the nonstandard way? That seems to be the only problem left about superpositions turning into mixtures once nonunitary Hamiltonians are admitted; and that, I urged, is a pseudo-problem. Different individual operators have many different mathematical features. Why should we think that Nature cares about any one or another of these? In particular why should we think that Nature cares whether the mathematical operator that represents an evolution is unitary or not, any more than she cares whether it has three zeros in it or has eigenvalues that are all prime numbers but one?

Trivially, if we can identify a physical characteristic that is captured by an aspect of our mathematical representation, then we have *ipso facto* a reason to take that aspect to represent something physically significant. But it is a mistake to go the other way around. Our mathematical representations have a huge amount of excess structure and we must be careful to avoid attributing physical significance to "mere" mathematics—and especially careful to ensure that any "predictions" we derive from a theory depend only on well-warranted physically significant features and not on excess mathematical structure. This is a view I have never relinquished and it is reflected in my current pleas for physics to produce more—vastly more—representation theorems.[2]

This said I should add three minor remarks. First, if we assume that in some cases of scattering a mixture of momentum eigenstates evolves rather than the "corresponding" superposition, then we do not lose particles, as I think Falkenburg suggests. There are just as many counts in the detectors, and at just the same scattering angles, with the mixtures as with the superposition. Second, if we have evidence, as she claims, that in some scattering situations a mixture evolves, then on those occasions we ought to represent the evolution with a nonunitary operator. The job of physics is to write down the "right" operator to represent what happens. If I am right, it may sometimes be unitary, sometimes not, and we have no reason to think that Nature cares. Third, I attempted at the time of this paper to work entirely with descriptions from within quantum theory, so at the time I certainly had no truck with trajectories or anything else that would violate the uncertainty principle. Indeed, at the same time I worked on the problem of joint distributions for position and momentum in quantum mechanics. I showed that the most natural candidate—the Wigner distribution—known not to be a probability because not nonnegative—could be turned into a proper joint probability but only by smoothing over regions in p,q space of size $\hbar^2$ (Cartwright 1976).

I also suggested in the *Artefact* paper that evolution from superposition to mixtures did not seem all that unusual. It seems to happen in many cases of decay, of scattering, and generally anywhere we would be inclined to apply the informal description "preparation". I do not see why this is "question begging". My basis for the claim is my usual one. To see what a theory "says"—at least what it says that is warranted—we look not to claims made by its advocates but to what assumptions are supposed in its empirical success. Successful treatments, I observe, do not start with the huge entangled state of the universe. They assign quantum states to the specific bits of the world, just as Falkenburg does when she talks about a single-particle beam (i.e. an eigenstate of the number operator) passing through polarizers. Nor do I understand how Falkenburg's own story of the polarizers, or the quantum erasers, contradicts this claim. The kind of polarizer experiment she describes does not seem able to tell us whether after passing, say the first two polarizers, the beam is in a mixture rather than a superposition of 'one-particle with 45° polarization' and 'no particles' states.

Does Falkenburg take the eraser experiment to show that in that setup there cannot be a mixture after interaction with a which-way detector because putting in a second which-way detector reveals interference? If so, the evolution must be modelled with a unitary operator, and something further must be added to the account to justify using a beam that passes the single which-way detector as if it had been prepared in the corresponding eigenstate.

When it comes to my more recent views that we can have our cake and eat it too—in fact empirically successful quantum mechanics always does—Falkenburg's report of the views is not really accurate. Probably it is because of her own concern to 'explain why individual events and processes happen', coupled with her assumption that they do happen as a result of measurement, that she focuses on postmeasurement situations as the location where I must be attempting to defend the joint ascription of quantum and classical descriptions. She says

> She [Cartwright] suggests dealing with the superposition problem following W. Lamb, in terms of a semi-classical approach. Her semi-classical solution consists in identifying the state before measurement with a quantum state and the state after measurement with a classical state. The solution consists in the simple rule: Whether before measurement there was a superposition, just replace it by a classical state after measurement.
>
> (Falkenburg this volume: 24)

This is decidedly not what I intended to suggest. This is far too close to what we are told when we concentrate on abstract mathematical theory and ask how we might interpret it. I urge instead that theory is what theory does. Look to see how quantum ascriptions are assigned in empirically successful treatments.

In reconstructing theory this way we should not suppose that quantum terms need, or can, receive an interpretation. In particular the rule that goes from quantum states to probabilities for events that never happen—the "Born rule"—can be supposed to hold only where it can be found to be essential in successful practice, which in my studies of how quantum theory is used to get a laser or a transistor or a SQID to work is not very often. Notice in particular that Falkenburg's account of my view does not posit the simultaneous ascription of quantum and classical states but, rather, quantum before some event and classical after. But in successful practice it seems we often need to assign both at once. In the Lamb theory of the laser, for instance, we provide a quantum description of the atom as well as treating it as a classical dipole oscillator in order for it to interact with the electromagnetic field.

The point here is that I no longer believe we have to deal with the superposition problem. I still maintain that we must assign the right Hamiltonian to account for the facts, and sometimes it will be unitary and sometimes nonunitary. But, in argument with Falkenburg, I would not suppose that that provides access to either the classical events she seems concerned with nor to any of the classical descriptions I see applied in empirically successful treatments alongside quantum descriptions.

Falkenburg tends to describe my examples of the joint use of quantum and classical descriptions as "semiclassical". Perhaps there is a suggestion that the need for classical assumptions will disappear in a "fully quantum" treatment, but nothing in my surveys of empirically successful accounts supports this. There are of course situations where in order to provide an empirically successful treatment we must assign the field a quantum state. In those cases—when they genuinely lead to significant predictions empirically borne out—I am happy to take the quantum state as a true description of the field. And if assigning the field a classical description in those cases leads to bad predictions, then that assignment is probably false. That does not show that in the first set of situations the classical description is false. Nor does it show that our successful treatments of the second set of situations will not employ classical descriptions elsewhere. I have never seen a treatment of real empirical phenomena that does that. Conversely, the use of classical descriptions as well as quantum descriptions at some points in a treatment does not show that the quantum descriptions are false. And happily so because we do need them, as Falkensburg suggests, not only for the older technologies of laser and SQIDs but also for the newer technologies of quantum computers and detectors for use in astrophysics.

It is probably important to stress that the issues of having your cake and eating it too, that is having quantum theory and classical as well, are orthogonal to issues of unitarity and reduction of the wave packet. I agree with Falkenburg that nothing *in* quantum mechanics—including non-unitary evolutions or reductions to new eigenstates—allows the ascription of classical descriptions. And we both think that we had better be able to ascribe classical descriptions sometimes. I take that Falkenburg does so because she would like to see some "real events" and not just evolutions of quantum states. I do so because I think we are warranted in doing so and warranted by the only means we can gain warrant for scientific ascription—that are warranted by the empirical success of treatments using these ascriptions.

NOTES

1. There cannot be nonquestion begging "rigorous proofs" in quantum mechanics that show that this cannot happen, so this certainly is not what Peter Mittelstaedt has shown.
2. *Cf.* my forthcoming 'In Praise of Representation Theorems', in P .Suppes (ed.).

REFERENCES

Cartwright, N. (1976) 'A non-negative Wigner-type distribution', *Physica*, 83A.

16 Getting the Causal Story Right

Hermeneutic Moments in Nancy Cartwright's Philosophy of Science

Alfred Nordmann

For all I know, the term "hermeneutics" appears nowhere in Nancy Cartwright's books and articles. Any attempt to appreciate hermeneutic moments in her work therefore requires special justification.[1]

Even though there is by now a long tradition of studies and reflections on the hermeneutics of science, it has not been able to dispel serious reservations about the transfer of a textual, if not literary mode of analysis to the domain of science and nature. First, even though it has been acknowledged that in scientific experience we do not encounter things in themselves but something that is structured by conceptual, instrumental, and sensory modalities, reality is not therefore inert and fabricated like a text. Second, the hermeneutic process is said to consist in the integration of a text within a horizon of meaning, and as this integration is never seamless it requires adjustments such that the reader of the text emerges as a different person (Gadamer 1975; Ricoeur 1981). This presupposes an individualistic conception that is hardly suitable for the collective work of science. Third, though one can say that scientific data require interpretation, this kind of "interpretation" is surely much more constrained than, say, the interpretation of a literary work.[2] Fourth, while in the paradigmatic case of literature hermeneutics generally refers to the relation between reader and text, the hermeneutics of science follows Kuhn in that it is less interested in the reader of a scientific text and rather more in the scientific community as a community of interpreters that reads nature in a certain way. The hermeneutics of science thus appears stuck between a rock and a hard place: It needs to either consider nature as a text and encounter the first objection above, or it must account for the curious fact that scientific texts defy hermeneutics in that they do not require exegesis but disclose themselves immediately. Indeed, it is a hallmark of membership in a scientific community that the texts of one's peers can be taken literally and are rarely subject to interpretation. Science and nature and scientific texts and their readers have thus appeared to be the moving targets of hermeneutic equivocation. Fifth and finally, the hermeneutic process is said to lead into a hermeneutic circle according to which there is no outside to the activity of interpretation. Bas van Fraassen elaborated how the scientific enterprise moves within such a hermeneutic circle: Because the empirical content of a theory is specified by the theory itself, theories can only save their phenomena and have no further-reaching claim to truth (van Fraassen 1980: 56–9; see Cartwright 1983: 88). Nancy Cartwright's work, however, is an attempt to meet van Fraassen's challenge and to show a way out of the circle at least for causal explanations.

> Van Fraassen [. . .] offers more of a challenge than an argument: show exactly what about the explanatory relationship tends to guarantee that if *x* explains *y* and *y* is true, then *x* should be true as well. This challenge has an answer in the case of *causal* explanation but *only* in the case of causal explanation. [. . .] In causal explanations truth is essential to explanatory success.
>
> (Cartwright 1983: 4, 10, see 89–99, 159)

For the most part the scientific enterprise may well be caught up in van Fraassen's nonvicious hermeneutic circle. However, we should not underestimate 'the very special case of causal explanation' (Cartwright 1983: 10). Empiricists like van Fraassen have tended to discount it; realists take it to be the paradigm of successful ordinary science. Cartwright seeks a middle ground: The very special case of causal explanation can teach us about the work that is required for scientists to achieve this peculiar kind of success. By showing that causal explanation results from a felicitous alignment of phenomena, models, and theories, she introduces her readers to the toolbox and resources of science. Cartwright thereby presents scientific work as a hermeneutic process of sorts and, along the way, counters the various objections to the very idea of a hermeneutics of science.

MIDDLE GROUND

In other respects, too, Cartwright locates her own position between that of various received views. Only a few instances of this need to be mentioned here. They help define *ex negativo* where Cartwright stands, allowing us to then appreciate the centrality of the hermeneutic moments in each of her three main works.

Cartwright explicitly claims for herself a 'middle ground in the dispute' between realist and constructivist accounts of the success of science (Cartwright 1999: 47). According to the first of these, science 'reveals [. . .] directly the language in which the Book of Nature is written' (Cartwright 1999: 46). This direct revelation issues in statements that are straightforwardly true or false, that can therefore be taken literally and require no mediation by a hermeneutic process of interpretation or negotiation. According to the second account, the success of science is trivial in that one cannot first construct a world and then act surprised that certain constitutive principles apply to it. This constructivism posits a realm of human practice that is hermetically self-enclosed and is not measured against anything outside it.[3] In contrast, Cartwright emphasizes that science requires work, that is, practical human engagement with a world of immensely varied concrete situations. Whether scientific work succeeds is no matter of simple procedure, methodology, or routine. The success of science consists in the establishment of a more or less local, more or less robust alignment of phenomena, models, and theories. Indeed, as will be shown below, this success coincides with the achievement of literalness: Once everything fits together, the hermeneutic mediations of scientific work give way to straightforward truth or falsity.

Cartwright also seeks a middle ground regarding "modalization" (Cartwright 1989: 158–170). She is sympathetic to empiricist attempts to "modalize away" causal laws, that is, to refer to the formal mode of mere linguistic representation what, as a manner of speaking, is misleadingly cast in the material mode. According to Cartwright, laws 'are generally pieces of science fiction, and where they do exist they are usually the objects of human construction, objects to be explained, and not ones to serve as the source of explanation' (Cartwright 1989: 218, see 229). Cartwright is also sympathetic, however, to the attempts by scientific realists to distinguish causal laws from merely accidental generalizations (Cartwright 1989: 7, 36, 131–136). Here, Cartwright claims as middle ground that one cannot modalize away capacities and their power to productively bring things about (i.e. singular causation). Those who wish to distinguish between laws and generalizations are onto something, namely capacities, even as they are wrong about causation, truth, explanation, and law.[4] They tend to be confused, in particular, about the relation between the formal mode of theoretical representation and the material exhibition of the capacity in the model. While they think of this relation as one of inclusion, Cartwright argues against the notion that the materially concrete is an instance of something like a general fact. In her view, properties like "being subject to a force" or "doing work" do not exist in the abstract but can exist only when, by way of models, they are referred to concrete situations like "being located at some distance to a charge" or "washing dishes" (Cartwright 1999: 40–46).

Cartwright finally detaches models both from phenomena and from theory in the sense that there are no determinative relations among them. On the one hand, this opens an indeterminate space for a wide-range of models (and this, in turn, has prompted wide-ranging discussions): She considers phenomenological or representative as well as theoretical or interpretive models; she allows for experimental situations, schematic and block diagrams, equations, conceptualizations, and simulations to serve as models. Models can have various degrees of idealization and abstraction, and some models are models of models. If there is a significant shared feature of interest in Cartwright's discussion of models, it is that they can exhibit the causal structure in which capacities come alive and manifest their productivity (Cartwright 1989: 223). Models figure prominently in the story of how one moves from phenomena all the way up to theory, and equally prominently in top-down accounts that take us from theories to the phenomena. While they stand at an intersection of the roads that lead from phenomena to theory and from theory to the phenomena, there are no antecedent guarantees that they will successfully coordinate theory and phenomena. Indeed, phenomenological or representative models may fail to concretize or realize theoretical concepts, and it may require a rather tenuous process to relate theoretical or interpretive models to the phenomena (compare Cartwright et al. 1995). However, it is also possible for phenomenological and theoretical models to be aligned or even to coincide. In those instances, it becomes possible for scientists to routinely traverse in both directions between the abstract and the concrete. Cartwright rejects any philosophy of science that takes those cases as its paradigms and thereby ignores the work that is involved in relating phenomena, models, and theories to one another (Cartwright 1983: 17, 162; 1999: 43, 47). At the same time, whenever Cartwright considers in her own terms the movements back and forth between the abstract and the concrete, she arrives at what I here call "hermeneutic moments". At these moments, the models are the stage on which the negotiations take place and on which the top-down and bottom-up approaches become calibrated to each other. Moreover, her hermeneutic characterizations treat the model not only as the site at which those negotiations converge, but in an interesting sense they turn the model into a protagonist of sorts, namely into a device that interprets, measures, or reads phenomena and theory and that promotes the attunement of concrete and abstract properties.

MIXED METHOD: HOW TO READ MARX, SCHRÖDINGER (AND MILL)

In *Nature's Capacities and their Measurement*, Nancy Cartwright endorses not only Mill's discussion of tendencies but, along with it, his mixed method and its proposed middle road between inductivism and hypothetico-deductivism. To the extent that both methodologies take laws to be exceptionless statements about what things regularly do, neither does justice to her and Mill's view that laws are about the tendencies or capacities of things even where these are manifested only in highly irregular circumstances. To show how one arrives at knowledge of these capacities, Cartwright quotes the following passage in which Mill contrasts the inductivism of the so-called "practicals" with the mixed method that is adopted by the "theorists." As Cartwright emphasises, Mills' theorist does not conjecture a theory in order to deduce a testable prediction. Instead, he draws on his knowledge of capacities and extrapolates from this knowledge:

> Suppose, for example, that the question were, whether absolute kings were likely to employ the powers of governments for the welfare of or

> for the oppression of their subjects. The practicals would endeavour to determine this question by a direct induction from the conduct of particular despotic monarchs, as testified by history. The theorists would refer the question to be decided by the test not solely of our experience of kings, but of our experience of men. They would contend that an observation of the tendencies which nature has manifested in the variety of situations in which human beings have been placed, and especially observations of what passes in our own minds, warrants us inferring that a human being in the situation of a despotic king will make a bad use of power; and that this conclusion would lose nothing of its certainty even if absolute kings had never existed or if history furnished us with no information of the manner in which they had conducted themselves. (Cartwright 1989: 171)[5]

The theorist's mixed method here refers on the one hand to the experience of men and thus to knowledge of our tendency to exercise power over others—a kind of self-knowledge—and on the other hand it refers to the experience of kings by way of the conjecture that kings are men like other men. Together, introspective acquaintance with a tendency and the deductive consequence of a hypothetical generalization yield the conclusion about the despotic king's abuse of power. This is how Nancy Cartwright goes on to generalize Mill's example:

> [O]ne looks for what is true in an ideal model in order to establish an abstract law. But there is a difference between what is true in the model and the abstract law itself. For the ideal model does not separate the factors under study from reality but rather sets them into a concrete situation. The situation may be counterfactual; still it is realistic in one sense: all the other relevant factors appear as well, so that an actual effect can be calculated. What is ideal about the model is that these factors are assigned especially convenient values to make the calculation easy. (Cartwright 1989: 191)

By the observations that pass in his own mind, Mill's theorist has experience of how men use their powers generically in regard to the factor of social standing, that is, for any situation where someone has power over another. An ideal model represents such an experience. From the truth contained in the model one may then advance to an abstract law which states something about the uses or abuses of power over others, and this law does not need to refer to social standing at all. The ideal model thus sets the factor of social standing in a way that "makes the calculation easy"; one can concretize it by adding the factors back in and assigning them more definite values, for example by considering the case of an absolute king.

Idealizations thus remain realistic in the sense that, in principle at least, they afford a way back to the phenomena. All one can ever do in a concretization, however, is to give definite value ("absolute monarch") to a factor that remained generic in the idealized model ("power over others"): One can fill in causal structure (Cartwright 1989: 223). However, one cannot undo the abstraction from the materiality of all situations that takes place in the abstraction to laws and in models of theories: In a concretization from theory, "theory gives out" sooner or later (Cartwright 1989: 211, see 207, 226). As distinct from idealization, abstraction from factors in the material world is no longer realistic in that it subtracts the factors altogether, rather than merely assigns them an idealized, convenient value in a counterfactual, yet concrete situation. Indeed, the terms of a theory implicitly provide a list of all those factors that can be concretized. As this list always abbreviates the total number of factors involved in any concrete situation, 'this kind of process will never result in an even approximately correct description of any concrete thing. For the end-point of theory-licensed concretization is always a statement true just in a model' (Cartwright 1989: 207; compare Suárez 1999: 180–182).[6]

In the case of Mill's example, any materially concrete historical situation contains more than what is contained in our historical experience of kings and in the experience of our tendencies in exercising power over others. In particular, it may contain factors that counteract our tendencies in the exercise of power. Even if it is true that a certain social structure which concentrates power in an absolute monarch produces a lack of social justice, this truth does not serve to describe the concrete situation of a religious state or of the enlightened despot who lets fairness rule by his grace or whim: The truth doesn't explain much.

Where, now, lies the hermeneutic moment in this negotiation of the abstract and the concrete by way of Mill's mixed method? A first clue is provided by Cartwright's reliance in her account of abstraction and concretization on Leszek Nowak's *The Structure of Idealization: Towards a Systematic Interpretation of the Marxian Idea of Science*, i.e. a hermeneutic exercise par excellence.[7]

> Nowak's story involves the obvious idea that one must add corrections and additions as one moves from the abstract to the concrete. It is critical to the account that these corrections should not be ad hoc addenda just to get the final results to come out right: they must be appropriately motivated. I take it that means they must genuinely describe other causes, interferences, impediments, and the like. But it follows from that that the scheme can only work if we are already in control of a rich set of non-Humean, capacity-related concepts. (Cartwright 1989: 202, see 206)

In the case of Marx's *Das Kapital*, this requires an interpretive reconstruction such that 'a more detailed account of the nature of the corrective factors vis-à-vis the principal ones can be given'. Once this account is obtained, 'it is Marx's theory that tells what kinds of factor have been eliminated in arriving at the law of value' and what sequence of corrective steps will take us closer to a concrete historical situation until 'theoretical corrections run out and the process must be carried on case by case' (Cartwright 1989: 206, 209):

> The same is true for quantum mechanics. The Hamiltonian for a "real" hydrogen atom is supposed to be arrived at by adding correction terms to the ideal Hamiltonian, where the correction terms come from the theory itself, from the list of other acceptable Hamiltonians. (Cartwright 1989: 207, see 205)

Cartwright and Nowak thus show that this widely, perhaps standardly used scientific method gives rise as a matter of course to the perceived explanatory weakness of Marx's economic theory in particular and of social or political science in general (Cartwright 1989: 204). In terms of explanatory weakness or strength, quantum mechanics fares no better than Marx's theory: In both cases, if we want to move from abstract theory toward concrete phenomena, the theory has to be read or interpreted such that it tells us not only what is true in certain idealized circumstances but also what factors have been eliminated by it.

Here, the reader or interpreter need not and perhaps should not be an individual scientist who subjects the theory to some sort of exegesis. Nor is it an abstract entity like the scientific community as a whole or "science itself" that provides such a reading of the theory. Instead, just as abstract properties exist only in models, the abstract scientific reader and interpreter of theories also exists in the model.[8]

It holds for both, Marx's *Das Kapital* and the Schrödinger equations, that 'it needs to be made clear that in this or that concrete situation the designated factors are indeed correctives or preventatives, as required for the reconstruction, and also why that is true' (Cartwright 1989: 206). However, only in the first case someone like Leszek Nowak is required to interpretively tease apart the corrective factors vis-à-vis the principal ones. In the case of Schrödinger's equations, the principal factors are identified by the theory itself and the ongoing work of quantum physics adds to the list of other acceptable Hamiltonians that can serve the corrective purposes. The subjective or personalized reader thus drops out in quantum physics. Similarly, the place that was occupied in Mill's account by the theorist himself is taken over in Cartwright's account by the model. Mill's theorist begins with introspective self-knowledge of his tendencies in regard to the exercise of power over others.[9] In science more generally, this knowledge of tendencies is exteriorized and instead of a person, the model provides the causal structure in which tendencies manifest themselves, capacities do their work, and abstract properties come alive.

Although this idea of "the model as reader" needs to be substantiated further, it is already apparent how it addresses a central problem for any hermeneutics of science.[10] The hermeneutic process is typically said to involve the disclosure of self and world in the interpretive encounter of a reader with a text. As Gyorgy Markus pointed out, the place of the reader in this encounter has traditionally been occupied by a solitary, individual nineteenth-century subject confronted by a poetic text (Markus 1987). Hermeneutics has thus been ill-equipped to acknowledge the depersonalized knowing subject of science. To the extent that Nancy Cartwright's nonsubjective models can take the place of the personalized reader, she has met one of our initial objections against hermeneutic approaches to science.[11]

Cartwright's account of the abstract and the concrete addresses another major difficulty of any hermeneutics of science, namely the problem of literalness. It always appeared to be a hallmark of successful science that it requires no interpretation but issues statements that are straightforwardly true or false.[12] To the extent, however, that hermeneutics denies literalness, it will have to explain how the appearance of a transparency of meaning can come about in the case of science. Nancy Cartwright offers such an explanation by showing literalness to be a specific accomplishment of science. Accordingly, just as she rejects any philosophy of science that takes successful causal explanation to be its paradigm rather than an important special case, she rejects any approach that presupposes rather than explains literalness. Instead of taking literalness for granted, Cartwright begins by pointing out that theoretical laws cannot be literally true because their very purpose is to consider causal processes in isolation (Cartwright 1983: 12). An abstract law about the use and abuse of power might not refer to social standing, let alone the mitigating or aggravating influence of prevailing religious sentiments. The abstract properties it identifies have no given literal referent but will only exist in a model which provides at least an idealized situation such as our own tendencies when we imagine to have power over others. Only once the antecedent of the abstract law has thus been filled in with the relevant detail, one gains a concrete law 'that can be read as literally true or false in the most straightforward sense' (Cartwright 1989: 199). This concrete law assigns phenomenal content to the abstract law, and only the collection of all such concretizations would provide the more or less homogeneous phenomenal content of the abstract law in its entirety.[13] As Cartwright's discussion of Nowak has shown, concretizations stop short of the phenomena and cannot fully undo the material abstraction of the theoretical law. Therefore it is the particular concretized laws that "lie" about concrete situations such as that of the Enlightened despot (compare Cartwright 1989: 199–212). In contrast, the abstract laws cannot be literally true because, taken by themselves, they have no literal meanings that could be judged true or false.[14] Since only their interpretation by an ideal model creates empirical truth-conditions, the interpretive model constitutes meaning and Nancy Cartwright has described a hermeneutic process that yields literalness.[15]

To be sure, just like "the model as reader," this "hermeneutic construction of literalness" requires further substantiation. The hermeneutic moments of *How the Laws of Physics Lie* and *The Dappled World* provide this.

FITTING TO: PREPARED DESCRIPTIONS IN THE THEATRE OF PHYSICS

As we have seen, Cartwright argues that the process of concretization can rarely be completed "once theory runs out". In *Nature's Capacities and their Measurement* she suggests that this is the point where science stops and only engineers can bridge the remaining gap between concretized models and real-life situations (Cartwright 1989: 211). However, both her earlier and her later work have more to say on how to bridge that gap, namely from the bottom up in *How the Laws of Physics Lie*, and from the top down in *The Dappled World*. Both invoke the metaphor of "fitting"—fitting facts *to* theory and fitting *out* theories by dressing them up as statements of fact.

According to Ludwig Wittgenstein, causal accounts establish a fit between causes and effects, and what he had in mind is a kind of mechanical "fit": The machinery of causal interpretation must not idle and will only work if its various parts engage properly (Wittgenstein 1993). Cartwright, in contrast, takes a realistic view of causation: Capacities productively bring things about and need not be fitted to effects. Accordingly, she does not restrict herself to a mechanical notion of "fit" when she requires that many levels of description need to be fitted together in order for abstract laws, models, and materially concrete situations to work together in a scientific explanation. Instead, in Cartwright's case the judgement of proper fit concerns appropriateness and how the different levels are attuned to one another. Beyond that, however, there are important difference between "fitting to" and "fitting out."[16] While "fitting out" is the subject of the next section, facts are "fitted to" theory by being prepared properly:

> At the first stage of theory entry we prepare the description: we present the phenomenon in a way that will bring it into the theory. The most apparent need is to write down a description to which the theory matches an equation. But to solve the equations we will have to know what boundary conditions can be used, what approximation procedures are valid, and the like. So the prepared descriptions must give information that specifies these as well. [. . .] The first stage of theory entry is informal. There may be better and worse attempts and a good deal of practical wisdom helps, but no principles of the theory tell us how we are to prepare the description. We do not look to a bridge principle to tell us what is the right way to take the facts from our antecedent, unprepared description, and to express them in a way that will meet the mathematical needs of the theory. The check on correctness at this stage is not how

> well we have represented in the theory the facts we know outside the theory, but only how successful the ultimate mathematical treatment will be. (Cartwright 1983: 133–134)

There is a certain ambiguity in this passage regarding the notion of "prepared description", an ambiguity that will go away once "fitting out" is considered along with "fitting to". By extending the talk of "prepared descriptions" beyond the realm of quantum mechanics to science more generally, Cartwright invites the analogy to the preparation of a sample, say in microscopy. Now, this is preparation for scientific observation and eventually for theoretical treatment, and surely it is always constrained by disciplinary or theoretical interests (compare Cartwright 1989: 209). However, to prepare a sample in microscopy is not necessarily a preparation for a particular theoretical, let alone mathematical, treatment. The sample is not normally prepared specifically for the theory which is expected to deliver the explanation of the phenomena. In the preparation of samples, "theory" enters only in a generic fashion, it sets the parameters of the stage which the prepared description enters as an actor and on which it will eventually become a well-defined character.

> Imagine that we want to stage a given historical episode. We are primarily interested in teaching a moral about the motives and behaviour of the participants. But we would also like the drama to be as realistic as possible. In general we will not be able simply to "rerun" the episode over again, but this time on the stage. The original episode would have to have a remarkable unity of time and space to make that possible. There are plenty of other constraints as well. These will force us to make first one distortion, then another to compensate. Here is a trivial example. Imagine that two of the participants had a secret conversation in the corner of the room. If the actors whisper together, the audience will not be able to hear them. So the other characters must be moved off the stage, and then back on again. But in reality everyone stayed in the same place throughout. [. . .] We cannot replicate what the characters actually said and did. Nor is it essential that we do so. We need only adhere "as closely as possible to the general sense of what was actually said".
>
> Physics is like that. It is important that the models we construct allow us to draw the right conclusions about the behaviour of the phenomena and their causes. But it is not essential that the models accurately describe everything that actually happens; and in general it will not be possible for them to do so, and for much the same reasons. The requirements of the theory constrain what can be literally represented. This does not mean that the right lessons cannot be drawn. Adjustments are made where literal correctness does not matter very much in order to

> get the correct effects where we want them; and very often, as in the staging example, one distortion is put right by another. That is why it often seems misleading to say that a particular aspect of a model is false to reality: given the other constraints that is just the way to restore the representation. (Cartwright 1983: 140)

"The requirements of the theory constrain what can be *literally* represented", and indeed, nonliteralness increases representational salience as one tries to remain "as realistic as possible". By transforming the phenomena into actors fit for a morality tale, physicists create a setting in which new conditions for literalness are set. The model is the stage where the productivity of capacities can be witnessed. It thus institutes a hypothetical "as if" condition where the "as if" does not signify fictitiousness but uses theory to create a perfectly real situation that is counterfactual only in respect to the ordinary course of natural events.

When Cartwright thus discusses 'Physics as theatre' (Cartwright 1983: 139), she refers to the theatre specifically in order to distinguish the "as if" of the novel from the "as if" of the stage where perfectly real events unfold in space and time. The difference between these two uses (and placements) of the "as if" operator makes for different hermeneutic processes.[17] A historical novel may refer to real agents and say of them that they behaved *as if* they were in rage, that is, it may treat their mental states as if these were accessible to us. At the same time, the meaning of the novel can be recovered only by means of interpretation (what is literally true of the novel is limited to the appearance of signs on the page). In contrast, the performance in the theatre of a historical play puts the "as if"-operator "all the way up front" (see Cartwright 1983: 129). A person appears on stage *as if* he were someone who acts in rage. Here, it may well be literally true that the stage action is a manifestation of rage.

This difference between novel or script on the one hand and the theatre on the other is due to the fact that the theatre is already a reader of the script (compare Nordmann 1996). The performance renders the text of the play as a score for the public exhibition of certain movements and events. Similarly, the model takes the theory as an occasion to exhibit certain physical occurrences. Performance and model are thus impersonal readers of a text (the script, abstract theory) by creating representationally salient (though not descriptively true) conditions of literalness: Theories cannot be literally true about the phenomena, but they can be true and false in the models, that is, in the setting in which these phenomena are prepared for the stage like actors. While this once again presents the "model as reader" and takes the hermeneutic situation of the theatre to exemplify the "construction of literalness," theatre and model are not just readers of script and theory but also of the world. Indeed, they mediate between the abstract and the concrete precisely in that, as readers of both, they establish their commensurability.

Theatre and model are readers of the world in the sense that they mobilize or prepare phenomena for the performance, that is, by making them speak. The model turns phenomena into stage-actors by giving them a setting in which they can perform and become eloquent. In *Nature's Capacities and their Measurement* Cartwright describes this setting as the causal structure which renders capacities salient. This setting is hermetic in that, like a text or theatrical performance, it offers no way out and is no longer transparent to the conditions of its creation. Just as the theatrical performance by an actress allows no direct inference to her private character, one cannot recover raw data from prepared descriptions or a biological cell from a slide. Knowing how samples or descriptions are prepared may give us some tools for reconstructing the original phenomenon but this reconstruction will remain speculative or must draw on circumstantial evidence in order to subtract the various effects of preparation.[18]

Like a theatrical performance, therefore, the model has more reality than what it ostensibly refers to—the reality "behind" it is just as derivative as the laws that are prompted by or extracted from it (compare Morrison 1999).

FITTING OUT: FABLES AND MODELS

In *The Dappled World* Cartwright shows that the model does more than fit the phenomena to a causal structure such that their capacities can perform and bring things about. The model also assimilates theory into its setting. Only in this setting, she argues, does the theory or do things like "force" concretely exist. In doing so, the model effects a further transformation. After the phenomenon has been prepared to act in the causal structure provided by the model, the phenomenon-qua-stage-actor now becomes a character in a play. After all, for the purposes of teaching a moral it is not enough that the phenomena are fitted to the task of displaying their capacities. The actors also have to be fitted out such that they are sufficiently stereotyped characters to convey the moral of a fable.

This final hermeneutic moment draws on the work of the eighteenth-century playwright, critic, and philosopher Gotthold Ephraim Lessing and his 1759 *Abhandlungen über die Fabel* (Lessing 1854; see Cartwright 1999: 37–44). Lessing is best known in the theory of the arts for his essay *Laocoön: On the Boundaries of Poetry and Painting*. By determining these boundaries, Lessing shows what is suitable for each medium of representation. For example, while the expression of Laocoön's pain is suitable to poetry and any art-form that develops its subject in time, only a sublimated attitude of suffering is suitable to sculpture and any art-form that freezes a moment for all time. Consequently, whether or not the historical Laocoön really wailed in anguish or suffered his pain with stoic nobility cannot be inferred from its representations in poetry and sculpture. Since these representations may sacrifice descriptive accuracy for the sake of realism, an inference from representation to concrete historical situation would confuse formal and material modes (Lessing 1962).

Similarly, Lessing's treatments of the fable determine its constitutive boundaries in contrast primarily to allegory (see Cartwright 1999: 39): The fable's moral is not disguised or expressed by the fable, nor is the moral inferred from the similarity of concrete character in the animal or human world and certain abstract properties like strength of weakness. The grouse in the fable of grouse, marten, fox, and wolf is not merely similar to the weakest but *is* the weakest. Accordingly, the fable provides a story that instantiates the moral: The moral is couched in the story or the story fits out (*einkleiden*) the moral (Lessing 1854: 243, 255; see Cartwright 1999: 39). Of course, the grouse *is* the weakest only in the concrete situation provided by an ideal model, namely a situation that brings together only wolf, fox, marten, and grouse. And yet, though this situation provides a concrete instance of what it means to be the weakest, the meaning of weakness as an abstract property can be articulated also on the level of theory, for example by saying that the weaker are always prey to the stronger.

In her *Study of the Boundaries of Science*, Cartwright considers the relation between theoretical law and concrete model in Lessing's terms.[19] The notions of "work" or "force" do not exist on the level of theory, but theory can articulate the meaning of these terms, for example by opposing work and leisure on the one hand and by associating force with acceleration and mass on the other (Cartwright 1989: 40). To the extent that these are linguistic representations, it would be a categorical mistake to speak of the action in a model as being similar or dissimilar to the relation of terms in a theory[20]: Meaning is produced differently in the formal mode of theory (e.g., by way of definition or location in an axiomatic structure) and the material mode of the model (e.g., by instantiation, preparation, or mediation).

> Turn now from the Gascon and the fox to the stereotypical characters of the models which "fit out" the laws of physics. Consider F = ma. I claim this is an abstract truth relative to claims about positions, motions, masses and extensions, in the same way that Lessing's moral "The weaker are always prey to the stronger" is abstract relative to the more concrete descriptions which fit it out. To be subject to a force of a certain size, say F, is an abstract property, like being weaker than. Newton's law tells that whatever has this property has another, namely having a mass and an acceleration which, when multiplied together, give the already mentioned numerical value, F. That is like claiming that whoever is weaker will also be prey to the stronger.
>
> In the fable Lessing proposes, the grouse is the stereotypical character exhibiting weakness; the wolf, exhibiting strength. According to Lessing we use animals like the grouse and the wolf because their characters are so well known. We only need to say their names to bring to mind

> what general features they have—boastfulness, weakness, stubbornness, pride, or the like. In physics it is more difficult. It is not generally well known what the stereotypical situations are in which various functional forms of the force are exhibited. That is what the working physicist has to figure out, and what the aspiring physicist has to learn. (Cartwright 1999: 43)

Cartwright's and Lessing's dappled world is a product of work that is performed in a piecemeal fashion by exhibiting capacities in models, by rendering particular models as stereotypical situations that can teach a general lesson, and by sometimes managing to do both at once. Just as poetry and sculpture set their own rules of representation for the achievement of realism and thus claim Laocoon's suffering differently, each scientific discipline will constitute its domain by seeing what phenomena it can claim in the terms of its theories (compare Cartwright 1983: 13, and Cartwright 1989: 209). The success of science therefore cannot consist in the reduction of complexity or the unification of domains. Instead, it owes to the rightness or appropriate fit of particular causal accounts. If we are interested in descriptive adequacy, Cartwright argues, we are better off not caring 'about the tidy organization of phenomena'. Instead, we should be interested in how scientists are 'getting the causal story right. This interest 'is new for philosophers of science' (Cartwright 1983: 160, 162), as analytic philosophers have traditionally distinguished the goodness of stories from the rightness of knowledge. By asking what it takes to get a story right and thus to successfully mediate in particular cases the formal relations among abstract concepts and causal processes in the world, Cartwright confronts 'Physics as theatre' (Cartwright 1983: 139–142), the reconstruction of *Das Kapital* and Schrödinger's equations in terms of 'Abstraction and concretization' (Cartwright 1989: 202–212), and 'Fables and models' (Cartwright 1999: 35–48).

CONCLUSION

This survey of the three hermeneutic moments in Cartwright's books prompts again the opening question of how it can be justified to treat Cartwright's contribution in the terms of hermeneutics at all. It can only be part of the answer that this treatment afforded a reconstruction in Cartwright's own terms of her approach as a whole and that it thereby helped to clarify the broad outlines of this approach. As opposed to traditional philosophy of science, she does not provide formal reconstructions of causal stories but asks just what it takes to get the causal story right in the first place. Instead of taking as her paradigm of science just those cases where scientists traverse easily and successfully between abstract theories and concrete phenomena, she shows how these are the very special cases that are the hardest to understand. Similarly, she presupposes neither the impersonal knowing subject of science nor the literalness of scientific language but shows how these are constituted only as phenomena are fitted to models and theories fitted out by models.[21]

This clarification of her project also indicates what further work may need to be done. In particular, to the extent that Cartwright helps undermine the notion that phenomena are constituted by theories or paradigms, the preparation of phenomena for scientific or disciplinary treatment ought to be distinguishable from their calibration to a particular theory with its particular formalism. In the "physics-as-theatre"-analogy this is the difference between training concrete individuals to become actors on stage and then fitting them out as stereotypical characters that can convey a moral. This distinction might be clear enough conceptually or programmatically, but it remains to be seen whether it can be used to tease apart what has become amalgamated at least since the time of Kuhn, namely the demands of a discipline and the demands of a central theory.

A second critical opportunity arises from Cartwright's reticence to distinguish between physically instantiated models (e.g., experiments) and conventionally formalized models (e.g., schematic and block diagrams). It needs to be shown how a block diagram, too, can provide the causal structure in which capacities can bring things about and in which we can see, for example, what lasers tend to do or how they tend to behave (Cartwright 1989: 226). This project gets help from two very different corners. On the one hand, it can be advanced by attention to the intermediate case of simulations in which schematizations take the place of experiments. On the other hand, one can now draw on hermeneutic conceptions of a "text": When Paul Ricoeur, for example, considers actions as a text, he does not take texts to be inert but appreciates their power to bring things about, and in particular to bring about a changed alignment of self and world (Ricoeur 1981). Even our ordinary language can be more and less finely attuned to concrete situations and the resulting, more or less conventional, verbalizations can afford or resist a seamless integration into a larger horizon of expectation and meaning. Just like Mill's theorist we can learn about causal capacities from the stories we tend to tell ourselves and not just from experiments. How we learn this, in each case, requires more detailed study.

These are the various heuristic benefits of taking Cartwright to suggest that scientific modeling corresponds to a hermeneutic process, and the approach can be justified further: In the course of reconstructing this process, the five initial objections toward any hermeneutics of science have become insubstantial. As for the first objection that the object of scientific inquiry surely must not be likened to a text, matters are obviously not as simple as that. Whether nature can be considered as a text depends on whether texts are thought to be inert and fabricated in the first place. In Cartwright's case, however, and in regard to the general discussion of the mediations in modeling, it is not so clear that "nature" is the immediate object of scientific inquiry at all. Instead, the model takes the place of the phenomenon—its reading of the world provides the text for the general lesson that is to be developed.

As we have seen, Cartwright meets the second objection regarding the individualism inherent in the relation of reader and text by having the model stand in for the impersonal knowing subject of science. If the hermeneutic process consists in the integration of a text into a horizon of meaning, and if this integration requires a new alignment of reader, text, and world and thus changes how the reader relates meaningfully to the world, this process is now transposed into the model as the site at which these mediations take place. And in the literary as well as scientific case, "interpretation" is neither more nor less than attending to concrete situations and abstract concepts and fitting them to one another. This fairly inconspicuous empiricist notion of interpretation meets the third objection according to which the term should have a fundamentally different use in the contexts of science and literature.

The fourth objection maintained that by equivocating between the book of nature and the texts that are produced by scientists, hermeneutics fails to address the appearance of literalness and thus the decidedly antihermeneutic self-presentation of science. Nancy Cartwright shoulders this explanatory demand by showing how literalness emerges from a hermeneutic process.[22]

This leaves the final and perhaps most difficult question, namely whether scientific inquiry leads into a hermeneutic circle. Again, any answer depends on what, precisely, this notion is taken to mean. In van Fraassen's account, the hermeneutic circle appears nonviciously in the context of justification. Perfectly capable of absorbing into it our experience of an outside physical world, the circle merely indicates that observational content cannot be specified independently of theory and that the truth of a theory cannot be claimed on top of its ability to save the phenomena (van Fraassen 1980: Ch. 3, 5). As we have seen, Cartwright contradicts van Fraassen for the special case of causal explanation. It is here, perhaps, where her distinction between phenomenological or representative models and theoretical or interpretative models is most significant. As we have seen, each type of model constitutes a hermeneutic process of its own (fitting to and fitting out), and it is entirely nontrivial how these are fitted together in order to allow for scientists to traverse by way of these models back and forth between concrete situations and abstract theories. This suggests that Cartwright breaks out of the hermeneutic circle by positing various circles of interpretation that are in some measure external to each other.[23]

While appreciating central hermeneutic moments in Cartwright's philosophy of science, I have nowhere suggested that her account of science is derived from or even similar to any extant position in the hermeneutic tradition. Instead, I am suggesting that, without trying, she succeeds where most hermeneutic accounts have failed, namely in making sense of scientific activity as a hermeneutic process.[24]

NOTES

1. All the more so, as the author of this chapter would not describe his own interests or background as that of hermeneutics, either. The choice of the label reflects the difficulties not of understanding Nancy Cartwright's work but of accounting for its originality. The various contributions to the Konstanz workshop on 'Nancy Cartwright's philosophy of science' identified salient issues, but most treated these in terms of similarities and differences to a "received view" (as Cartwright herself tends to do, for example Cartwright 1999: 183). This chapter adopts the heuristic of setting her work quite apart by focusing on central passages in her three main works, which state her position in a germane and idiosyncratic fashion. As it happens, all these passages consider clearly identifiable hermeneutic situations.
2. This is the weakest of the various objections against a hermeneutics of science. It sets up as a straw man the hermeneutic notion of interpretation as if somehow it *must* mean more than an appropriate fitting of some input into a given context.
3. According to the constructivists, scientists 'do not take laws they have established in the laboratory and try to apply them outside. Rather, they take the whole laboratory outside, in miniature. They construct small constrained environments totally under their control. They then wrap them in very thick coats so that nothing can disturb the order within' (Cartwright 1999: 46).
4. Just like her concern with *ceteris paribus* conditions in *How the Laws of Physics Lie*, her discussion of the distinction between causal laws and mere generalizations serves Cartwright as 'a kind of ladder to climb out of the modalization programme, a ladder to be kicked away at the end' (Cartwright 1999: 169). This chapter attempts to characterize where she ends up after the ladders are kicked away.
5. Cartwright is quoting (1985) *The Economics of John Stuart Mill*, Oxford: Blackwell: 325.
6. For a non-theory-licensed concretization from a schematic diagram compare (Cartwright 1989: 225).
7. While Nowak's title promises a book on "idealization", Cartwright points out that, according to her terminology, he deals with abstraction (Cartwright 1989: 202; see Nowak 1980).
8. Also, just as models provide for the measurement of capacities, they measure or judge theory. Moreover—as remains to be shown—the two measures become commensurable in the model. The model can take on this productive task of producing commensurability because of the tension between the various functions of the model coupled with the aim of science to overcome this tension and establish as direct a link as possible between theory and the phenomena. The characterization of Mill's mixed method identified two different functions of models. To the extent that the model provides an idealized setting in which one gains acquaintance with a tendency, it supports the causal explanation of concrete phenomena. To the extent that the model results from the concretization of a materially abstract theory until the theory runs out, it instantiates an abstract relation that supports theoretical explanation. Already in *How the Laws of Physics Lie* Cartwright speaks of this 'tension between causal explanation and theoretical explanation. *Physics aims to give both*, but the needs of the two are at odds with one another. One of the important tasks of a causal explanation is to show how various causes combine to produce the phenomenon under study. Theoretical laws are essential in calculating just

what each cause contributes. But they cannot do this if they are literally true; for they must ignore the action of laws from other theories to do the job' (Cartwright 1983: 12) emphasis added.

9. In light of Cartwright's analysis one should say more precisely that Mill's theorist constructs a model in his own mind in order to manifest a tendency which he then observes.
10. It would appear that this idea assimilates Cartwright's view ever more closely to Margaret Morrison's notion of models as mediators or instruments, that is, of physical models that '*can* take on a life of their own as a way of mediating between technology, theory, and phenomena;' (Morrison 1998: 70). Mauricio Suárez identified three features of mediating models and added a fourth: (1) they are not derivable from theory, (2) they are not necessitated by empirical data, (3) they can replace the phenomena themselves as the focus of scientific research and thus become a quasi-autonomous source of knowledge, (4) they fix the criteria used to refine theoretical descriptions of the phenomena (Suárez 1999: 169–171). Cartwright doesn't speak of mediating models but distinguishes, instead, between representative and interpretive models. However, the *modelling* that on her account is done with these two kinds of models satisfies all four criteria of mediation.
11. From the point of view of the philosophy of science there is another way of formulating this achievement: Like Karl Popper's 'Epistemology without a knowing subject' (Popper 1972), Cartwright provides us with a depersonalized epistemology according to which '"p" says that p' (Wittgenstein 1922: 5.542). Unlike Popper, Cartwright develops the tools with which to analyze the mediations between theory, model, and world, i.e. with which to appreciate hermeneutic processes in science.
12. It is said that the hermeneutic approach equivocates between the scientific interpretation of nature and scientists' interpretations of scientific texts. However, this equivocation is part already of the decidedly antihermeneutic self-understanding of science. First and foremost, theories and hypotheses and descriptions and predictions are to be literally true of their object—they do not allude to, evoke, or illuminate nature; they do not enter into a dialogue with the world. Even when it is said that science reads the book of nature, this is not to be the kind of reading which effects a change in the reader who attempts to constitute symbolic meaning in the encounter with the text. Secondly and by the same token, the claims of science, including its so-called interpretations of data, are to be taken literally—they do not require interpretation by those who have learned to read them. Therefore science appears to be most successful where it manages to become entirely unselfconscious about its means of representation and where it establishes conditions under which nature itself appears merely to produce imprints, traces, or effects, i.e. where it leaves its mark and inscribes itself into our representations. This is the view according to which science 'reveals [. . .] directly the language in which the Book of Nature is written' (Cartwright 1999: 46). While analytically distinct, the two notions of literalness mutually support each other and only together achieve the ideal of unselfconscious immediacy of agreement between mind and world and among minds.
13. Compare Cartwright's discussion of the requirement of "contextual unanimity" in (Cartwright 1989: 143–148).
14. Cartwright endorses, for example, Leszek Nowak's claim that Marx's law of value applies to an economic system that 'resembles ideal gases, perfectly rigid bodies', that is, an empirical domain in which it is 'satisfied vacuously' (Cartwright 1989: 203).

15. Again, from the point of view of the philosophy of science there is another way of formulating this achievement: Following Wittgenstein, Thomas Kuhn offered a generic account of literalness as a construction that to the members of the scientific community does not appear to be constructed. According to Kuhn, membership in a scientific community requires the acquisition of a shared language. By learning to speak the same language, scientists become socialized into an interpretive community where agreement and disagreement about empirical matters no longer appears to involve interpretation at all (Wittgenstein 1958: remark 241). Again, Cartwright improves on this generic account by showing that this interpretive community should not be presupposed in our accounts of normal science but that the acquisition of a shared language and the training of scientists go hand in hand with concrete knowledge of the conditions of literalness for abstract theories (Cartwright 1999: 43).
16. "Fitting out" is introduced by Cartwright as a translation of the German word "einkleiden," whereas "fitting to" corresponds to the German "an-" or "einpassen" or "annähern".
17. Cartwright first discussed these differences in a seminar with Paul Grice on metaphysics in which 'we talked about pretences, fictions, surrogates, and the like' (Cartwright 1983: 129).
18. Of this reconstruction, Cartwright says it is an engineering task rather than scientific (Cartwright 1989: 211.)
19. 'Lessing said about his examples, "I do not want to say that moral teaching is expressed (*ausgedrückt*) through the actions in the fable, but rather . . . through the fable the general sentence is led back (*zurückgeführt*) to an individual case." In the two-body system [. . .] Newton's law is "led back" to the individual case' (Cartwright 1989: 44).
20. This is the point of Cartwright's *simulacrum* account of explanation: Theory is applied to the construction of the models and the similarity of the models with concrete situations is then determined or established (Cartwright 1983: 143–162).
21. See notes 11 and 15.
22. Moreover, the equivocation was seen to be endemic not to hermeneutics but to the scientific claim to literalness (see notes 12 and 15 above).
23. Such an investigation would probably show up differences between Cartwright's argument against van Fraassen in *How the Laws of Physics Lie* and her account of representative models in *The Dappled World*. By relying in the latter work on R. I. G. Hughes's notion of representation, Cartwright attributes to the representative model more clearly the characteristics of a hermeneutic circle (see Cartwright 1999: 192; and Hughes 1998: 128).
24. It would go beyond the scope of this chapter and the expertise of its author to relate Cartwright to the considerable variety of positions in the hermeneutic tradition (e.g., Gadamer, Heelan, Bubner, Ihde, or—to the extent that he wishes to be counted in—Hacking). A fairly general understanding of the hermeneutic project allowed me to identify the five obstacles to its applicability in the case of science and nature. Like the majority of Cartwright precursors and readers I am perhaps falsely assuming that I would have heard of a hermeneutic account that overcomes these obstacles and has yet something to say about the peculiar dynamics of scientific inquiry. Unlike most of her precursors and readers I am neither shocked nor surprised that someone firmly rooted in the "analytic tradition" of the philosophy of science has managed to do so. —I thank various critical readers of earlier drafts, especially Davis Baird and Jan Schmidt.

REFERENCES

Cartwright, N. (1983) *How the Laws of Physics Lie*, Oxford: Clarendon.

———. (1989) *Nature's Capacities and their Measurement*, Oxford: Clarendon.

———. (1999) *The Dappled World: A Study of the Boundaries of Science*, Cambridge: Cambridge University Press.

Cartwright, N. et al. (1995) 'The tool box of science: Tools for the building of models with a superconductivity example', in W. E. Herfel et al. (eds) *Theories and Models in Scientific Processes*, Amsterdam: Rodopi.

Gadamer, H. G. (1975) *Truth and Method*, New York: Seabury.

Hughes, R. I. G. (1998) 'The Ising model, computer simulation, and universal physics', in M. Morgan and M. Morrison (eds) *Models as Mediators: Perspectives on Natural and Social Science*, Cambridge: Cambridge University Press.

Lessing, G. E. (1854) 'Abhandlungen über die Fabel' in *Gotthold Ephraim Lessings gesammelte Werke*, Leipzig: Göschen, 4: 231–314.

———. (1962) *Laocoön; An essay on the limits of painting and poetry*, Indianapolis: Bobbs-Merrill.

Markus, G. (1987) 'Why is there no hermeneutics of natural sciences?', *Science in Context*, 1:5–51.

Morrison, M. (1998) 'Modelling nature: Between physics and the physical world', *Philosophia Naturalis*, 35: 65–85.

———. (1999) 'Models as autonomous agents', in M. Morgan & M. Morrison (eds) *Models as Mediators*, Cambridge: Cambridge University Press.

Nordmann, A. (1996) 'Blotting and the line of beauty: On performances by Botho Strauss and Peter Handke', *Modern Drama*, 39: 4.

Nowak, L. (1980) *The Structure of Idealization: Towards a Systematic Interpretation of the Marxian Idea of Science*, Dordrecht: Reidel.

Popper, K. (1972) *Objective Knowledge*, Oxford: Clarendon.

Ricoeur, P. (1981) 'The model of the text: Meaningful action considered as a text', in *Hermeneutics and Social Sciences*, Cambridge: Cambridge University Press.

Suárez, M. (1999) 'The role of models in the application of scientific theories: Epistemological implications', in M. Morgan and M. Morrison (eds) *Models as Mediators*, Cambridge: Cambridge University Press.

van Fraassen, B. (1980) *The Scientific Image*, Oxford: Clarendon.

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*, London: Routledge & Kegan Paul.

———. (1958) *Philosophical Investigations*, New York: MacMillan.

———. (1993) 'Cause and effect: Intuitive awareness', in J. Klagge and A.Nordmann (eds) *Ludwig Wittgenstein: Philosophical Occasions*.

Reply to Alfred Nordmann

Alfred Nordmann offers a hermeneutic reading of my accounts of theories, models and empirical success that I much welcome and for one special reason that I shall explain. Often the question arises, am I a scientific realist. It arises not least because I claim in *The Dappled World* that I had earlier wanted to attack realism—particularly the claim that our best scientific laws are approximately true. By contrast, in *The Dappled World* I take many of the laws as true—so long as we affix the right kind of *ceteris paribus* clause to them: The laws are true so long as the right kind of arrangement and interaction of capacities to generate them is in place and operates without interference.

Stuart Hampshire criticized me for this. Not for the case studies and the detailed lessons I draw from them. Nor for the strictures about warrant and trusting in what some one or another scientific group takes to be the dictates of "well-established" theory for a concrete case without a very great deal of different kinds of corroborating evidence. Rather he criticized me for indulging in questions of "realism" and for supposing it to be worthwhile to ask whether and how theory really describes the world. This is just the kind of metaphysics that he thought he and his colleagues at Oxford—Ayer, Ryle, Austen, Berlin, and others—in league, but naturally not in total agreement, with those elsewhere had left behind. Anglophone philosophy, he had believed, could never turn to them again, just as he thought that the ideas and commitments of the government of "the good Mr Attlee" were a turning point for Britain from which we would never turn back. Perhaps it is because of the too-close association of the political and the philosophical histories, and of my own work with Thomas Uebel and Jordi Cat on the linked shifts in political and philosophical thought in the Vienna Circle, and Peter Galison's work on *Aufbau-Bauhaus* that I have felt particularly shaken by Hampshire's criticisms.

Hampshire himself was no special friend of hermeneutics. Nevertheless I think that the hermeneutic reading that Nordmann proposes of my views show them in a light far more acceptable to the kind of in-the-world empiricism and particularism that we might ascribe to Hampshire, and that I would wish to emulate, than does the framing in terms of realism, universalism, unification, simplicity, and the like familiar in contemporary philosophy of science.

Nordmann says that models, like performances, become impersonal readers of the text of abstract theory, and they create the conditions for literal truth and falsity of theory. But they are not just readers of the text of theory; they are also readers of the text of the world. As readers of both, models establish the commensurability of theory and "the world"—or better, the world as read through what I call "unprepared descriptions" but which Nordmann points out are already prepared 'for scientific or disciplinary treatment' though not yet calibrated to a particular theory, as are what I call "prepared descriptions" (Nordmann this volume: 383).

So why is talk of an "impersonal reader" better than talk of "realism", "fundamentalism", and "unity of nature"? Because it allows a description like the following from Nordmann:

> . . . there are no antecedent guarantees that they [models] will successfully coordinate theory and phenomena. Indeed . . . models may fail to concretize or realize theoretical concepts, and it may require a rather tenuous process to relate . . . models to the phenomena. . . . However, it is also possible for . . . models to be aligned or even coincide. In those instances, it becomes possible for scientists to routinely traverse in both directions between the abstract and the concrete. Cartwright rejects any philosophy of science that takes those cases as its paradigms and thereby ignores the work that is involved in relating phenomena, models, and theories to one another. . . . At the same time, whenever Cartwright considers in her own terms the movements back and forth between the abstract and the concrete, she arrives at what I here call 'hermeneutic moments'. At these moments the models are the stage on which the negotiations take place and on which the top-down and bottom-up approaches become calibrated to each other. Moreover, her hermeneutic characterizations . . . turn the model . . . into a device that interprets, measures, or reads phenomena and theory and that promotes the attunement of concrete and abstract properties.
>
> (Nordmann this volume: 372)

This seems to me an entirely apt and accurate description of what is going on and without any references to the world that are probably, on closer inspection, nonsense, as Hampshire suspected and Neurath certainly believed. We align theory and the world often through the process of simultaneously building the model, building the system it models—literally building, or shielding or substituting a different system with more agreeable characteristics (as in Gähde's account in this volume of Halley who took Jupiter to act only when on one side of the sun and not the other), as well as making the theory say what we need it to by exploiting the flexibility of the mathematical representations and the looseness of the constraints for fixing physics descriptions. The models are the centre point at which the processes get aligned—as best they can.

Other views on models too can be happily rid of any metaphysical overtones they might have been ascribed and read with Nordmann's hermeneutical interpretation. For instance, his claim that 'it is not so clear that "nature" is the immediate object of scientific enquiry' echoes Mary Morgan's idea that in many cases models themselves have become the object of experimental enquiry (Nordmann this volume: 383). This is patent in the case of model organisms, like fruit flies and laboratory rats, and prepared systems, on slides and in test tubes. But it is equally true of the kind of fictional models that we make up and write down.

Nordmann highlights the hermeneutic elements in my story of how models become the objects that theory can describe and make predictions about; Morgan tells of how they become the objects of experiment. We experiment on the models and not on reality; indeed, it is hard to learn from models except by experimenting on them. Morgan's chief examples are from her own field of economics and from biology. But, it is true in spades of much of our contemporary mathematical physics where, Peter Galison tells us, mathematics is the new laboratory.

So I am happy to adopt the description of models as impersonal readers of both theory and the world, both for my own views and those of many others. And I especially embrace Nordmann's descriptions of science—really good science—that take us away from discussions of Truth, Unity, and Beauty, which I ought to have had no truck to begin with, to something far more modest: 'The success of science,' Nordmann tells us, 'consists in the establishment of a more or less local, more or less robust alignment of phenomena, models, and theories' (Nordmann this volume: 371).

Notes on Contributors

Daniela Bailer-Jones (1969–2006) led the research group "Kausalität, Kognition und die Konstitution naturwissenschaftlicher Phänomene" at the University of Heidelberg, Germany. She studied Physics and Philosophy in Freiburg, Oxford and Cambridge, where she received her PhD in History and Philosophy of Science in 1997, and taught at the Universities of Paderborn and Bonn before moving to Heidelberg.

Nancy Cartwright holds the Karl Popper Chair in Philosophy, Logic and Scientific Method at the London School of Economics. She is also Director of LSE's Centre for Philosophy of Natural and Social Science. Her work in philosophy of science is the inspiration for the papers in this volume.

Michael Esfeld is Professor of Philosophy of Science at the University of Lausanne, Switzerland. His main areas of research are metaphysics of science, philosophy of physics and philosophy of mind.

Brigitte Falkenburg is Professor of Philosophy at Technische Universität Dortmund, Germany. She received a Ph.D. in Philosophy from the University of Bielefeld and a Ph.D. in Physics from the University of Heidelberg. Before moving to Dortmund, she held a position at the University of Heidelberg. She also won a Heisenberg Fellowship (supported by the German Research Foundation) and spent a year at the Institute for Advanced Study in Berlin as a Fellow.

Ulrich Gähde is Professor of Philosophy and Department Head at the University of Hamburg, Germany.

Ronald N. Giere is Professor of Philosophy Emeritus as well as a member and former Director of the Center for Philosophy of Science at the University of Minnesota. His current research focuses on agent-based accounts of models and scientific representation and on connections between naturalism and secularism.

Carl Hoefer has been a Professor at the University of California, Riverside, at the London School of Economics, and is currently an ICREA Research

Professor at the Autonomous University of Barcelona. A former student of Cartwright, he specializes in philosophy of space and time, and philosophy of probability.

Iain Martel is Sessional Lecturer in the Department of Philosophy at the University of Toronto.

Margaret Morrison is Professor of Philosophy at the University of Toronto. She works primarily in history and philosophy of physics and is the author of *Unifying Scientific Theories: Physical Concepts and Mathematical Structures* (2000) CUP and co-editor of *Models as Mediators: Perspectives on the Natural and Social Sciences* (1999) CUP.

Alfred Nordmann is Professor of Philosophy of Science at Darmstadt Technical University, Adjunct professor in the Philosophy Department at the University of South Carolina, Columbia, USA, and President of the Lichtenberg Society.

Stathis Psillos is Associate Professor in the Department of Philosophy and History of Science at the University of Athens, Greece. His book *Causation and Explanation* (Acumen 2002) received the British Society for the Philosophy of Science Presidents' Award. He is also the author of *Scientific Realism: How ScienceTracks Truth* (Routledge 1999), *Philosophy of Science A–Z*, (Edinburgh University Press 2007), and editor (with Martin Curd) of the *Routledge Companion to the Philosophy of Science* (2008).

Julian Reiss is Assistant Professor of Philosophy at Erasmus University Rotterdam, The Netherlands, and Research Associate at the Centre for Philosophy of Natural and Social Science at the London School of Economics.

Christoph Schmidt-Petri is Lecturer in Practical Philosophy at the Institute for Philosophy of the University of Leipzig, Germany.

Mauricio Suárez is Professor Titular (Associate Professor) at Complutense University of Madrid and Research Associate of the Centre for the Philosophy of Natural and Social Sciences at London School of Economics. He has previously held positions at Oxford, St. Andrews, Northwestern, and Bristol.

Paul Teller is Professor Emeritus of Philosophy at the University of California at Davis.

James Woodward is J.O. and Juliette Koepfli Professor of the Humanities at the California Institute of Technology.

Index