

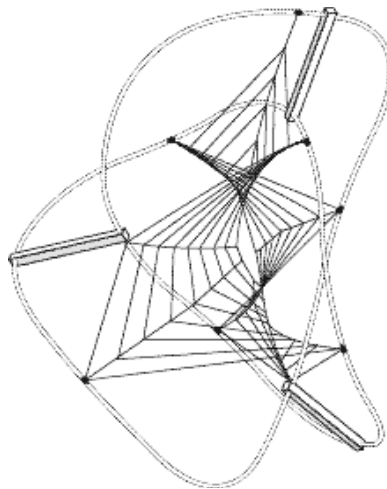
*Centre for Philosophy of Natural and Social Science*

*Contingency and Dissent in Science*

*Technical Report 04/07*

*Causal Powers:  
What Are They? Why Do We Need Them?  
What Can Be Done with Them and What Cannot?*

*Nancy Cartwright*



Series Editor: Damien Fennell

The support of The Arts and Humanities Research Council (AHRC) is gratefully acknowledged.  
The work was part of the programme of the AHRC Contingency and Dissent in Science.

Published by the Contingency And Dissent in Science Project  
Centre for Philosophy of Natural and Social Science  
The London School of Economics and Political Science  
Houghton Street  
London WC2A 2AE

Copyright © Nancy Cartwright 2007

ISSN 1750-7952 (Print)  
ISSN 1750-7960 (Online)

All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of the publisher, nor be issued to the public or circulated in any form of binding or cover other than that in which it is published.

## Contents

Causal Powers: What Are They? Why Do We Need Them? What Can and Cannot Be Done with Them?.....	2
I - Causal laws, policy predictions and the need for genuine powers.....	6
II - In Favour of Laws that are not Ceteris Paribus After All .....	31
III - What makes a capacity a disposition? .....	46
IV - Are RCTs the gold standard?.....	58
V - Economic Models: No Capacities, No Inductions .....	70

## Permissions

Essay I is forthcoming in Handfield, T (ed.), *Dispositions and Causes*, Oxford University Press.

Essay II is reprinted with permission from *Erkenntnis*, Volume 57, no. 3 , pp.425-439, (2002), Springer Netherlands.

Essay III is reprinted with permission from Kistler, M. and Gnessousou B., *Dispositions and Causal Powers*, (2007), Ashgate.

Essay IV is reprinted with permission from *Biosocieties*, Volume 1, no. 1, pp.11-20 , (2007), Cambridge University Press.

Essay V is forthcoming in *Philosophy of Science*, Proceedings of the 2006 Biennial Meeting of The Philosophy of Science Association Part II: Symposia Papers.

**Causal Powers:  
What Are They? Why Do We Need Them?  
What Can Be Done with Them and What Cannot?**

What are causal powers and why should we believe in them? Causal powers are now a central topic in metaphysics but my defence of them does not begin there, but rather in studies of the practices of the sciences, especially in my case, of physics and economics. Both of these use the analytic method: they ascertain the behaviour that would result from the operation of a cause ‘in isolation’; then take this behaviour to provide the ‘contribution’ that that cause makes to the behaviour that occurs even when the cause is not in isolation. What counts as ‘isolation’? And how is the notion of ‘contribution’ to be understood?

Start with ‘isolation’. At the most general level the best we can say is that a cause is isolated (for purposes of ascertaining its causal power) when everything that can interfere with the production of the natural effect of the related causal power is absent (or, more realistically, calculated away) and all the helping and triggering factors necessary for the natural effect associated with the power to be produced are present. ‘Contribution’ is even harder to characterize at a general level. This is what the cause ‘tries’ to produce even when not in isolation. We could try something like: for any situation there is a systematic difference in the result between what occurs when the cause operates and when it does not; that difference is determined by the contribution. This is too strong for general use though. Consider the pin stuck in the dust between the floorboards. The magnet *may* raise it, but I want to allow that given the same situation again – if that concept even makes sense – the magnet may raise the pin on one occasion but not on the other, and without any fixed probabilities. After the fact in a successful case we can see the difference as determined by the contribution of magnetic attraction. But we cannot say that there *will* be a systematic difference, merely that there *can* or *may* be.

Some may object to these characterizations because they are so non-reductive, using both funny concepts – like ‘interference’, ‘triggering’ and ‘necessary helping factors’ – and funny modalities like ‘can be’ or ‘may be’. I would defend them against this kind of objection (and have done so regularly<sup>1</sup>). My objection is that they are too abstract to be of practical use. If we are in debate about whether a targeted factor has a putative causal power, these characterizations (and others like them) do not provide much help for how to settle the matter.

---

<sup>1</sup> Cf. Cartwright (1989) and (1994) as well as essays in this collection.

We can do much better when we move to a more concrete level. The forces of classical particle mechanics are a paradigm of causal powers. Here we have a detailed theory that provides a more practical account. On the standard story about classical mechanics (though not as I see it<sup>2</sup>) a cause, such as a mass like the earth or a magnet, is in isolation when no other forces are at work; no triggering is needed; nor are any helping factors necessary if the cause is truly isolated. The contribution to motion is the mass of the affected body times the acceleration that it undergoes in the experiment – i.e., in this case the force due to the gravity of the earth or to the presence of the magnet. The notion of the systematic difference produced by a cause in a given situation is cashed out in the vector addition of its contribution – say the force of gravity or the force of magnetic attraction – to the contributions of the other causes in that situation, i.e., the forces exerted by the other causes.

The other paradigm I have discussed frequently is linear structural equation models in economics. Suppose that the equations in the model represent causal laws for a given range of situations. This will impose a number of constraints on the variables that appear in the equations, on the form of the equations, and on how the equations are to be interpreted (e.g., left of the equal sign is the effect with a full set of causes on the right).<sup>3</sup> The coefficients in these equations, for instance the price elasticity of demand, represent the contribution of the causal power with which a given value of the associated variable affects the effect represented in the equation. The differences made by the various contributions present in any situation are additive. A cause acts in isolation to produce a given effect in any situation in which the values of all the other variables in the equation for that effect are zero. In principle this is the appropriate method for ascertaining what the contribution is, though in practice sophisticated econometric methods are used – but methods, note, that are only justified, I would argue, on the assumption that it is indeed causal powers that are being estimated.

I point to these two examples to stress that causal powers play a central role in serious science, where the abstract notions that help characterize them can be cashed out in concrete concepts with real scientific bite (relative to the assumptions of the discipline, of course). I do so for two reasons. The first is scientific.

---

<sup>2</sup> The standard account has it that nothing can interfere with a force except another force. This assumes that causes like the wind blowing about an object falling under the influence of gravity can always be represented properly as forces, which, I argue, is more a metaphysical assumption than one assured by solid empirical evidence. Cf the essays here and also in Cartwright (1989).

<sup>3</sup> For sufficient conditions for a set of linear equations to represent causal laws, see Cartwright (1989).

Power terms appear regularly in discourse related to the sciences but often without the rich theoretical development we see for forces or the features like the demand elasticity of economic modelling. For instance, ‘efficacy’. Evidence-based policy, which is now mandated throughout the US and the UK, calls for scientific evidence of efficacy before a policy is agreed to, and by far the best evidence for efficacy, we are told by government and other agencies in both places, is the randomized controlled trial (RCT). These trials are provably good at testing *context-dependent* causal laws<sup>4</sup>, but these do not directly bear on other contexts. Nor is this generally what advocates in the policy communities focus on. They deal rather in the currency of ‘efficacy’. The idea is that the RCT measures an average difference that the cause under test contributes; the average is taken across the range of subpopulations represented in the test, where in each subpopulation all the possible interfering factors take on some fixed arrangement of values.

That is how the average contribution across the population in the test is measured.<sup>5</sup> But what does ‘contribution’ mean here? In the mechanics and economics examples a contribution makes a systematic difference, where the science in question supplies the rule for how to calculate what the difference is. In the case of policy, however, ‘efficacy’ seems to be left as an abstract, cross-disciplinary term. But should we not be able, on each occasion of use, to cash it out within some concrete scientific rubric in which it can be properly scientifically defined? Otherwise it looks as if we are left with advice like that for the pin and the magnet: the cause *may* (or *can*) make a difference, but when, how often and what the difference is remains unpredictable. Is this good enough? Is a better scientific base required to legitimize the concept on any occasion that we are to put it to use in policy and if so, can we find the requisite bases?<sup>6</sup>

These are questions that matter for the relation between pure and applied science and the first reason I stress the paradigms for causal powers in mechanics and in economic modelling is to raise questions like this. My second reason is to focus the attention of philosophers. Causal powers play a serious role in serious science, so it behoves us to take them seriously, to try to provide a proper philosophic account of them, just as with other

---

<sup>4</sup> See Essay IV in this volume.

<sup>5</sup> See also my discussion in Essay IV that argues that RCTs can at best measure the size of efficacy but do not bear in any direct way on whether it is really an *efficacy* that is being measured; that is, nothing in the usual RCT design works to show that the average difference that the cause contributes in the experimental population has any kind of constancy or invariance that can be expected to be contributed elsewhere (unless the design itself, unusually, extends to providing good reason to think that the test population is a true representative sample of the target population).

<sup>6</sup> This question is similar to the complaint I cite in Essay V from Anna Alexandrova about capacity claims derived from economic thought experiments.

scientifically central concepts such as ‘law of nature’, ‘invariance principle’ or ‘theory acceptability’. That is why I have put together the essays in this volume. Some are reprints and others are preprints of papers to be published elsewhere. Together they constitute my recent efforts, from a number of different viewpoints stretching from metaphysics through economic thought experiments and RCTs, to come to grips with the notion of causal power that is so central across scientific arenas, both pure and applied.

### References

Cartwright, N. (1989), *Nature’s Capacities and their Measurement*, Clarendon Press: Oxford.

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

## Essay I

### **Causal laws, policy predictions and the need for genuine powers\***

#### Abstract

Knowledge of causal laws is expensive and hard to come by. But we work hard to get it because we believe that it will reduce contingency in planning policies and in building new technologies: knowledge of causal laws allows us to predict reliably what the outcomes will be when we manipulate the factors cited as causes in those laws. Or do they? This paper will argue that causal laws have no special role here. As economists from JS Mill to Robert Lucas and David Hendry stress, along recently with philosophers like James Woodward and Sandra Mitchell, they can do the job only if they are invariant under the manipulations proposed. But then, I shall argue, *anything* that is invariant under the proposed manipulations will do this job equally well. There seems to be nothing special about causal-law knowledge in and of itself that makes it particularly valuable for policy and technology prediction. What seems to matter is invariance alone, not causality.

But what guarantees invariance and how do we know when it will obtain? Here certain kinds of causal laws do have a special place – those underwritten either by what I have called ‘nomological machines’ or by what I have called ‘capacities’. Capacities and nomological machines have a double virtue that makes them invaluable for policy planning. First, the causal laws they give rise to will be invariant so long as they obtain; and second, they typically have visible markers we can come to recognize that tell us when they obtain. The markers for nomological machines are shakier than those for capacities, though, since capacities are often tied to markers by well-established empirical laws. Capacities have their own drawback however, which is the final topic of this paper: the causal laws that are guaranteed by a capacity connect the obtaining (or triggering) of a capacity with its exercise. But Hume argued (mistakenly I suggest) that no distinction can be made between the obtaining of a power and its exercise.

#### 1. Introduction

Knowledge of causal laws is hard to come by; it is expensive. Yet we make enormous effort to obtain it. Why? What is the use of knowing causal laws? I have always supposed that knowing causal laws will help us to change the

---

\* Research for this paper was supported by the AHRC grant for ‘Contingency and Dissent in Science’. I am grateful to the AHRC for this support and to the LSE research group associated with the project for help in thinking through the issues discussed here.

world and in a very immediate way: we can read off from the causal laws ways to manipulate the world to achieve the effects described in the laws. Nor have I been alone in this supposition; it is, I think, at the heart of various manipulation theories of causality.

Unfortunately I now believe it is mistaken. Knowledge of causal laws cannot play the special role I expected of it. In this paper I will attempt to explain why. I will also provide two special cases where causal knowledge is backed up in just the right way to make it of use in the way envisaged: the case of nomological machines and the case where *capacities* are at work. There is a cost with respect to the latter, however, at least for the die-hard Humean. For the notion of capacities that does the job is a metaphysically heavy notion that cannot be cashed out in terms of lawful relations among ‘occurrent’ properties, a notion with one of the central characteristics that made Hume despise powers. Much of my discussion will focus on my worry about the usefulness of causal-law knowledge for policy prediction. I turn to capacities, with their similarities to Hume’s much despised powers, as a solution in the last few sections.

For purpose of illustration I shall consider deterministic causal laws that are expressed in a system of linear equations of a familiar sort:

*System of linear deterministic causal laws*

$$x_1 c= u_1$$

...

$$x_n c= \sum_1^{n-1} a_{ni}x_i + u_n.$$

Here the  $u$ ’s represent quantities not caused by any of the  $x$ ’s, and the symbol ‘ $c=$ ’ means that the left and right-hand side are equal and that the factors on the right are a complete set of causes of those on the left. (Reference to the population and circumstances is repressed as is usual in presentation.) In the case of yes-no variables the analogue is J.L. Mackie’s famous formula for INUS conditions, supposing that all factors on the right-hand-side are genuine causes of the one on the left:

*Boolean deterministic causal law*

$$E c\equiv A_{11}A_{12}\dots \vee A_{21}A_{22}\dots \vee \dots A_{n1}\dots A_{nm}.$$

Analogously with  $c=$ , the symbol  $c\equiv$  means that the left- and right-hand sides are equivalent and the factors on the right are all causes of those on the left.

## 2. What is a causal law?

I argue that causal laws have a number of specific features.<sup>7</sup>

- They are *population-relative*. They describe relations that hold among quantities in a particular kind of population in particular kinds of circumstances, though this relativisation to population and circumstance is often left to context to supply.
- *The primacy of singular causation*. The causal relation they report is not primitive, say an abstract relation between universals. Rather, causal laws describe what singular causal processes will or can occur in the specified circumstances whenever the prescribed causes are instantiated.
- In keeping with the fact that a causal law reports what singular causal processes will or can occur, a causal law picks out at least one *complete set of causes*: the specified set is *enough* to produce the effect and nothing more need be added.
- The causes reported in a causal law may produce the prescribed effect with a *variety of frequencies*. The laws may be deterministic – the causes always produce the designated effects in the prescribed populations in the prescribed circumstances; they may be probabilistic, where the effect is produced with some fixed probability; or the effects may even be produced by hap, with no fixed probabilities. (It should be noted that what we say about probabilistic laws depends on our views about probabilities. Do we want, for instance, to say that the effect *will* occur with the relevant limiting relevant frequencies in some mythical infinite sequence or instead that it *can* occur or that it has a certain *propensity* to occur?)
- Causal laws may use a *peculiar modal form* familiar in daily life but not much studied in philosophy: the specified causes *can* produce the designated effect – they are enough to do so – but they *may not*, with nothing more to be said. There may be no fixed probabilities with which the effect occurs nor any further reasons that work in a systematic way to pick out when it should occur and when not.
- *Factual effects*. Causal laws specify that effects will (or can, in the relevant sense of ‘can’) genuinely obtain when the specified causes obtain. In particular causal laws do not describe counterfactual effects that would (or could) obtain were the causes (or the populations) different from those specified. Nor do they presuppose that the specified causes are produced in any special way unless that is stated in the description of the situation to which the law is relativised. (Note that, as I shall discuss in Section 7, this does not mean that the effects need be occurrent in some narrow ‘Humean’ sense of ‘occurrent’)

---

<sup>7</sup> Cf Cartwright (1989).

To illustrate, a law  $x_n = \sum_{i=1}^{n-1} a_{ni}x_i + u_n$  from a linear deterministic system relative to situation  $S$  says that in every individual instance in  $S$ ,  $x_1$  taking the value  $X_1$  and ... and  $x_{n-1}$  taking the value  $X_{n-1}$  and  $u_n$  taking the value  $U_n$  whenever instantiated causes  $x_n$  to take the value  $\sum_{i=1}^{n-1} a_{ni}X_i + U_n$ .

This account of causal laws has a number of important virtues.

1. It tells us what causal laws say. Many contemporary treatments of causality in philosophy of science do not do so, including common versions of the probabilistic theory of causality, related Bayes-nets theories and some versions of manipulation and invariance-under-manipulation accounts.

Consider Judea Pearl's account,<sup>8</sup> which begins with a set of causal equations like those of the linear deterministic system of Section 1. His work supposes that the equations say at least that in any system of the kind under study the values of the quantities in the equations are always related as the equation describes. But there is more, an asymmetry; the causal equation tells us that the quantities on the right are *causes* of those on the left. But what does a law say in telling us that? The Bayes-nets axioms, relating causal laws and probabilities, put constraints on the set of causal laws: only some sets of causal laws are admissible for any given probability distribution. This is not enough to answer the question. Beyond asserting that the specified functional relation holds in the system under study no matter what values of the quantities are instantiated, what more does the causal equation say? I have already explained my answer.

Wolfgang Spohn supplies a different answer,<sup>9</sup> one that will be dear to the hearts of those who think causal language is a veil. Roughly, Spohn maintains that causal laws are summaries of the kinds of complicated patterns of conditional probability relations among time-ordered quantities represented in a Bayes net. So it becomes important to Spohn to make the axioms relating causal laws and probabilities strong enough so that only a single set of causal laws is consistent with a given probability. Otherwise different, and incompatible, sets of causal laws would be equally appropriate for summarizing the same probabilistic facts. An alternative might be to claim that a causal law says that a certain abstract relation holds between universals, the universals that are represented by the variables in the equations.

Consider too the probabilistic theory of causality for yes-no variables. One fairly good attempt at formulating it says that  $C$  is a cause of  $E$  in test

---

<sup>8</sup> Pearl (2000).

<sup>9</sup> Spohn (2001).

population  $K$  for situation  $S$  (i.e.  $C$  appears on the right-hand side of a Boolean deterministic causal law for  $E$  relative to  $K, S$ ) if and only if in  $S$ ,  $P(E|C\&K) > P(E|\neg C\&K)$ , where a test population is one in which all other sources of variation in the probability of  $E$  are held fixed barring  $C$ .<sup>10</sup>  $C$  causes  $E$  in  $S$  *simpliciter* if it does so in any test population in  $S$ . Hence it can be true in  $S$  both that  $C$  causes  $E$  and that  $C$  causes  $\neg E$ . (Once some particular form for the causal laws in question is specified it becomes possible to say more concretely what counts as other sources of variation of the probability of the effect.)

There are at least two understandings of this theory.<sup>11</sup> On the one hand we can see it as an answer to my question about what a causal law says. The law ‘ $C$  causes  $E$  in  $S$ ’ says that  $P(E|C\&K) > P(E|\neg C\&K)$  for some test population  $K \subseteq S$ . On the other hand we can see it as a sufficient condition for the truth of a claim that says something else more immediately about causation. That’s what I try to do. I offer an account of what the law says in terms of singular causings that will (or ‘can’) occur ‘in the long run’. Then I show for various kinds of causal systems that when the probabilistic theory is formulated appropriately for them, the increase in conditional probability is indeed sufficient for the related causal law to be true.

These two understandings of the probabilistic theory can also help make clearer my concern that we should, if we can, offer an account of what causal laws say. Many discussions nowadays settle for describing some central ‘characterizing’ features of causal laws. Often these descriptions themselves refer to causal laws, as in the probabilistic theory, which refers to other causal laws true in the situation in order to characterize  $K$ . In this case the descriptions cannot serve as definitions. That is not the concern I am raising, however. I myself take this kind of ‘circularity’ to be a virtue of an account since, I argue, we have good reasons to think that causality is endemic, that many causal relations are as basic as anything can be and that we have good epistemic access to various kinds of causal relations. What I want to know is what a causal law says, and what it says could very well include something about other causal laws. On neither understanding can the probabilistic theory provide a noncircular definition of a causal law. Only the first provides an answer to my question, albeit an answer that seems to me to be mistaken.

---

<sup>10</sup> The formulation I give here still isn’t quite right because  $K$  must not hold fixed any causal intermediaries by which  $C$  causes  $E$  on a given occasion. My own best attempt relies on reference to singular causings even in the formulation of the probabilistic theory. See Cartwright (1989) and Cartwright (2007) for a fuller discussion.

<sup>11</sup> I would like to thank John Worrall for pointing out to me the importance of making clear these two different understandings of the probabilistic theory.

2. The account I offer of causal laws is empiricist in what I take to be the most important sense, that stressed by Otto Neurath: it is *this-worldly*. Causal laws do not describe relations between universals nor structures behind the happenings; they describe what happens when the causes are instantiated. (Though note again that, as I shall point out in Section 7, there are a great many more things that I take it we have good reason to count as ‘this-worldly’ than a narrow ‘Humean’ does.)
3. This account of what causal laws say dovetails with a panoply of our most favoured methods for testing causal laws: various statistical methods used in econometrics and other social sciences, the mark method, randomised control trials, controlled experiments and tests looking for invariance under controlled interventions. For a number of these, formal proofs can be provided that positive results on the test – if the test is ideally conducted in the appropriate setting – is sufficient for the truth of the causal law as I interpret causal laws.<sup>12</sup> I take this to be a strong argument in favour of this interpretation of what causal laws say since it seems to me essential to good philosophy and to good scientific practice that metaphysics and methodology can be shown to be mutually supporting.

I am at pains to explain what I take causal laws to say since causal laws are the topic of my concerns in this paper. I am worried that, contrary to expectation, causal laws cannot after all provide us with the kind of information we need to predict what will happen as we attempt to change the world. For this reason an account of causal laws that meshes with methodology matters, since what we seem to assume is that what we establish with our best methods for testing causal laws carried out in the best circumstances is knowledge that we can use directly: in knowing the causal law we know how to change effects by changing their causes and we can make precise predictions about the results of so doing. Perhaps the reader will have a different view about what causal laws say. Much of my argument in what follows applies to other views as well, but here we have at least one articulated view to keep in mind that maintains the all-important connection between metaphysics and methodology.

---

<sup>12</sup> The proofs naturally require assumptions about features of singular causings. Also the singular causal reading of causal laws is not unique in having this virtue, at least for some of the methods. Holland and Rubin (1988) for instance show that in the right kinds of populations, positive results in a randomised controlled trial are sufficient for a causal law, supposing that the law makes claims about the occurrence of singular counterfactual differences.

### 3. Add-on v intrinsic accounts of causal laws

Economists at least from the time of JS Mill have worried about the usefulness of causal laws for policy prediction due to their instability. Mill worried about naturally occurring variations in both the background arrangement of causes and in the underpinning structures that support causal laws. Early econometricians and more recently Chicago School economists like Robert Lucas worried more about the likelihood that active policy intervention would undermine the structural arrangements that support the causal laws. The worries I raise here about the usefulness of causal laws for policy prediction echo these earlier worries about the stability of causal laws by economists.

To explain my worry I shall review a few recent accounts that aim to provide central characterizing features of causal laws, focussing on accounts that treat economic causes since these are most alert to the problem. Begin by considering the probabilistic theory of causation, described in Section 2, which I claim provides a provably sufficient condition for the designated cause to appear in a true causal law.

For a given population  $S$  let us focus on some particular test (sub)population  $K$  so we can repress  $K$  in the conditional probability. Consider what use we can make of knowledge of the law that  $C$  causes  $E$  in  $K$  for policy predictions about the effects in  $K$  on  $E$  of changing  $C$ . The probabilistic account of causality seems ideal for providing an answer: increase the probability of  $C$  and the probability of  $E$  will increase because

In  $K$ ,  $P(E) = P(E|C)P(C) + P(E|\neg C)P(\neg C)$ .

So if  $P(C)$  increases in  $K$  so too should  $P(E)$  supposing  $P(E|C) > P(E|\neg C)$ .

But this is not the case. The formula above is for a given probability measure  $P$ .  $P(C)$  cannot change without the measure  $P$  changing and if  $P$  changes to some new probability  $P'$ , the fact that  $P(E|C) > P(E|\neg C)$  determines nothing about the relation between  $P'(E|C)$  and  $P'(E|\neg C)$ . Knowing that  $C$  causes  $E$  in  $K$  and hence knowing that  $P(E|C) > P(E|\neg C)$  does not help us predict the effects on  $E$  of manipulating  $C$  in  $K$ .

It turns out that this is exactly the problem that concerns econometrician *David Hendry* when he gives an account of causation.<sup>13</sup> As with the theory of probabilistic causality, Hendry focuses on the conditional probability of the

---

<sup>13</sup> See Hendry (2004).

effect on the hypothesized cause,  $P(E|C)$ . His primary attention is to cases where  $C$  is strictly exogenous to  $E$ . This means that the conditional probability,  $P(E|C)$ , can be estimated without attention to the marginal probability,  $P(C)$ . For the relation between  $C$  and  $E$  to be causal however, Hendry requires *super-exogeneity*:  $P(E|C)$  must be invariant across the envisaged policy changes. So Hendry stipulates that ‘ $C$  causes  $E$ ’ is not to be counted as a causal law unless  $P(E|C)$  remains fixed across the manipulations envisaged to change  $C$ .

This is an example of what I call an *add-on* account of causality par excellence.<sup>14</sup> A criterion that is sufficient to guarantee a causal law on its own is provided – as I indicated, the probabilistic theory is provably sufficient for a causal law to obtain. Then invariance is added on top as a second demand before the label ‘causal’ is allowed. One can of course quarrel with my account of what a causal law says, insisting that the account is not complete until invariance is added on. This is essentially what Hendry does (and also James Woodward as we shall see shortly). But recall my claims to a connection between the content of a causal law and our best methods of testing for causal laws. The additional invariance requirement is not part of most standard methods, not just for those based on the probabilistic theory but also for a host of conventional methods for establishing causal laws, from the mark method to randomised controlled trials.<sup>15</sup>

*James Woodward* also has an add-on account.<sup>16</sup> When it comes to causality Woodward focuses on systems of linear equations of the kind I described in Section 1. For Woodward two demands must be fulfilled before equations like these can properly be labelled ‘causal’. First

- *Level invariance*: the equation must remain invariant under any changes on right-hand-side variables ‘by intervention’.

‘Intervention’ is hard to define properly; it is something like a ‘miracle’ from the Lewis account of counterfactuals, a change in the cause at the last stage that affects nothing other than the cause and things causally downstream from it.

Woodward defends this as a condition on causality with a number of examples in which it is violated by relations known to be spurious. What we would like to know is whether all spurious relations violate it; that is, is level invariance a sufficient condition for causality? I have a kind of representation

---

<sup>14</sup> See Cartwright (2006) and Cartwright (2007).

<sup>15</sup> Though some methodologists, especially in economics, now frequently add on tests of invariance.

<sup>16</sup> Woodward (2003). See also Sandra Mitchell (2003) who stresses the need for invariance without a detour through causality.

theorem to show that it is.<sup>17</sup> I begin with some axioms that a set of causal laws should satisfy – like asymmetry, irreflexivity and the assumption that any functional relations that hold are generated by genuine causal relations.<sup>18</sup> The theorem shows that any functional relation generated by a set of causal laws will be one of those causal laws if and only if it is level invariant.

So I am happy with level invariance as criterion of causality. But Woodward is not. He demands in addition

- *Modularity*: there must be at least one way to change the other causal relations in a system that leaves any genuine causal relation invariant.<sup>19</sup>

The effect of this requirement is that each variable in a system<sup>20</sup> can be changed (by changing the law that governs it) without changing anything else except the effects of that variable. This again is a ‘miracle’-like change. What justifies this as a condition on causality? Woodward is clear:<sup>21</sup> this addition allows us to use the relation in question for manipulation. That I take it is why Woodward calls his account of causality indifferently an ‘invariance’ account and a ‘manipulability’ account.

As an answer to my worries it is far less satisfactory than Hendry’s condition however. For it underwrites the use of causal laws for predicting the outcomes of manipulations not for the manipulations we might be envisaging, as with Hendry, but only for the very special kinds of ‘surgical incisions’. These are the kinds of manipulations that are demanded in a controlled experiment. That, I believe, is how they come to play such a special role in Woodward’s account. They are good for a very special way of testing for causal laws. But they are no good for showing why knowledge of causal laws is useful for policy predictions.

---

<sup>17</sup> See Cartwright (2003) reprinted in Cartwright (2007).

<sup>18</sup> I take these axioms to be fairly innocuous and to be true of causal laws even if my singular-causings account of causation is mistaken. I should note that there has been some objection that the axioms are not so innocuous because a transitivity axiom is included. But the transitivity axiom assumes only that if x appears as a cause of y in a linear deterministic causal system, we still have a causal law for y if we substitute for x the right-hand side of any causal law that has x as effect. I think this is necessary unless we are willing to assume that causation in nature is not continuous in time, so that there is a notion of direct causal law (the ‘last’ law in operation before the effect is produced) that is not representation relative and that it is this notion of direct causal law that we are trying to characterize.

<sup>19</sup> See Woodward’s definition of modularity in Woodward (2003), p. 329.

<sup>20</sup> I.e., any variable that appears as an effect in a law in the system of laws.

<sup>21</sup> Actually, he gives the same reason – causes must be usable to manipulate their effects – for both level invariance and for modularity. I cite it only for modularity because level invariance does not provide manipulability unless modularity is added and I at any rate have an alternative defence of level invariance.

Woodward’s account also shares the central defect of Hendry’s when it comes to addressing my worries. Modularity, just like the requirement of *superexogeneity*, is an add-on. Level invariance already provides a sufficient condition for licensing causal laws. But if we want to use those laws for predicting what happens as we manipulate the causes, to think of causes laws as useful for policy prediction, we must *add on* invariance.<sup>22</sup>

To reinforce this, turn to an account of causality offered by macroeconomist and methodologist *Kevin Hoover* that is not an add-on account.<sup>23</sup> Hoover defines causality directly in term of the effects that can be achieved by manipulation.

*Hoover: C causes E iff anything we can do to fix C partially fixes E but not the reverse.*

Although this definition secures a connection between causation and manipulation and does so with no add-ons, the kinds of relations it calls ‘causal’ would not count as causal in everybody’s books – like probabilistic theories of causation, causal process theories or Lewis-style counterfactual accounts. Figure 1 provides an example of a simple mechanism to illustrate, where the *u*’s are ‘policy levers’ – quantities we can manipulate, and the solid lines with arrows depict pure ‘mechanical’ causation, like pushing on a lever at one end to trip a switch at the other. The dotted line depicts ‘Hoover’ causation.

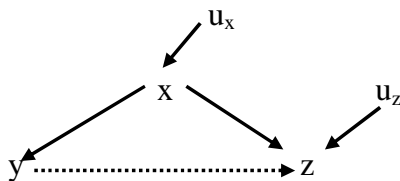


Figure 1 – Mechanical and Hoover-Causation

Comment [TH1]: Please supply.

<sup>22</sup> Though note that Woodward’s condition doesn’t really do the trick since we are guaranteed by him just that there is at least one way to change the cause that leaves the causal law invariant and this may well not be among any of the ways we envisage implementing policy. Moreover for Woodward the one way that is guaranteed could be just an ‘in principle’ way to change the cause, not a way that is at all accessible to us. If so, no connection at all with policy prediction is secured. What the ‘in principle’ manipulation is good for is an ‘in principle’ test that the causal law obtains.

<sup>23</sup> See Hoover (2001). I should note that the description I give here of Hoover’s account is not one he is happy with. I claim it is what his definitions say and take the kind of causal relation described by the definitions as a very important one different from more ‘mechanical’ kinds of causal relations. He maintains that he intends his account to cover the more conventional notion of ‘mechanical causation’ and that various caveats he offers allow his definitions to do so. For further discussion, see Cartwright (2007), section III.4.

So Hoover does not need add-ons. If he is right about what causation is, we can see why causes by their very nature can provide predictions about strategies for manipulating the world. But many will not want to allow that Hoover's is an account of causation at all.

Where then is the expected connection between knowledge of causal laws and our ability to predict the consequences of manipulating causes? It seems it is not there. The requisite predictive capacity is not provided by a causal law. It is achieved by adding on a further assumption, an assumption whose sole motivation seems to be to achieve this connection.

#### 4. What then is special about knowledge of causal laws?

So, what's wrong with adding on a Hendry-like invariance assumption? We can add it on as an additional requirement for the *use* of knowledge of a causal law. 'Warning: do not use this product for predictions about the consequences of manipulations without ensuring that it is invariant under those manipulations.' But it would be a bad idea to add it on as a requirement on the truth of causal laws for a number of reasons.

- As I have already stressed, this would break the connection between the metaphysics of causal laws and our standard methods for testing them.
- It relativises the concept to the manipulations under which it is supposed to be invariant. We could perhaps suppress the relativisation if the set contained only one kind of member, like Woodward's miracle-like manipulations. But this does not go very far in establishing the usefulness of causal knowledge for predictions about the results of manipulating causes.
- It breaks with standard usage of the concept of causal law that we put to other purposes. We want to know about the causal laws for systems that are shaky, where attempts to manipulate the causes may undermine the laws: snowflakes and spider webs, undefusable bombs and old-fashioned children's toys that break if wound up too tightly; and also for systems that we never intend to manipulate or never could manipulate, like the processes in the sun that produce light and heat, the planetary system and the innards of my car engine and computer, where I firmly intend never to go.

We want to know because we want ways to ascribe responsibility about what is happening in the system, because what happens there may be analogous to what happens elsewhere, because we want to be able to repair the system or dismantle it, because how the system operates may have side effects, because it is a central feature of human nature to be consumed with curiosity about how the world around us operates, etc etc etc.

Knowledge of causal laws is thus valuable in dozens of different ways, even if it does not provide direct predictability about how we can achieve effects by manipulating causes, as we might have unreflectively supposed. So forsaking it would be foolish and relabelling it by demanding the add-on requirement of invariance under some set of manipulations leaves us in need of another label for what seems to me to be the original concept.

All this leads me to conclude that adding-on is a bad idea. But knowledge of causal laws without adding on invariance does not help with policy predictions in the desired way. But the situation is far worse even than this for causal knowledge. Once we allow invariance to be added on, causal knowledge seems to have no special role at all. *Any claim that is invariant under a proposed manipulation of some factor will provide correct predictions about the results of manipulating that factor.*

Look again for instance at a law from a linear deterministic causal system, where for years many of us have been keen to insert a symbol like “c=” not just an equality sign:  $x_n \text{ c=} \sum_1^{n-1} a_{ni}x_i + u_n$ . If we insist that the functional relation be invariant under a particular manipulation, say of  $x_i$ , the fact that the equation expresses a causal law is irrelevant to what we can predict about  $x_n$ . If we insist on invariance of the relation, any relation is as good as a causal relation for predictions about results of the manipulation. And if we do not insist on invariance of the relation, causal relations have no guarantee to do the job. So we return to my central worry – what is so special vis-à-vis policy prediction about knowledge of causal laws?

Perhaps it is an empirical fact that causal laws are always stable under manipulations, even if this is no part of the concept of causal law and no part of what the causal law says. This is clearly false. I think even Woodward’s far weaker – and far less useful – claim that there is always some manipulation that leaves the causal law invariant is false. Moreover, if true it goes a long way to showing that exactly the same thing is true of certain spurious relations that then have the same claim to usefulness as genuine causal laws.

Consider again the Hoover diagram, Figure 1, with causal laws  $x \text{ c=} u_x$ ,  $y \text{ c=} a_{yx}x$  and  $z \text{ c=} a_{zx}x + u_z$ . Then the ‘spurious’ relation  $z = a_{zx}(y/a_{yx}) + u_z$  is also functionally true. According to Woodward’s modularity thesis, there is some manipulation that can change  $x$  and leave the two causal laws between  $x$  and  $y$  and between  $x$  and  $z$  unchanged. In the Hoover diagram that manipulation could be by manipulating  $u_x$ . But then the very same manipulation is also a way of changing  $y$  whilst leaving the spurious relation between  $y$  and  $z$  unchanged. So if the manipulation ( $u_x$ ) is accessible to us and provides a way to change  $x$  without threatening the power of the causal law  $z \text{ c=} a_{zx}x + u_z$  to

predict what happens to  $z$  as we change  $x$ , it is equally a relation that is accessible to us and that provides a way to change  $y$  without threatening the power of the spurious relation  $z = a_{zx}(y/a_{yx}) + u_z$  to predict what happens to  $z$  as we change  $y$ .

This manipulation does not of course satisfy Woodward's definition of an intervention on  $y$ . It is not a miracle-like change of  $y$  since it also involves a change of a cause,  $x$ , of another variable,  $z$ , in the system that is not an effect of  $y$ . That means it is not good for testing whether  $y$  is a cause of  $z$ . But testing is not our issue. We are here interested in the usefulness of causal knowledge not in the testing of it. And if the manipulation ( $u_x$ ) makes the causal law  $z = a_{zx}x + u_z$  useful for policy predictions about the effects on  $z$  of changing  $x$  (via  $u_x$ ), it does exactly the same for the spurious relation  $z = a_{zx}(y/a_{yx}) + u_z$  for predicting the effects on  $z$  of changing  $y$  (via  $u_x$ ). So even if we disregard the facts that Woodward's modularity demand may well be false as an empirical assumption and that the one manipulation it claims to exist may well not be one we can – or wish to – exploit, the modularity requirement provides no special role vis-à-vis policy prediction for knowledge of causal laws over knowledge of spurious relations. Yet again we have failed in our hunt for what is special about causal knowledge.

There is further reason to be suspicious about a widespread difference in stability of causal laws over spurious relations of this kind. Nature is rife with what I call *nomological machines*,<sup>24</sup> from clocks and vending machines to seeds and caterpillars. The machines of interest here involve a relatively stable arrangement of parts which gives rise to a number of interconnected causal processes inside the machine plus some kind of skin or shield that limits access to the internal variables under a variety of common circumstances. We put the coins in and get out a packet of crisps; we do not perform key-hole surgery on the vending machine to jiggle the levers and chutes inside. We water the seed and plant it in the right kind of soil at the right time; we do not reach in and shift about the internal make-up that will produce the seedling.<sup>25</sup> In these cases the causal processes inside can be very stable. But then all the spurious relations are stable as well, and for the very same reasons.

Often even our efforts to achieve a desired result are directed by our recognition of a spurious relation in total ignorance of the causal laws. For instance, I was paid to keep the needle centred on a dial at the cyclotron lab in Pittsburgh. The point was to keep the beam on target. When the beam drifted so did the needle; and what one did to adjust the needle adjusted the

---

<sup>24</sup> See Cartwright (1989) and Cartwright (2000).

<sup>25</sup> Though my recent gardening experiences suggest that what I am doing with sweet pea seeds must be like taking a sledgehammer to the vending machine.

angle of the particle gun producing the beam. No-one was expected to manipulate the particle gun in the right way directly, we just had to keep the needle centred. The relation we relied on to get the job done was entirely spurious. There are a vast number of other examples as well. Medicine provides many: we frequently successfully treat a mild symptom in order to relieve a severe symptom of the basic malfunction without understanding the causal processes involved. The structure is like the common cause structure from the Hoover diagram I considered a few paragraphs back in discussing modularity. We rely not on causal knowledge but on knowledge of a stable correlation, a correlation that is stable presumably because within a reasonable range whatever we do to relieve the mild symptom passes through the basic malfunction that causes the more severe symptom.

I take it then that causal laws are not universally stable and that a great many situations that guarantee the stability of causal laws also guarantee the stability – and sometimes the equal usefulness for policy prediction – of non-causal relations. Perhaps instead the answer to my worry is just that causal laws are more frequently stable than other relations. Maybe so, but without more ado this does not provide much predictive certainty, especially if there are no markers to indicate which are stable and which are not. I think we can do better if we focus instead, not on the frequency with which causal laws are stable, but rather on what kinds of situations make for stability and, very centrally, on whether we can recognize them.

Now that I have become gripped with despair about the general uselessness of our hard-won causal knowledge but have come to see some hope in the idea of recognizable markers, I realize that I have studied two different kinds of situations that can relieve the problem: situations in which capacities are at work and nomological machines. Nomological machines are not my central concern here so I will turn to them later and only briefly, concentrating instead on capacities.

## 5. Capacities

*Capacities*, I claim, are a source of stable causal laws. When a capacity is properly triggered it will regularly exercise itself in a canonical way. For example, I am *irritable*; I have the capacity to be angered easily: if I find my daughter has piled the dishes in the kitchen and not washed up, I am apt to lose my temper. Triggering may not be essential for a capacity however. For example, the capacity of one massive object to attract another seems to act continuously. Nor does the capacity described in quantum mechanics for an atom in an excited state to emit a photon seem to need triggering even though it is exercised only sporadically.

As I use the term *capacity*, the connection between a capacity and its exercise or the potential for its exercise is analytic.<sup>26</sup> For many capacities we do not need the second phrase, ‘potential for its exercise’, since they are universally exercised when properly triggered; the gravitational capacity to attract a massive object is a case in point – and it does not even need triggering.

But the exercise of a capacity need not occur universally upon triggering even when nothing interferes. Some capacities are probabilistic, like the (quantum) capacity of an excited atom to emit a photon. (As with causal laws themselves, what we say about probabilistic capacities vis-à-vis their exercise depends on our views about probabilities. Do we want, for instance, to say that these capacities *will* be exercised with the relevant limiting relevant frequencies in some mythical infinite sequence?) And some may just be exercised by hap, without even fixed probabilities. In these cases the analytic connection is only with the potential for exercise and to talk about them we use a modal form familiar in daily life but not well studied by philosophers: triggering my irritability *can* produce anger but it *may not*; and perhaps there is nothing more to be said. There may be no real probabilities for the irritability to be exercised nor further reasons that work in a systematic, reliable way to pick out when it is exercised and when not. It may even happen that the capacity is there all my life and never exercised: I have moral luck; I am irritable and occasions arise that can trigger the irritability but this chancy capacity happens never to be exercised.

The important point is that the presence of a capacity guarantees that the matching causal law obtains, whatever be the form of that law: triggering the capacity causes (or causes with some fixed probability or can cause) the exercise of the capacity; and sometimes, as with gravity, just the presence of the capacity can cause it to be exercised.

This is not enough to allow us to predict what happens under manipulations, however. To do that we need to have some way of ascertaining when a capacity obtains and when it does not. But for many capacities we do have just this kind of information. *Mass* is associated with the capacity to attract massive objects; *negative charge* with the capacity to attract positive charges

---

<sup>26</sup> What then of what have come to be called ‘finkish’ dispositions (see Martin, C.B. (1994))– those that are thwarted the instant they start to work? From my point of view we can treat these in at least three different ways, at least the first two of which have long been available in the philosophy of science literature. 1) The disposition is not really there although its putative marker is. The marker is merely putative; a more correct description of it would exclude the possibility of interference (though there may be no way to characterize the exclusion other than ‘nothing interferes with the operation of the disposition’. 2) Finkish dispositions are really cases where one and the same factor or arrangement has mixed dispositions that cancel out each other’s effects. 3) The finkish disposition is indeed exercised; what fails is the manifest result.

and with the capacity to repel negative charges; etc. Capacities can be useful to us just because in many cases Nature supplies secure ways to recognize them: there are features that we have independent means of identifying that guarantee (or make probable or possible<sup>27</sup>) the presence of the capacity. Unlike the connection between a capacity and its potential for exercise, I take it that this connection is not analytic.

There are a number of important metaphysical issues that ultimately need to be resolved but that we can sidestep for many philosophy of science purposes and especially for the purposes for which I introduce capacities here. These include

1. *What is a capacity?* A property? A second-order property? A material mode concept that expresses only a formal mode distinction about how we identify properties? ...?
2. *What is the connection between the capacity and the ‘occurrent’ property with which it may be associated?* Must there be such an occurrent property for every capacity? Must the association be universal or can it be *ceteris paribus*, probabilistic or chancy?

What is central for my purposes here are the two facts I have described:

- Capacities are analytically connected with causal laws.
- For a vast number of capacities we have learned how to tell when they are present and when not. They are associated in a systematic way that we know about with features we know how to identify independently.

Together these two facts guarantee that we can make reliable predictions about the results when we manipulate the capacities. We manipulate the presence or absence (or the degree) of the associated ‘occurrent’ feature. The presence of the feature guarantees (or makes probable, etc.) the presence of the capacity and that in turn guarantees a causal law to the effect that the capacity will be exercised if present or if properly triggered (or that it will with some designated probability or that it can be exercised). So capacities make knowledge of causal laws useful for predicting the results of manipulations.

So too, I have claimed, do nomological machines. But beware: it is not the relations inside the machine that matter. Although nomological machines support the stability of both causal and non-causal relations inside the machine, these, as I have stressed, are not generally of direct use to us. Many

---

<sup>27</sup> For instance, “Eating Shredded Wheat can improve your heart health”. Just as the connection between the triggering of a capacity and its exercise might be chancy, I see nothing to rule out the possibility that the connection between the more readily identifiable markers and the capacity might also be chancy. Even if chancy the connection can still be important to know, especially if we do not know any more reliable connections.

however also support input-output relations that are of considerable use.<sup>28</sup> We water the seed in the right temperature and light conditions and that causes a seedling to grow. We put a pound coin in the vending machine and that causes a packet of crisps to come out. In the next section I shall review in more detail what is in common about capacities and nomological machines that makes for usable causal-law knowledge.

## 6. What makes knowledge of causal laws useful?

Capacities and nomological machines share several important features that allow them to serve as guarantors of the predictions that causal laws suggest about the results of manipulations, some of which I have already talked about with respect to capacities:

1. *Characterizability*. There are available reasonably good characterizations of what a capacity and a nomological machine are. A philosophical account of capacities piggybacks on an account of dispositions and, despite the fact that we know there are many outstanding problems about dispositions the notion is well enough developed that we can have confidence that some analogue of it holds in reality. Similarly with nomological machines. These may not be well enough characterized in my work but we can also draw on discussions of mechanisms by philosophers like William Bechtel and Jan Elster<sup>29</sup> for reassurance that it is not an empty notion.
2. *Underwriting of causal laws*. Both guarantee that specific kinds of causal laws obtain and continue to obtain so long as they are in place. So long as a capacity obtains, so too will the causal law that connects the obtaining or triggering of the capacity with its exercise. So long as a nomological machine is intact, not only are the relations inside maintained, both causal and non-causal, but so too are the input-output relations.
3. *Identifiability*. In many cases there are markers by which the presence of the capacity or the nomological machine can be identified.

---

<sup>28</sup> Recall that on my account causal laws are population and situation relative. Many causal laws, I argue, arise from the operation of a nomological machine; the input-output laws that many machines give rise to are an example. In this case the relativisation of the causal laws is to the proper operation of the nomological machine. For capacities the situation is more complicated. Since I take the connection between the capacity and the related causal law to be analytic, the causal law that connects the obtaining or triggering of a capacity with its exercise cannot be population or situation relative. In this case the laws that are population or situation relative are the empirical laws that hold in many cases, linking the presence of a capacity with some other independently identifiable feature. These are indeed very often population or situation relative even if a few (like the examples I cite from basic physics) may be thought to hold *tout court*.

<sup>29</sup> Others views of 'mechanism', like Stuart Glennan's (2002) or Woodward's (2002) may not be close enough to help here.

4. *Recognizability*. Very often both the markers and the conditions for safe use – use that leaves the relevant causal laws intact – can be recognized independently. I have argued extensively that discovering these markers in the case of capacities is one of the central achievements of science. (Recall as an example that negative electric charge brings with it the capacity to repel other negative charges and the capacity to attract positive charges.) For nomological machines the case is more complicated. Some – seeds and caterpillars and the planetary system for instance – are given by Nature; we learn to recognize them, and to recognize what can and cannot be done to them without harm, both by experience and by theory. Others we build ourselves and for many, identification is made easy; we put labels on them and even instructions for proper use. ('Do not over wind', 'Keep in a cool place'.)
5. *Usability of associated causal laws*. In many cases the causes of the causal laws guaranteed by a capacity or a nomological machine are ones we can recognize and know how to manipulate without undermining the existence of the capacity or the machine that gives rise to the law that we will use for predicting the outcomes of our manipulations.

So both these categories provide simultaneously a metaphysical and an epistemological basis for the usability of causal-law knowledge.

- When a capacity or a nomological machine obtains, there will not only be a stable causal law, but there will also be a *reason* why the law is stable and in many cases we can recognize when this reason holds and when not and what kinds of manipulations will jeopardize it. We do not learn just that this law seems stable under this set of manipulations but not that; that another law has been observed stable under a particular manipulation but we have no basis to speculate about others; that yet another seems not very stable at all. Where capacities or nomological machines obtain, our knowledge about the stability of this or that law – if we have it at all – is not haphazard and inexplicable. It becomes systematic, thus easier to remember, to understand and to generalize from.
- When there are recognizable markers we can come to learn when this system of knowledge applies and when not.
- With some understanding of the sources of stability we can come to learn and to respect the ways in which our manipulations must be carried out. (For instance, we do not rely so confidently on the assumption that striking the key labelled 't' on a computer keyboard will produce a 't' on the screen if when we strike the key we inadvertently knock over a cup of coffee onto the keyboard.)
- When there are understood sources of a causal law, we can often repair the causal law when it breaks down or remove it if it becomes undesirable.

Still, life is not easy. Nomological machines must not be breached if they are to do their job and sometimes it is difficult to tell if they have been damaged. That's my trouble with sweet pea seeds. I think I have kept them dry, not too hot, not too cold, have not mashed them nor irradiated them nor... Yet when I put them in soil of the right temperature at the right time and water them in the right way, they still do not germinate. Something has gone wrong inside but I have no clue what.

The difficulty of recognizing when the support for a causal law has been breached can be considerably less problematic with capacities. Nature seems to have provided many with sure markers that we have learned to identify reliably – consider again the negative charge that marks the capacity to repel other negative charges, or mass, which marks the capacity to attract other massive bodies. But capacities introduce their own troubles, which I mentioned in the Introduction, at least for those who believe in some robust notion of occurrent properties. I do not know what these are really supposed to be but if they are anything at all, the exercise of a capacity should not be among them – and it is the exercise of the capacity that appears as the effect in the causal law that the capacity underwrites. I turn to this issue in the next section.

## 7. Capacities as powers<sup>30</sup>

Return now to my warning in Section 1. Why do I say that the notion of capacity necessary to understand how scientific knowledge supports policy and planning is a metaphysically heavy power notion? Hume taught that it makes no sense to distinguish between the obtaining of a power and its exercise. But this is just what is required to characterize the facts we need about capacities. By 'metaphysically heavy' here I mean that the notion of capacity and the use to which I put it requires the three-fold distinction between

- The obtaining of the capacity
- Its exercise
- The manifest ('occurrent') results.

---

<sup>30</sup> A lot of the material in this section repeats earlier work. But I think it is worth repeating here for two reasons. First, many readers will probably not be familiar with it. Second, and more important for my theses in this paper, I have never before seen or described explicitly how capacities serve as the guarantors of the usability of causal laws. Indeed I have not seen before how useless causal-law knowledge is in and of itself. My earlier work on capacities stressed instead how capacities explain how we can export knowledge from special 'experimental' situations to ones with new settings (not particularly the same settings with new methods for introducing causes) and how we can deploy knowledge of what happens in these 'experimental' settings to construct totally new situations in which totally new laws emerge.

Gravity has the capacity to make heavy objects fall. The attraction of a heavy body constitutes the exercise of the capacity; the motion of the heavy body is the actually manifested result when the capacity is exercised.

What matters here is that the canonical behaviour after which the capacity is named may be seldom if ever the manifest result and the actually manifested result may have no systematic connection with the presence of the capacity. The systematic connection is between the obtaining of the capacity and its exercise. Massive objects – that is, objects with the gravitational capacity – always attract other bodies even should the other bodies never move closer.

We should not be misled by this over familiar example. In the case of gravity the Humean is apt to retort that the effect produced by the presence of a massive body is that other massive bodies become subject to a gravitational force. I agree. Indeed that is my point. What is the causal law at stake? Not that the gravitational capacity in an object, which is guaranteed by empirical law to be there whenever the object has a mass, causes this or that specific motion in another object – the manifest or ‘occurrent’ result. Rather the presence of the gravitational capacity is enough to cause that capacity to be exercised, that is, to cause a ‘force’ to be created; what motions occur depend on the environment in which the capacity is exercised.

Irritability is the same. What my being irritable guarantees is that if triggered I can get angry easily. That is the causal law. What are the manifest results when the capacity is exercised, that is when I ‘get angry’ – my feelings, my behaviour, my words, my body language, my facial expressions – all depend on the environment in which the capacity is exercised. ‘Getting angry’ here is like ‘force’. It does not name some inner feeling or some spectrum of outer behaviours. Rather it labels the exercise of the capacity of irritability, which exercise combines with other facts about the environment to account for the manifest results, my feelings and behaviours, which may differ dramatically from one occasion of exercise to another.

Perhaps we can think of force as the ‘exercise of a power’, it might be argued, but if we do, it does not violate Hume’s strictures since being subject to a gravitational force is an ordinary occurrent or manifest result, like having a mass or moving with a certain velocity. This raises the sticky issue of what an occurrent property is supposed to be, which I maintain has no interesting answer. Of course the force ‘occurs’ in many senses of ‘occur’. (It is, for example, certainly factual as opposed to counterfactual.) Most notable perhaps is the fact that it can be measured, and very precisely. But we measure it ‘indirectly’, as we do in science, by looking for its causes – yes, the massive body was there to produce it, or by looking for its effects – the second body moves differently from how it would move were the first body not there (unless something tricky is substituted for it). We do not, though,

ever measure it by inspecting the impressions it creates in us. So I am keen to admit that the exercise of a capacity is a genuine empirical ‘occurrence’. But then I do not think there are any cogent arguments against admitting all sorts of ‘non-Humean’ features into the knowable empirical world in the first place, features like causings and powers, and exercises of powers and interferences with those exercises.

I do not think, however, that those who find ‘occurrent’ an interesting category – let’s call them ‘Humeans’ here for the sake of a label – should be so glib about sweeping all these exercisings of capacities into that category. Recall that Gilbert Ryle in *The Concept of Mind* warned against positing a single thing – say ‘grocing’ – that occurs when a grocer performs any of the myriad jobs (the manifest results) that she does qua grocer – weighing coffee, wrapping cheese, stocking shelves. We do this all the time in science, however. And I think we do it just because it allows an accurate statement of the causal laws we have discovered. That though does not make these happenings ‘occurrent’ in the right sense for the ‘Humean’, I expect, whatever that sense is. To see the source of my hesitation, look at two more cases in science, one where the exercise does not seem to be occurrent in whatever is the intended sense and one in which it is.

In the case of force, what it seems reasonable for a ‘Humean’ to count as an occurrent or manifest result in terms of force is the *total* force. That after all is linked by well-understood laws with the actual motions. What then of the force of gravity or the Coulomb force, which I take to be the exercisings of the gravitational and electromagnetic capacities respectively? These are called ‘component’ forces and many ‘Humeans’ take this as a justification for the claim that they are occurrent in the appropriate sense: they are there because they are the parts that make up the total force. But are they there in the right sense? Let us look at some capacities in circuitry that seem to be exactly analogous but where to answer ‘yes’ to the same question would be a real stretch.

Capacitors have the capacity to affect the flow of electricity in a circuit; it is called the ‘capacitance’ and is given in a well-established formula. Inductors too have a capacity to affect the flow, called the ‘inductance’, and resistors, the ‘resistance’, both with their characteristic formulae. When these act together in a complicated circuit there is no simple way to ‘add’ their effects, no analogue of the ‘total’ force in which the separate ‘effects’ can be naturally seen as parts that make up the whole. Rather the net effect on the current is calculated from the structure of the circuits and the formulae for inductance, capacitance and resistance by a series of rules that reduce complex circuits to simple ones; then a final rule calculates the net effect in the simple circuit. This is a case where there are a number of distinct capacities at work. The formulae give the causal laws that connect the

triggering of the capacity with its exercising. The rules give a way to calculate what the current will be when all the exercisings occur together. Of course the exercisings occur in some sense or there would not be the same final effect. But there is no sensible way in which we can see them as part of that effect. Nor can I imagine any other way to count them as ‘occurrent’ in some restricted ‘Humean’ sense.

Contrast a case where the exercisings really do occur in the same sense and same way as the overall effect, that is, as the analogue of the motions in the case of force or the final characteristics of the current in the case of circuitry. Many economic models describe economic processes by sets of simultaneous equations. Each is said to represent a separate ‘mechanism’. For instance consider a simple, familiar supply-demand equilibrium model:

$$\begin{aligned}Q_s &= aP + u_s \\Q_d &= bP + u_d \\Q_s &= Q_d\end{aligned}$$

where  $Q_s$  is quantity supplied,  $Q_d$  is quantity demanded,  $P$  is price;  $a$  is positive,  $b$  negative; and  $u_s$  and  $u_d$  represent other factors than price that affect the quantity supplied and quantity demanded respectively.

I take it that the first equation describes the exercising of a supply capacity. This is a particular capacity that obtains in (specific kinds of<sup>31</sup>) equilibrium situations whenever the price is  $P$ . It tells us that the exercise of this capacity is a quantity effect of size  $aP$ . Similarly for the second equation. It tells us that the effect of the demand capacity that obtains whenever the price is  $P$  is a quantity effect of size  $bP$ . Here the ‘Humean’ is perfectly entitled to infer that both effects are occurrent in almost any strong sense of ‘occurrent’ that might be intended. That’s because the model is a simultaneous-equations model. The rules for what happens when a number of capacities are exercised at once is that all the separate equations must be satisfied together. What happens overall must be consistent with the actual obtaining of each of the separate effects.

Notice how different this is from the vector addition of forces. Say we have two negatively charged masses, with the gravitational and Coulomb forces just balanced. Motions, which are the overall outcome, the manifest result, in this case, are uncontroversially occurrent in anyone’s books. The gravitational capacity should produce a motion of the two towards each other; the Coulomb, motion away from each other. In fact the two are

---

<sup>31</sup> Exactly what the causal laws are relativised to is generally not discussed, which gives rise to a standard problem in philosophy of economics – to what situations in the real world is a given economic model supposed to apply?

motionless. Are we prepared to say the two separate motions exist in the motionlessness? If they do, it is certainly in a much weaker sense than the two quantity effects exist in the equilibrium model. And for the circuits, the claim that the separate effects are ‘occurrent’ in some restrictive ‘Humean’ sense is even more farfetched.

## 8. Conclusion

Work by economists like Robert Lucas and David Hendry and more recently by philosophers like James Woodward and Sandra Mitchell, reminds us that for policy and planning we need something invariant, something that can be relied on across the policies to be implemented.<sup>32</sup> But the idea that causes, or casual laws, are linked to strategy directs us to look in the wrong place for the desired invariance. There is nothing about the fact that one thing makes another happen in a situation that means that it will continue to do so if we start changing the methods by which the cause is introduced. Stability is nice when it happens and it is useful to know. But it does not follow from the causal relation itself and that is reflected in the fact that standard accounts of causality need to add it on if they want to get it.

Yet we do succeed in prediction and planning – and by using our knowledge of causal laws. Other people grow glorious sweet peas from sweet pea seeds, all of us rightly expect our lights to turn on when we flip the switch, it is surely correct to expect that the magnet may retrieve the metal earring from between the floorboards and even I, despite my failures at gardening, can generally succeed in getting a packet of crisps from a vending machine. Causal laws – but causal laws with very special sources – vouchsafe these predictions. Nomological machines generate causal laws between inputs and predictable outputs. Capacities guarantee causal laws as well, though in this case we shall have to adopt some metaphysical distinctions that Hume abjured. In both cases, happily, there are frequently recognizable markers by which we can tell that the causal law in question can be relied on for the manipulations we propose to make.

So – causal-law knowledge is not in general useful for policy and planning, despite our huge investments in obtaining it. It can, however, be useful if the causal laws are generated in the right way. But what then about the link between methodology and metaphysics? If we think of causal laws as I do, then there is a clear link between what a causal law says and how we go about justifying causal-law claims. But I have argued that knowledge of causal laws is not then of immediate use in the way we generally suppose unless the laws are generated in very special ways. Where is this reflected in our methodology?

---

<sup>32</sup> Sandra Mitchell (2003).

Econometricians like Hendry who insist on add-on tests for invariance, or social scientists who worry about the ‘external validity’ of claims from the environments in which they are established to those where they may be used, worry about just this question. But the efforts here are thin and unsystematic. In methodology we have manuals and courses on how to establish causal laws. The US government with respect to its ‘No Child Left Behind’ legislation is firm in its policing of what will and will allow to count as evidence for a causal claim; so too is the UK government, for instance in NICE’s manuals of best medical practice. Philosophers too are right in the centre of the fray. We have at least a dozen different accounts of causal laws on offer, most of which are read off from some one or another methodology for licensing causal laws. But where is the methodology for determining if the laws are generated in the right ways to make them stable and usable for policy predictions? And what can we philosophers say – beyond the handful of sketchy suggestions here – about when causal knowledge will be stable – and recognizably so – and when not?

### References

Cartwright, N. (1989), *Nature’s Capacities and their Measurement*, Clarendon Press: Oxford.

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

Cartwright, N. (2003), ‘Two Theorems on Invariance and Causality’, *Philosophy of Science*, 70, 203-224.

Cartwright, N. (2006), ‘Where is the Theory in our “Theories” of Causality?’, *Journal of Philosophy*, Vol. CIII, no. 2, 55-66.

Cartwright, N. (2007), *Hunting Causes and Using Them: Approaches from Philosophy and Economics*, Cambridge: Cambridge University Press.

Glennan, S. (2002), ‘Rethinking Mechanistic Explanation’, *Philosophy of Science*, 69, S342-S353.

Hendry, D. (2004), ‘Causality and Exogeneity in Non-stationary Time Series’, *Causality: Metaphysics and Methods Technical Report*, CTR 18-04, Centre for Philosophy of Natural and Social Science, London School of Economics.

Holland, P. and Rubin, D. B. (1988), ‘Causal Inference in Retrospective Studies’, *Evaluation Review*, 12, 203-231.

Hoover, K. (2001), *Causality in Macroeconomics*, Cambridge: Cambridge University Press.

Martin, C.B. (1994), 'Dispositions and Conditionals', *The Philosophical Quarterly*, 44, 1-8.

Mitchell, S. (2003), *Biological Complexity and Integrative Pluralism*, Cambridge: Cambridge University Press.

Pearl, J. (2000), *Causality: Models, Reasoning and Inference*, Cambridge: Cambridge University Press.

Spohn, W. (2001), 'Bayesian Nets Are All There is to Causal Dependence' in Constantini, D., Galavotti, M.C. and Suppes P. (eds.), *Stochastic Causality*, Stanford: CSLI Publications.

Woodward, J. (2002), 'What Is a Mechanism? A Counterfactual Account', *Philosophy of Science*, 69, S366–S377.

Woodward, J. (2003), *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.

## Essay II

### **In Favour of Laws that are not *Ceteris Paribus* After All\***

#### Abstract

Opponents of *ceteris paribus* laws are apt to complain that the laws are vague and untestable. Indeed, claims to this effect are made by Earman, Roberts and Smith (2002). I argue that these kinds of claims rely on too narrow a view about what kinds of concepts we can and do regularly use in successful sciences and on too optimistic a view about the extent of application of even our most successful non-*ceteris paribus* laws. When it comes to testing, we test *ceteris paribus* laws *in exactly the same way* that we test laws without the *ceteris paribus* antecedent. But at least when the *ceteris paribus* antecedent is there we have an explicit acknowledgment of important procedures we must take in the design of the experiments – i.e., procedures to control for “all interferences”, even those we cannot identify under the concepts of any known theory.

#### 1. Introduction

I am generally taken to be an advocate of *ceteris paribus* laws throughout the sciences, even in physics. But what are *ceteris paribus* laws? According to John Earman, John Roberts, and Sheldon Smith (2002), the distinctive feature of *ceteris paribus* laws is that they do not entail any strict or statistical regularities in the course of events. Nor do they entail any predictions, categorical or probabilistic. Earman, Roberts, and Smith also suppose that a *ceteris paribus* law is not explicit about what precise conditions have to obtain for the regularity after the *ceteris paribus* clause to hold; alternatively that the *ceteris paribus* clause is vague and cannot be stated in a precise form or a precise and closed form.

If that's what it takes, then what I have defended are not *ceteris paribus* laws.<sup>33</sup> The laws I talk about either can be stated in precise and closed form or they entail strict or statistical regularities in the course of events or both. The matter hinges, of course, on what we take to constitute a “precise and closed” description. This returns us to the old issue of how we should police the language of science. I suspect that I am far less strict about what is admissible as a description than are Earman, Roberts, and Smith. My reason is that I find that the less restrictive language is the kind of language that is

---

\* Research for this paper was conducted under a grant from the Latsis Foundation, for which I am very grateful. Thanks also to Christoph Schmidt-Petri for help.

<sup>33</sup> I do not mean to imply that I am opposed to them; simply that they are not the kinds of laws I have been thinking about and defending over the last decade.

regularly employed in exact science; and that attempts to reconstruct this language away produce scientific claims that are at odds both with the ways we test our scientific theories and with the ways we put them to use.

There are two kinds of formulation that I use to reconstruct scientific laws that I think Earman, Roberts, and Smith would reject, the first because of the language it uses, the second because of the limitations it supposes on the descriptive power of theory. I shall discuss each in turn in sections 2. and 3. In section 4. I shall take up the issue of testing. What I say about testing will not only defend the kinds of laws I discuss but also *ceteris paribus* laws as more generally conceived. Section 5. answers some criticisms that Earman, Roberts, and Smith make against a connection I trace between induction and capacity.

## 2. Powers vs. *Ceteris Paribus* Laws

The language I use in reconstructing a number of scientific laws in the exact sciences (most notably physics and economics) is the language of *powers*, *capacities* or *natures* and related concepts such as *interfere*, *inhibit*, *facilitate*, and *trigger*. Those of us raised in the joint shadow of the Vienna Circle and British Empiricism were taught that these kinds of concepts must not appear in science. I was puzzled about this from early on since it seemed to me that many of the most important concepts I learned in physics are power or capacity concepts, *force* being the first, simple example.

A big obstacle to debate here is the problem of characterization: what criteria distinguish capacity concepts from OK concepts? Surely we do not want to adopt the characterization of the early British Empiricists that OK concepts are those built out of ideas that are copied from impressions.

Operationalization was on offer for a while, but it seems to cut too narrowly since it rules out many central theoretical concepts. Nor do I think we can be content with Carnap's similarity circles and the *Aufbau*.

In my own early attempts to understand these empiricist strictures, I proceeded differently, in a way that generally works best for 'trouser' words (that is, for concepts whose primary meaning comes from what they rule out): figure out what is supposed to be wrong with the illicit concepts; the OK ones are those that don't have those problems. What then is supposed to be wrong with power concepts? One central worry comes from the fact that power concepts seem to be tied either too closely or too loosely to their related effects. This in turn is thought to lead to problems in testing claims about powers. I shall discuss these latter in section 4.

That powers and their effects are tied too closely was the complaint of the old Mechanical Philosophers against Scholastic concepts.<sup>34</sup> What makes heavy bodies fall? Gravity. What is gravity? That which makes heavy bodies fall. For those like Ernst Mach,<sup>35</sup> who wished to provide explicit measurement procedures for the concepts of physics, Newton's second law seems to suffer the same difficulty.  $F = ma$ . What is it for a body of mass  $m$  to be subject to a total force of size  $F$ ? A mass of size  $m$  is subject to force  $F$  iff its acceleration is  $a$ .

On the other hand, when the power is not defined in terms of the occurrence of its effects, there seems to be no fixed connection between the existence of the power and the occurrence of its effects, neither strict (i.e., universal) nor statistical. Aspirins have the power to relieve headaches; that is surely consistent with the fact that they do not always do so and perhaps there is no fixed statistical relation either. This objection to powers echoes an objection that Earman, Roberts, and Smith make to *ceteris paribus* laws. There is a familiar way to fix this problem: insist that the effect is there after all whenever the power is.<sup>36</sup> One well known case of this arises in discussions of the problem of evil. God is omnipotent: He has the power to create any kind of world at all. Couple this with the auxiliary assumption that He is all good. The effect to expect is a benign world, full of delights. Instead we see plagues and poverty and vice. One stock response is that the world is all good despite appearances. We simply fail to see or perhaps to understand the situation properly. I take it that this kind of claim must be judged unacceptable by standards employed in successful science. The world does not appear good; it does not pass any of the standard tests for being good; and its effects are not the effects we are entitled to predict from a world of virtue and perfection.

Or consider Freudian claims, which we know distressed many followers of the Vienna Circle. Consider a crude version of one Freudian example. Freud maintained that the childhood experiences that the Ratman sustained have the power to make one desire the death of one's father. Freud says: "... he [the Ratman] was quite certain that his father's death could never have been an object of his desire but only of his fear ... According to psychoanalytic theory, I [Freud] told him, every fear corresponds to a former wish which was now repressed; we were therefore obliged to believe the exact contrary of what he had asserted ... He wondered how he could possibly have had such a wish, considering that he loved his father more than any one else in the world."<sup>37</sup> But the Ratman did not recognize this desire in himself, he

---

<sup>34</sup> Cf. Glanvill (1661).

<sup>35</sup> Mach (1893).

<sup>36</sup> Or, to be more fair to the proponent of powers, 'whenever the power obtains and the circumstances are propitious for its exercise.'

<sup>37</sup> Freud (1909), p. 39.

appeared to others as a loyal and loving son and he had behaved just like someone concerned to ensure the welfare and safety of his father. But that's alright on Freud's view. The desire is really there; it is just unconscious and thus does not manifest itself in the usual ways.

Turn now to what Earman, Roberts, and Smith call "special force laws", like the law of universal gravitation (A system of mass  $M$  exerts a force of size  $GMm/r^2$  on another system of mass  $m$  a distance  $r$  away) or Coulomb's law (A system with charge  $q_1$  exerts a force of size  $\epsilon_0 q_1 q_2 / r^2$  on another system of charge  $q_2$  a distance  $r$  away).<sup>38</sup> These are not strict regularities. Any system that is both massive and charged presents a counterexample. Special forces behave in this respect just like powers. This is reflected in the language we use to present these laws: one mass *attracts* another; two negative charges *repel* each other. *Attraction* and *repulsion* are not among what Ryle called 'success' verbs.<sup>39</sup> Their truth conditions do not demand success:  $X$  can truly attract  $Y$  despite the fact that  $Y$  is not moved towards  $X$ .

But perhaps, as with the delights of our universe or the Ratman's desire for the death of his father, the requisite effects are really there after all. Earman, Roberts, and Smith feel that the arguments against this position are not compelling. I think they are: the force of size  $GMm/r^2$  does not appear to be there; it is not what standard measurements generally reveal; and the effects we are entitled to expect – principally an acceleration in a system of mass  $m$  a distance  $r$  away of size  $GM/r^2$  – are not there either.

Contrast a different case.<sup>40</sup> In simultaneous equations models in economics each equation is the analogue of a special force law: each describes the operation of a single cause. When more than one cause is present, all the equations must be satisfied at once. So the pattern of behavior that occurs is one consistent with each equation separately. Unlike mechanics, the 'special force laws' in economics really are strict regularities (if true at all). The effects demanded by each law separately are really there – and they meet standard requirements for doing so: the effects of each appear to be there; standard measurements reveal them; and the effects of these effects are the ones we are entitled to expect.

The price level in economics is a contrast. It is calculated by summing the 'contributions' of a variety of different causes. But we do not want to think of the price level as literally composed of a lot of distinguishable parts as a wall is composed of its stones – the level from  $w$  to  $x$  is that due to the stock of

---

<sup>38</sup> Note that throughout I take the special force laws to ascribe *forces* and not *motions* to situations.

<sup>39</sup> Ryle (1949).

<sup>40</sup> See Cartwright (1989), ch. 4.

money; from  $x$  to  $y$ , that due to the velocity of money; etc. This seems a highly unnatural reading to me. And it seems even more so when we move to engineering examples – say the construction of complex machines from simple ones or of circuits from combinations of resistors, capacitors and impedances – where the rules for how to calculate what happens when the parts act together are not by simple addition as they are in the case of *force* or *price level*.

Kevin Hoover in his extended study *Causality in Macroeconomics* backs up my point by criticizing the assumption of linearity of causal influences; that is, the assumption that “the influence of  $Y$  can be added to the influence of  $M$  and so forth”<sup>41</sup> in calculating their effect when operating jointly. (In Hoover’s example,  $Y$  is income,  $M$ , money and the effect is the interest rate.) Hoover complains, “But linearity is unduly restrictive.”<sup>42</sup> Hoover is particularly concerned with the non-linearities arising from rational expectations theory, which imply that the influences of macroeconomic causes cannot be calculated just by addition. He illustrates with a mechanical example:

A gear that forms a part of the differential in a car transmission may have the capacity to transmit rotary motion from one axis to another perpendicular to it... The capacity of the differential to transmit the rotation of the engine to the rotation of the wheels at possibly different speeds is a consequence of the capacities of the gear and other parts of the differential. The organization of the differential cannot be represented as an adding up of influences nor is the manner in which the gear manifests its capacity in the context of the differential necessarily the same as the manner in which it manifests it in the drill press or in some other machine

<sup>43</sup>  
...

Does all this matter? Pretty clearly it does not matter to the economics or to the physics under discussion. But it does matter to the metaphysics, particularly to the topic under discussion here – *ceteris paribus* laws. To see why let me explain how I see the difference between physics and many of the human sciences.

We study capacities throughout the sciences. Many of the central principles we learn are principles that ascribe specific capacities to specific features that we can independently identify, from the capacity of a massive object to attract other masses to the capacity of maltreatment of a child to cause that child to maltreat its own children, or, to mention the example that Earman, Roberts, and Smith discuss in their 2002 paper, the capacity of smoking to cause lung cancer.

---

<sup>41</sup> Hoover (2001), p. 55.

<sup>42</sup> Ibid., p. 55.

<sup>43</sup> Ibid, p. 55f.

What is special about physics then? Not that it does not offer knowledge about powers or capacities but rather that it has been able to establish other kinds of knowledge as well, knowledge that we can couple with our knowledge of capacities to make exact predictions. This additional knowledge is primarily of two kinds: 1) We know for the powers of physics when they will be exercised (e.g., a massive object *always* attracts other masses); and 2) we have rules for how to calculate what happens when different capacities operate together (e.g., the vector addition law for forces). This kind of knowledge is missing for many other subjects. That is why they cannot make exact predictions.<sup>44</sup>

Now for *ceteris paribus* laws. Consider Earman, Roberts, and Smith's example

(S) CP, smoking causes lung cancer  
of which they say, "If some oncologist claims that (S) is a law, then, we maintain, there is no proposition that she could be expressing..."<sup>45</sup> I disagree. I take it that the proposition she is expressing is

(S') Smoking has the capacity to cause lung cancer.  
a claim exactly analogous to the special force laws of physics. This is a precise claim: it states a matter of fact that is either true or not; it is not vague; and it has no *ceteris paribus* clause that needs filling in. So it does not suffer from those faults Earman, Roberts, and Smith ascribe to *ceteris paribus* laws. More central to their objections, it is testable, it makes predictions, and it entails regularities in the course of events, in this case statistical regularities. This is the topic of section 4.

### 3. The Limits of Scientific Languages

In the discussion so far I have been more liberal in my reconstruction of the language of science than most modern empiricists. I allow it to cover more, to talk about powers and capacities. Now I shall propose a way in which I think the languages of the different sciences can describe less.<sup>46</sup>

Consider Newton's second law,  $F = ma$ . What does it say? Many, probably including Earman, Roberts, and Smith, take it to describe a strict regularity. I think that it does so only conditionally. The claim we are entitled to believe from the vast evidence in its favor is this: *if nothing that affects the motion operates that cannot be represented as a force*, then ... The two views

---

<sup>44</sup> But they often can make rough predictions or give good advice.

<sup>45</sup> Earman et. al, (2002), p.295.

<sup>46</sup> For a more detailed discussion see Cartwright (2000).

collapse together if all causes of motion can be represented as forces. Why do I think many might not be?

Newtonian mechanics, like many other theories in physics, has, I believe, very much the structure that C. G. Hempel taught us theories have.<sup>47</sup> It consists of internal principles, such as Newton's three laws of motion, which give relations among the central concepts of the theory, and bridge principles, which constrain how some of the concepts of the theory are applied. Many early logical positivists hoped that the bridge principles would lay out direct measurement procedures for all the concepts of the theory. They had to settle for less. The bridge principles match some theoretical concepts with concepts 'antecedently understood'.

In the case of Newtonian mechanics the primary bridge principles are the special force laws. These license a particular theoretical description – e.g. '... is subject to a force  $F = GMm/r^2$ ' or '... is subject to a force  $F = \epsilon_0 q_1 q_2 / r^2$ ' – given a description in the vocabulary of masses, distances, times and charges – e.g. '... is a mass  $m$  located at distance  $r$  from a mass  $M$ , or '... is a charge  $q_1$  located at distance  $r$  from charge  $q_2$ '.<sup>48</sup>

Bridge principles provide strong constraints. The theoretical descriptions – in our example the individual force functions from the special force laws – are allowed *only if* the corresponding descriptions in 'antecedently understood' terms are satisfied.<sup>49</sup> (For example, "The force on a mass of size  $m$  is  $GMm/r^2$  if and *only if*  $m$  is a distance  $r$  from a body of mass  $M$ ".) The same thing is true of quantum mechanics and its bridge principles as well as quantum field theory, quantum electrodynamics, classical electromagnetic theory and statistical mechanics.

It is because of the issue of evidence that I urge that the bridge principles of these theories are so strongly constraining. I have looked at scores of applications and tests of the theories in my list, applications and tests of the kind that we take to argue most strongly for the truth of these theories. In these cases the theoretical terms that have bridge principle are invariably applied via the bridge principles. This interpretation of the demands of bridge

---

<sup>47</sup> Hempel (1966).

<sup>48</sup> For purposes of this section we can remain neutral about my claim in Section 2. It does not matter for the points here whether we take the special force laws to ascribe capacities of a certain types or instead to ascribe an actually existing force.

<sup>49</sup> There are two caveats here. First, it can happen that exactly the same vectorial quantity  $F$  that normally is associated with one force law applies to a situation 'by accident' even when it does not satisfy the requisite description because of the particular values of the force properly ascribed to the situation by other special force laws. Second, I would like to remain neutral about how strict we need to be about when 'the description offered in the bridge principle is satisfied.'

principles in turn puts a strong constraint on the descriptive power of the theory. Force functions can be legitimately applied only to situations that are described in bridge principles. Similarly for quantum Hamiltonians, classic electric and magnetic field vectors, and so on.

Can all causes of motion be correctly described using just the descriptions that appear in the bridge principles of Newtonian mechanics? To all appearances, not. We have millions of examples of motions that we do not know how to describe in this way. Consider one case where we eventually were successful. For centuries we knew about electricity and magnetism: e.g. rubbing a glass rod against cat fur can cause human hair to move; loadstones can move iron filings; and so forth. But we could not add these in as forces in Newton's second law. Eventually we evolved the formal, precise concepts of electric and magnetic charge as well as the bridge principles that assign them force functions. In so doing we came to ringfence a host of macroscopic situations from all the rest of those that could cause motion but that we could not describe in Newtonian theory: *these* are ones involving attraction or repulsion between electrically or magnetically charged objects. But what of the vast remainder?

We have succeeded in applying Newton's second law to a vast, vast number of cases – but always of the same kinds: the ones that appear in our bridge principles. And there are still not many bridge principles included in Newtonian mechanics, even after 300 years.<sup>50</sup> We are not constantly expanding the theory, regularly producing new bridge principles to meet either new cases or the old familiar ones. Nor do we have continuous success in bringing these cases in under the old bridge principles. This suggests that these cases may well not fall under any descriptions for which there are force functions. And a handful of striking successes does not discount this worry.

My conclusion from these kinds of considerations is that we need to add to the basic 'equations of motion', like  $F = ma$  or Schroedinger's equation, a special constraining condition: The equation holds so long as everything that can affect the targeted effect is describable in the theory. This is the

---

<sup>50</sup> Paul Teller (personal communication) has objected to my claim that in quantum mechanics there are only a small number of bridge principles by pointing out that, as I myself urge, there are a good many 'derivative' bridge principles. These, however, almost never expand the scope of the theory, but rather contract it. For they are in fact, as their name says, *derived*. We start with a situation modeled with a combination of descriptions available from our basic bridge principles. Then we add *more* facts about the situation to derive a new force function for it, by *limiting* the original force function. The derived bridge principle then provides force functions for only a subset of cases that the original did. Of course sometimes we make approximations as we go along. But that, if anything, narrows the scope of the force function even more. For now it is no longer true even of the originally described situation but only of some approximation to it.

formulation of the law that we have strong evidence for. Notice that it is in ‘precise and closed form’ and hence does not look like a *ceteris paribus* law on one of the criteria of Earman, Roberts, and Smith. But how do we test it?

#### 4. Testing

How do we test my version of Newton’s second law or the Schroedinger equation? In *exactly the same way* that we would test them if they had no condition attached. The same is true *mutatis mutandis* for capacity ascriptions and for certain kinds of more conventionally rendered *ceteris paribus* laws. Although there are important differences, let us for the sake of brevity lump all these together and consider them to be of the form, ‘If nothing interferes, then ... (some strict or statistical regularity).’<sup>51</sup>

Suppose we wish to test  $F = ma$  in its unconditional form. We set up a number of different kinds of situations to which, using our bridge principles, we would naturally assign some specific force function. For instance we arrange two bodies of charges  $q_1, q_2$  and masses  $m_1, m_2$  a known distance  $r$  apart so that we can assign the force function  $Gm_1m_2/r^2 + \epsilon_0q_1q_2/r^2$  directed between them. We ensure as best we can that the situation does not explicitly meet any of the other descriptions to which we know how to assign force functions. Then we also ensure as best we can that there is nothing else about the situation that might be assignable a force function – there is no significant wind, no trucks rumbling by, no bright lights, ... Finally we look to see if the motions that occur in all these situations match those predicted from the equation.

Suppose instead that we wish to test ‘If nothing interferes with the operation of the force,  $F = ma$ .’ Everything in the description of what we do will be identical to the previous description except for five words. We substitute for the sentence ‘Then we also ensure ...’

a new one: ‘Then we also ensure as best we can that there is nothing else about the situation that might *interfere with the force’s operation*.’ And what we actually do to ensure this is the same in both cases.

‘But,’ one may ask, ‘how do we know what to eliminate in the second case?’ I think the question is more appropriately ‘How do we know what to eliminate in the first case?’ We do not look for features that figure in our bridge principles as we did in setting up the basic part of the experiment. Of course people who believe in the unconditional form of the law will assume

---

<sup>51</sup> In the case of the equations of motion, as we have seen, the caveat really refers to factors not describable in the language of the theory. Setting aside some niceties, we can assume that capacity claims of the form “A has the capacity to  $\Phi$ ” imply that if nothing interferes A will  $\Phi$ ; and probabilistic ascriptions “A has capacity of strength  $r$  to  $\Phi$ ”, to imply roughly that if nothing interferes the probability that A will  $\Phi$  is  $r$ . But, as I have argued in Cartwright (1999) capacity ascriptions can say a lot more as well.

that the features we are looking for are exactly the other things in the situation that can be assigned force functions. But that does not supply them with a method for picking these features out.

Here's what I think happens. We have seen a vast number of cases of forces at work to which we have tried to fit Newton's second law and over time we have established very strong rules of thumb about what can make trouble for it. That is, we have learned from a lot of experience what kinds of things might *interfere* with the operation of the force. That's what we control for.

Earman, Roberts, and Smith might think my testing strategy lets in too much. They consider what might seem an analogous reading of *ceteris paribus* laws:

It has also been suggested that we can confirm the hypothesis that CP, all F's are G's if we find an independent, non-ad-hoc way to explain away every apparent counterinstance... But this could hardly be sufficient. Many substances that are safe for human consumption are white; for every substance that is white and is not safe for human consumption, there presumably exists some explanation of its dangerousness ... but none of this constitutes evidence that CP, white substances are safe for human consumption.<sup>52</sup>

The reading that Earman, Roberts, and Smith offer for *ceteris paribus* laws is an excellent attempt at the logical positivist program of substituting acceptable formal-mode concepts for dicey material-mode ones. For example: *X causes Y* becomes *X explains Y*, which in turn becomes *Y can be derived from X given the claims of our theory*. Here we have analogously: *X interferes with (x)(Fx → Gx)* becomes *X explains why a, which is F, is not G*, which presumably in turn becomes *¬Ga can be derived from X given the claims of our theory (and perhaps Fa as well)*.

The rendering is a good try, but it does not work, as we can see from Earman, Roberts, and Smith's own example. It does not work for very much the same reason that the analogous formal-mode rendering of causation does not work. You can't get causality and its associated family of concepts out of laws unless the laws themselves are causal laws, not just laws of association, in which case the program of replacement fails anyway.

Moreover, the program is misguided to begin with. There is nothing unacceptable about the concepts of *causation* and *interference*. They are well understood, claims about them are testable and, as G.E.M. Anscombe argues,<sup>53</sup> some causal relations are directly observable; or, more guardedly,

---

<sup>52</sup> Earman et. al. (2002), p. 294.

<sup>53</sup> in her (1971).

causal concepts do not systematically fare worse in any of these respects than other concepts.

We have, I maintain, ample reason to think that there is as much a fact of the matter about whether it is a causal law that forces cause motions as there is about whether it is a law that  $F = ma$ ; that there is as much a fact of the matter about singular causal claims as there is about other relational claims; and as much a fact of the matter about whether  $X$  interferes with some process as about whether the process itself obtains.

The drawback to *interference* is not that there is something wrong with it ontologically; it is rather that we often have epistemic problems. First, we often cannot tell just by looking that  $X$  is interfering with  $\Phi$ , though even this is not always the case. (Sometimes, for example, it is easier to tell that you are, for example, interfering with someone's work than to tell that they are working. My friends Anne and Sandy, for example, sit at their computers typing just as I do now *for fun*.) Usually to make judgements about interference, we need to have a lot of specialized knowledge and a lot of experience; you can't just tell by looking.

Second, it seems that in most cases there are no systematic rules linking  $X$  *interferes with*  $\Phi$  to other descriptions in some special vocabulary that we prefer epistemically (unless the vocabulary is itself heavily laden with concepts that already imply facts about causality, such as *pushing*, *attracting*, *shielding* ...). We almost never have 'special interference laws' to tell us in, say the language of masses, charges, distances and times, when something interferes with something else in the way that we have special force laws to tell us when a particular force function obtains.

We should note though that the absence of 'special interference laws' is not so epistemically damaging as many suggest. The special force laws do tell us when a particular force function obtains, but only for very specific descriptions – the descriptions that appear in our bridge principles. For other descriptions that may be applied far more immediately, such as *a truck passing by* or *the press of the wind*, we are just as much on our own without the help of a system of rules as we are in deciding if we can label the truck passing by as an *interference*.

There are four facts I would like to underline:

- 1) The lack of systematic rules does not mean that we cannot have knowledge about whether a certain kind of occurrence constitutes an interference. Galileo after all knew to use smooth planes for his rolling-ball experiments because he knew he should eliminate the interference of friction with the pull of the earth. Similarly he knew to drop small compact masses and not feathers from the Leaning Tower. And that was

long before he could have had any idea whether friction or the wind exerted a *force* in the technical Newtonian sense.

- 2) The fact that we cannot identify what counts as interference with respect to a claim  $\Phi$  does not mean that we cannot test whether  $\Phi$  is true or not. Consider *Aspirins relieve headaches, if nothing interferes*. We regularly test claims like this in randomized treatment/control experiments.
- 3) Nor does it mean that it is too easy to dismiss disconfirmations.<sup>54</sup> When the predicted result fails to transpire, one can always *say* that something interfered. But saying does not make it true. And as epistemologists are always reminding us, saying, even when it is true, does not constitute knowledge, or even reasonable belief. We need a good reason for claiming that something is an interference. When we do not have any idea whether a nominated factor is an interference or not, then we equally have no idea how to classify the case. Our intended test is no test at all.
- 4) It follows that one needs a great deal of information about what might and might not interfere with a process before we can carry out serious tests on the process and that in turn means that we need already to have a great deal of information about the process itself. That just means that science is difficult, as we already knew, and that it is hard to get started in a vacuum of knowledge.

## 5. Capacities and Induction

In *Nature's Capacities and their Measurement*<sup>55</sup> I offer a number of defenses of capacities:

- a) Once we have rejected Hume's associationist view of concept formation, there is no good argument against the family of concepts connected with causes and capacities.
- b) Strengths of capacities<sup>56</sup> are measurable, just as is the strength of the electromagnetic field vectors or the energy of a system.<sup>57</sup>
- c) 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. (What I mean by 'in some sense' is that the rules for calculating what happens when a number of factors with different 'natural' effects operate together will differ according to subject matter.

---

<sup>54</sup> For a recent example of this kind of claim specifically in the context of the capacity laws I defend see Winsberg et al. (2000).

<sup>55</sup> Cartwright (1989).

<sup>56</sup> This includes their presence or absence.

<sup>57</sup> We cannot, of course, tell by the measurement itself that what we are measuring is a real capacity, anymore than we can tell by the procedures for measuring the electric field strength that what we are measuring is a real quantity. In both cases that requires a lot of theory.

Recall the examples of such rules in section 2.<sup>58</sup>) This procedure is not justified by the regularity law we establish in the experiment, namely ‘In  $I, Q \rightarrow E$ ’; to adopt the procedure is to commit oneself to the claim ‘ $Q$  has the capacity to  $E$ ’.

In *The Dappled World*<sup>59</sup> I add another. With the use of capacity language we can provide a criterion for when induction is reliable. This, I maintain, cannot be done with the use of OK properties and strict regularities alone. Earman, Roberts, and Smith object to this claim. This is not surprising because they also reject one of its major premises.

Imagine we set up a very complex and delicate design,  $D$ , an ideal experiment, to observe the precession of a gyroscope in order to test relativistic predictions about the effects of space-time curvature on precession. The result is  $R$ . To the extent that we believe our design is a good one and that we have implemented it properly, we expect that that result should be repeatable in just that experimental set-up, i.e. we believe that  $D \rightarrow R$  is a strict regularity.

What does it mean that ‘our design is a good one’? That is, what criteria must  $D$  satisfy if  $R$ , which occurs on one occasion of  $D$ , is to occur whenever  $D$  occurs? The crude answer is that  $D$  must control for all factors relevant to  $R$ . I read this as ‘ $D$  is an arrangement in which the capacity of the space-time coupling to produce precession  $R$  operates unimpeded.’ Those who do not like capacities will try other ways to explain relevance. I imagine they look to strict regularities and consider two levels at which they might look.

First, at a concrete level. Look through all the strict regularities involving very concrete features that have  $R$  as a consequent. All and only factors that occur in the antecedents of these are relevant and should be controlled for. My objection to this strategy is not that the list is too long but rather that it will not provide the information we need. Almost *anything* can appear in one of these laws, depending on the arrangement of the other factors; we could design our experiment in indefinitely many ways and still expect the result  $R$ . Any feature that was essential to any of these designs gets counted as relevant. Moreover, the long list of regularity laws with  $R$  in the consequent will not fix *how* a relevant factor should be controlled for. That will depend on the actual design,  $D$ .<sup>60</sup> So lists of strict regularities at a very concrete level

---

<sup>58</sup> In Cartwright (1999) I argued that nature might not have always provided such rules. Even in that case there is a cash value to knowledge about the capacity. The associated effects are more likely to occur when a feature with the appropriate capacity is present than when no such feature obtains. (Consider using a magnet to pull a pin from between the floorboards. It is a good idea to try the magnet even should there be no fixed rules for what happens to the pin in just exactly that combination of circumstances.)

<sup>59</sup> Cartwright (1999).

<sup>60</sup> In my (1988) and (1989) I give examples where two different laws employ the same factor in different ways.

cannot provide a criterion that  $D$  must satisfy if its results are to be repeatable.

A more plausible proposal is to look at a more abstract level, as Earman, Roberts, and Smith propose. The abstract formula for precession is

$$\text{Precession: } d(n_n^r)/dt = \Gamma^r n_s / \omega_s$$

This formula suggests that an adequate criterion is, ‘Eliminate all sources of torque ( $I$ ) except that arising from the space-time coupling as well as all sources of variation in the gyroscope’s moment of inertia ( $I$ ) and in its spin angular velocity ( $\omega$ )’. Let us concentrate on the torque, as Earman, Roberts, and Smith do. They suggest that we couple the formula for precession with ‘the laws relating force to precession and various special force laws’ to fix what  $D$  must be like. What is wrong with that from my point of view? Two things – the first familiar from section 2. and the second from section 3.

- 1) The special force laws are not strict regularities among OK features.<sup>61</sup>
- 2) As with Newton’s second law, we do not have sufficient evidence to ensure that the precession law can be read as a strict regularity. (A more cautious rendering includes a condition: “If nothing that cannot be described as a torque (or a variation in  $I$  or  $\omega$ ) interferes, then ....”)

In their discussion Earman, Roberts, and Smith deny 1), as we have seen. I suspect they would also deny 2). I have explained why I disagree with them. But if we grant either of these assumptions, we see that the job cannot be done with strict regularities alone. We need capacities. The generalization “ $D \rightarrow R$ ” is reliable because  $D$  is a kind of situation in which a stable capacity (the capacity of space-time curvature to affect precession) operates without interference.<sup>62</sup>

---

<sup>61</sup> This is the assumption that figured in the arguments of my (1999), p. 95, where I concluded, “The regularity theorist is thus faced with a dilemma. In low-level highly concrete generalizations, the factors are too intertwined to teach us what will and what will not be relevant in a new design. That job is properly done in physics using more abstract characterizations. The trouble is that once we have climbed up into this abstract level of law, we have no device within a pure regularity account to climb back down again”. The device we need includes the special force laws, which, I maintain, can not be rendered as statements of regularity.

<sup>62</sup> In my (1999) I dubbed situations like this ‘nomological machines’. This is to highlight both the need to eliminate interference, which I have stressed in this discussion, and the need to have the right kind of internal structure (one for which there are rules about how the contributions of the parts combine), which I do not discuss here.

## References

- Anscombe, G. E. M. (1971), *Causality and Determination*, Cambridge University Press, Cambridge.
- Cartwright, N. (1988), 'Capacities and Abstractions' in P. Kitcher and W. Salmon (eds), *Minnesota Studies in the Philosophy of Science, Vol. XIII: Scientific Explanation*, University of Minnesota Press, Minneapolis (1989), 349-355.
- Cartwright, N. (1989), *Nature's Capacities and their Measurement*, Clarendon Press, Oxford.
- Cartwright, N. (1999), *The Dappled World*, Cambridge University Press, Cambridge.
- Cartwright, N. (2000), 'Against the Completability of Science', in M. W. F. Stone and J. Wolff (eds), *The Proper Ambition of Science*, Routledge, London.
- Earman, J., J. Roberts, and S. Smith (2002), 'Ceteris Paribus Lost', *Erkenntnis*. 57, 281-301.
- Freud, S. (1909), 'Notes Upon a Case of Obsessional Neurosis', in P. Rieff (ed), 1963, *Three Case Histories*, Vol. 7 of *The Collected Papers of Sigmund Freud*, Collier Books, New York.
- Glanvill, J. (1661), *The Vanity of Dogmatizing*, London.
- Hempel. C. G. (1966), *Philosophy of Natural Science*, Prentice Hall, Englewood Cliffs.
- Hoover, K. D. (2001), *Causality in Macroeconomics*, Cambridge University Press, Cambridge.
- Mach, E. (1893), *The Science of Mechanics*, Open Court, La Salle.
- Ryle, G. (1949), *The Concept of Mind*, Hutchinson, London.
- Winsberg, E., M. Frisch, K. M. Darling, and A. Fine. (2000), 'Review of Cartwright (1999)', *Journal of Philosophy*, 97, 403-408.

## Essay III

### **What makes a capacity a disposition?**

#### 1. Introduction

Many, if not most, of our highly prized ‘laws’ of physics cannot be adequately rendered as statements of regular association among the values of ‘categorical’ quantities, I have argued.<sup>63</sup> This is true even if we do not balk at the concept of natural necessity and are willing to add that the associations hold ‘by law’. They are rather ascriptions of capacities. They tell us what capacities a system will have by virtue of having a given property. The law of gravity is one example. A system of mass  $M$  has the capacity of strength  $GMm/r^2$  to move another object of mass  $m$  a distance  $r$  away towards itself. I call this *the gravitational capacity*. My second thesis is a commonly shared one. Ascriptions of capacities do not reduce to conditionals involving only categorical properties.

I shall here discuss two questions about these theses: 1) Why think of capacities as akin to dispositions or powers; and 2) Why allow them in science? Before tackling the first question, I shall first try to figure out what features we expect to be characteristic of dispositions and powers themselves.

#### 2. What makes a disposition a disposition?

There are a number of features on account of which we might call something a disposition or a causal power. I shall discuss some that I think are particularly promising; that is, I think they are good starting candidates for characterizing a feature of the world that we have very good evidence for, especially in the overwhelming evidence of everyday experience.<sup>64</sup> The list I shall discuss here includes some characteristics that do not figure centrally in current philosophical discussion. That is because I think my irascibility is a paradigm disposition and I look for characterisations that are sure to include it. There is of course the alternative paradigm for philosophy of science—water-solubility; we may suppose that we also have good reason to accept dispositions of this kind because of their scientific credentials. My focus here, however, is on what I have called *capacities* and my aim is to understand why I and others should have assimilated them to dispositions. Are capacities—like the gravitational capacity—genuinely like dispositions in significant ways, or was the assimilation a mistake, generated perhaps by

---

<sup>63</sup> Cf. Cartwright (1989).

<sup>64</sup> Though of course we should not be wedded to the supposition that there is exactly one such feature we are aiming to characterize.

unthinkingly lumping all non-categorical features together? For this enterprise we may miss important aspects of similarities and differences if we focus our account of dispositions too narrowly.

Here then is my list of promising candidates for what is special about dispositions:

- a. *Substance causation.* The causal relata are not events but rather the cause is an enduring substance and the effect a change in another substance. This was Aristotle's view; so too Kant's according to Eric Watkins.<sup>65</sup> On account of this Watkins ascribes to Kant a causal power view. But this does not seem to be true of what I have called 'capacities'. It seems, rather, that it is the having of a mass by the first object at a given time that causes the motion of another at that time; and this looks far more like event causation than like causation by an enduring substance.
- b. *Latency.* Dispositions are not always on display. If that is so, the capacity of one mass to move another is not a disposition.
- c. *Conditionality.* Dispositions are generally thought to be closely connected with conditionals. The simplest connection is 'If C, then systems with disposition D will M' where C is a description of some specific conditions, including perhaps something that triggers the disposition, and M is a manifestation. Stuart Hampshire, however, has argued the contrary. Hampshire maintains that this connection with conditionals holds for what he calls 'causal properties' but not for descriptions of human character and disposition, which have mistakenly been assimilated to causal properties.

He argues, "Such causal properties of things, as being magnetized and being soluble in aqua regia, manifest themselves, if at all, in specific and definitely storable reactions, which can be produced in specific and storable conditions. The incidents which may count as manifestations of human dispositions – of intelligence, ambition, generosity, and honesty – are *essentially* various and these words are vague, summary, interpretive and indeterminate".<sup>66</sup> It may seem that the gravitational capacity is a paradigm causal property and hence satisfies Hampshire's conditionality condition. I think not and shall return to this topic.

- d. *Malleability.* There are a number of things that can usually be done to dispositions to affect their manifestations. There are three that seem

---

<sup>65</sup> See Watkins (2005), ch. 6.

<sup>66</sup> See 'Dispositions' in Hampshire (1972).

most central: interfering, triggering, and enhancing or retarding. The gravitational capacity does not need triggering – it seems to operate all the time. Nor, it seems, can it be enhanced or retarded. But it can be interfered with. This is a central feature of all the things I call ‘capacities’.

- e. *Two-sidedness*. There is a distinction between the occurrence of a disposition and its being manifested. There is no such distinction for categorical properties. Of course to use this as a characterization of a disposition we must at the same time also recognize the concept of a manifestation and manifestations themselves may be difficult to characterize. Consider: having the minimal unit of electric charge is a categorical property of an electron. But it is often described as ‘hidden’. We have to do something very special to see that it is there. One might say that in the right kind of experiment the charge ‘makes itself manifest’ in certain experimental results. *That* is not the sense of manifestation intended when we contrast dispositional properties, that have manifestations, with categorical properties that do not. Capacities are like dispositions in being two-sided in this way.
- f. *Missing from logbooks*. Hampshire says of a statement that refers to a disposition, “It could not be entered into a logbook of the day’s events opposite some time of the day, or in the annals of someone’s life opposite some definite date.”<sup>67</sup> This I take it is in large part what it means to say that the disposition is not categorical. I suppose it is tied up with the Aristotelian idea that the power is an enduring characteristic of a substance, not one that occurs at specific times.
- g. *Constancy of tendency*. This is the reason I introduced the idea of ‘capacities’ into my discussion of scientific laws in the first place. The outcomes that occur when the gravitational capacity operates are indefinitely various, but there is something fixed. The first mass is always *trying* to bring other masses closer to it; we say that a mass always *attracts* other masses no matter how the other masses actually move. Capacities are modelled on John Stuart Mill’s *tendencies*;<sup>68</sup> I used the word ‘capacity’ to underline that the tendencies I was discussing are tendencies to *cause* things to happen, not just more general tendencies to behave in some particular way (e.g. always to move in a straight line).

The first question that I am addressing in this paper is why assimilate capacities to dispositions I certainly did so but I am not alone. Here are

---

<sup>67</sup> Ibid., p. 34.

<sup>68</sup> Mill (1843).

just some recent examples from Australia, a hot-bed for dispositional analysis. Alan Chalmers says that the appeal to capacities, which we both endorse, “serves to capture what is implicit in common sense usage of utterances such as ‘glass is brittle’ or ‘acorns grow into trees’.”<sup>69</sup> A more thorough attempt to give an ontology of powers, dispositions and capacities appears in Brian Ellis’s book, *Scientific Essentialism*.<sup>70</sup> Peter Menzies in a commentary on my views<sup>71</sup> offers the same analysis for capacities, in terms of conditionals relating categorical properties, that he uses for dispositions.

Still, what have capacities, as I discuss them, to do with dispositions or powers? I suppose the connection has to do with failures not being entirely defeating conditions. One can exercise one’s power even though the result is not achieved. But that seems little connected with the central ideas defining power: control, influence, ascendancy, authority. Where is the authority? The sway over others? Nor does it have to do with what I isolated as the central feature of capacities – that the system always tries to do the same thing even if the results differ.

This leaves me with a puzzle. Mass indeed has the feature under discussion – constancy of tendency – but I do not any longer see what constancy of tendency has to do with either dispositions or powers. This puzzle is exacerbated by returning to Hampshire’s piece on dispositions. One may not agree with all Hampshire’s requirements but I think he is fairly well attuned to the features that characterize human dispositions and character traits, many of which are not part of our current philosophical discourse – and after all, human dispositions are paradigms of dispositions. Some of these are shared by the gravitational capacity and some are not, leaving a mixed verdict. Here are the other characteristics on his list that I have not already mentioned. (The titles and summary are mine.)

- h. *Non-episodic*. “There are short-term and long-term dispositions, but a disposition cannot come into being, then pass away and then come into being again very rapidly.”<sup>72</sup> The gravitational capacity shares this feature with dispositions.
- i. *Necessity of display*. “A disposition must be manifested and must show itself in actual incidents.”<sup>73</sup> Again, the gravitational capacity is like a

---

<sup>69</sup> Chalmers (2002), p.3.

<sup>70</sup> Ellis (2001).

<sup>71</sup> Menzies (2002).

<sup>72</sup> Loc. cit., p. 34.

<sup>73</sup> Ibid., p.35.

disposition by this criterion. Indeed recall my earlier worry that it is perhaps disqualified because it overfulfills this norm – it seems always to be on display.

- j. *Need for scrutiny in ascription.* To be confident in ascribing a disposition one must review actual incidents, looking especially for counter indications. One must thus have the opportunity for “prolonged and continuous study of the conduct and calculations of the person in question. When one has surveyed many incidents and found virtually no contrary evidence, one can say, for example, ‘He is certainly and indisputably generous.’ What is claimed as certain and beyond dispute,” according to Hampshire, “is that the word ‘generous’ is so far the right word to summarise the general trend or tendency of his conduct and calculations.”<sup>74</sup> The gravitational capacity satisfies this requirement, but the requirement does not really seem relevant to it in the right way.
- k. *Manifestations not necessarily behavioural.* “Most ordinary character descriptions refer compendiously to a tendency discernible equally in the behaviour, and in the thought and in the feelings, of the subject.”<sup>75</sup> This is clearly a feature relevant for human character traits and not for my capacities.
- l. *Wide-scope negation.* The opposite of *S has the disposition to X* is not *S has the disposition to not-X* but rather *X does not have the disposition to X*. This is true of capacities like the gravitational capacity.
- m. *Possibility of opposite behaviour.* “To attribute a disposition to someone is never to preclude that he may on some occasion act...in some way contrary to his general tendency or disposition...It is typical of human behaviour...that it allows of lapses...”<sup>76</sup> I take it that Hampshire means we are capable of lapses even without something actively interfering with the disposition in question. Sometimes we simply do not act in character but for no special reason. I am irascible and given to nagging, but I do not always explode about my daughters’ messy rooms – and the days I don’t need not even have been particularly good ones. The gravitational capacity is not like this; it never lapses. Without positive interference the canonical behaviour will always be displayed in the attracted object. Perhaps though it is misleading to focus on the gravitational capacity. An atom in an excited state has the capacity to deexcite but the display is chancy.

---

<sup>74</sup> *ibid.*, p. 35.

<sup>75</sup> *Ibid.*, p. 36.

<sup>76</sup> *Ibid.*, p. 36.

Should we then count a failure of the atom to deexcite in any particular case as a lapse?

Hampshire concludes from his list that statements of human disposition or character are “summarizing statements”. I am not sure if I agree. What I conclude from reflecting on my list and my worries about whether capacities are dispositions or causal powers is

*Not everything that is not categorical is the same.*

This is a familiar kind of lesson, but one I think we need to be reminded of since we have a tendency to look for *the* account of dispositions. There are character traits, dispositions, habits, capacities, powers; they are in humans or in non-humans; they usually derive from some underlying structure, but some may be fundamental. We could hardly expect that there are very many features that all these share.

If we wish to fix on one criterion as central it seems to me from studying my list that what must be common is

*Two-sidedness*: There is a distinction between the occurrence of a disposition and its being manifested.

I take it, though, that this is not the standard choice. The primary focus is generally on conditionality. That I think is a mistake. The reasons are Hampshire’s. Human character traits and dispositions are certainly as central a member of this family as anything else. Yet, as he argues, “The incidents which may count as manifestations of human dispositions ... are essentially various and these words are vague, summary, interpretive and indeterminate.” So no conditional (or set of conditionals) will capture the content of a dispositional ascription.

### 3. How much like other dispositions is a capacity?

If we take *two-sidedness* as the central criterion, capacities like the gravitational capacity do fall into the same family as dispositions, habits, character traits and the like. But how many of the other features associated with this family do capacities share? I want here to focus on two features from our list. The first is conditionality; this is the usual candidate for being the feature that, after two-sidedness, most firmly connects capacities with the other members of the disposition family. The second is malleability; this is my choice.

*Conditionality*. Many take this to be the central feature of the capacities that science studies. Peter Menzies for instance in discussing my notion of

capacity offers just such a conditional account.<sup>77</sup> It looks as if Hampshire too would count the capacity due to gravity as what he calls a causal property – satisfying conditionality – rather than assimilating it to a disposition. On the other hand, it does not satisfy his stringent criteria for the causal properties.

First, the incidents that count as manifestations of the gravitational capacity are indefinitely various. What actually happens to a second object when the capacity operates depends on the setting; the second can move anyway whatsoever. If we follow Gilbert Ryle's usage, though, this will make them not dispositions as opposed to causal properties, but rather 'highly generic' or 'determinable' dispositions. Ryle explains that verbs for highly generic dispositions

... are apt to differ from the verbs with which we name the dispositions, while the episodic verbs corresponding to the highly specific disposition verbs are apt to be the same. A baker can be described as baking now, but a grocer is not described as 'grocing' now, but only as selling sugar now or weighting tea now, or wrapping up butter now.<sup>78</sup>

Second, all conditional claims linking what happens to any categorical description of the setting are, I maintain, *ceteris paribus* laws. For many philosophers this immediately makes them vague. For instance, John Earman, John Roberts and Sheldon Smith maintain that the *ceteris paribus* clause is vague and cannot be stated in a precise form.<sup>79</sup> I claim on the contrary that it can be stated in a precise form: '*If nothing interferes, then...*' This claim is not vague, I argue, because the antecedent refers to a specific state of affairs that either obtains or does not obtain on a given occasion.<sup>80</sup>

This brings us, though, to a third possible reason why the gravitational capacity resembles Hampshire's dispositions more than his causal properties. Though not a vague word, 'interference' is an abstract word, like 'good' or 'work'. Whenever it truly describes a situation, some other more concrete descriptions will always apply as well. But what description that is depends entirely on context, and there are no rules that can be stated using entirely categorical expressions for setting which categorical descriptions will do in which contexts. For any context there is a fact of the matter about whether something is an interference or not; but there probably is no list—even for that context—of what are interferences. Interference is not only multiply realizable; as many philosophers stress, the list of realizations is open ended.

---

<sup>77</sup> Loc cit.

<sup>78</sup> Ryle (1949), p. 118.

<sup>79</sup> Earman et al. (2002).

<sup>80</sup> See Essay II in this volume.

Conditionality, I conclude, is not then a widespread feature of capacities, at least not in any sense in which conditionality goes beyond two-sidedness. Capacities and their manifestations are different, as two-sidedness demands. But there need be no set of conditionals that connects a capacity with specific manifestations.

*Malleability.* The central feature that locates capacities in the family of dispositions, I claim, is not conditionality but malleability. All other members of the family of dispositions, habits and character traits seem to have at least one of the three central features of malleability I mentioned: They need triggering, they can be enhanced or weakened or they may produce different manifestations, or no manifestations at all, if they are interfered with.

So too with capacities. They are all malleable in at least the last way. They can all be interfered with, and when they are interfered with, there is no guarantee that the canonical manifestations will occur. What does occur, if anything at all, will depend on the type and method of interference and there may be no system to either the possible interferences or the possible affect they have on the capacity's manifestations. This indeed, as we have seen, is one of the chief reasons that conditionality fails. It is also why, according to many, capacities can have no place in science. I shall return to this claim at the end. For now let us consider capacities vis-à-vis the two other ways in which dispositions are commonly malleable.

With respect to the other features of malleability, the gravitational capacity that has been my focus seems to sit at the far end. As I said, it does not seem to need triggering and we have no evidence that it can be strengthened or weakened. Can other capacities? First, we should consider what this means. It does not mean that the mass of an object, which brings with it this capacity, can be increased or decreased but rather that the effect of a given mass can be changed.

This kind of change, which may seem strange in fundamental physics, is commonplace to economists. Economic relations are now almost universally expressed in equations and the equations tend to be linear in their variables (though perhaps not in their parameters). For instance, here are some equations used by Kevin Hoover<sup>81</sup> to relate government spending at a time  $t$  ( $G_t$ ) to taxes at a time  $t$  ( $T_t$ ) and the rate of interest ( $R$ ):

$$G_{t+1} = \gamma + \delta[G_t - \gamma] + \varepsilon_{t+1}$$

$$T_{t+1} = T_t + [(R - 1)/(R - \delta)][G_{t+1} - \delta G_t + (\delta - 1)\gamma].$$

---

<sup>81</sup> Hoover (2001).

The Greek letters are parameters. Look for instance at  $\delta$ . It represents the strength of the capacity of  $G_t$  to influence  $G_{t+1}$ . It is typically supposed that the size of  $\delta$  can change. In fact the important point for Hoover of writing this as a parameter rather than as a variable (since he assumes that it can vary) is that *we* have it in our power to change it. Other economists may not be so sanguine about our powers to affect things, but the idea that parameters shift is universal. What that means is that the capacities involved can be enhanced or retarded for a *given* value of the quantity that brings the capacity with it.

For a less formal example, you might want to think about the familiar ‘crossed-sticks’ supply and demand curves. What do economists do with these? Typical questions are like this: What happens to price if the demand curve shifts upward or downward? If the supply curve shifts? Shifts in these curves are just shifts in the slope of the line, which are represented by the parameters in front of the variables representing supply and demand in the quantity equations; that is, they represent enhancements and retardations of the capacities of demand and supply to affect quantity produced.

The natural thought about the difference between the most fundamental capacities studied in physics and the capacities studied in economics is that the economic capacities are derived whereas those of fundamental physics are basic. Economic features have the capacities they do because of some underlying social, institutional, legal and psychological arrangements that give rise to them. So the strengths of economic capacities can be changed, unlike many in physics, because the underlying structures from which they derive can be altered.

Surely that is in some sense true. But there is a puzzle – at least if current economics practices are well-grounded. It is common in economics to assume that, for the most part, the parameters are independent of each other. Each can be altered leaving the others fixed. Hoover in fact takes this to be part of the characterization of a parameter. I say ‘for the most part’, but not ‘always’. When two parameters are not independent in this sense, however, there is a tendency (which is explicit in Hoover) to suppose that the two that are not independent can be written as functions of an overlapping set of parameters, each of which is independent of the others in the set. So if  $\theta$  can not be altered independently of  $\varphi$  then there is a set  $\{\alpha, \beta, \gamma, \dots, \chi\}$  such that  $\theta = f(\alpha, \beta, \gamma, \dots, \chi)$  and  $\varphi = g(\alpha, \beta, \gamma, \dots, \chi)$ , and every member of this set can be altered independently of every other.

How is this possible if the alterations of the parameters involve alterations of the underlying structure? Why is it typical rather than untypical, that we should be able to change the parameters one at a time when such changes involve us in mucking about with the underlying structure? We usually know

virtually nothing about this structure or how it operates. One would expect our interventions to be more like those with a sledge hammer than like surgical incisions. I shall leave this issue for discussion elsewhere since it seems more happily situated in a volume on philosophy of economics than in one on dispositions and powers.

There is one fact, however, that has come to the fore in our brief discussion of the malleability of economic capacities that is characteristic of dispositions in general. We see here in the economics case a threefold distinction that is typical for dispositions once they reside in even slightly complex systems and that is often conflated. There is i) the disposition, ii) the property in virtue of which the system has the disposition and iii) the underlying structure that ensures that that property is associated with that disposition.

#### 4. In defense of interference

I claim that it is characteristic of the capacities that science studies that they can all be interfered with. More strongly, the only way to state a true conditional with the canonical manifestation as consequent is roughly this:<sup>82</sup>

If the capacity is triggered properly and *is not interfered with*, then the canonical manifestation will result.

We are often told that in science we are not allowed to use terms like *interference*. Abstract or umbrella terms of this sort are admissible, but they must be accompanied by *bridge principles* that link them with more concrete terms which already have strict standards of application. The so-called ‘special force laws’ are the paradigm of bridge principles. The abstract force function  $F = GMm/r^2$  obtains when a mass  $m$  is situated a distance  $r$  from another mass  $M$ ; the abstract force function  $F = \epsilon_0 q_1 q_2 / r^2$  obtains when a charge  $q_1$  is located a distance  $r$  from a second charge  $q_2$ ; etc. We are also told that claims with the general rider ‘if nothing interferes’ in front are untestable, indeed vacuous – they allow anything. In closing I would like to make a number of remarks about these charges.<sup>83</sup>

- The absence of ‘special interference laws’ is not so epistemically damaging as many suggest. The special force laws do tell us when a particular force function obtains, but only for very specific descriptions – the descriptions that appear in our bridge principles. For other descriptions that may be applied far more immediately, such as *a truck passing by* or *the press of the wind*, we are just as much on our own

---

<sup>82</sup> My formulation here is similar to that of Joseph (1980).

<sup>83</sup> For more detailed discussions see Cartwright (1989) and Cartwright (1999).

without the help of a system of rules as we are in deciding if we can label the truck passing by as an *interference*.

- The lack of systematic rules does not mean that we cannot have knowledge about whether a certain kind of occurrence constitutes an interference. Galileo after all knew to use smooth planes for his rolling-ball experiments because he knew he should eliminate the interference of friction with the pull of the earth. Similarly he knew to drop small compact masses and not feathers from the Leaning Tower. And that was long before he could have had any idea whether friction or the wind exerted a *force* in the technical Newtonian sense.
- The fact that we cannot identify what counts as interference with respect to a claim does not mean that we cannot test whether that claim is true or not. Consider *Aspirins relieve headaches, if nothing interferes*. We regularly test claims like this in randomized treatment/control experiments.
- Nor does it mean that it is too easy to dismiss disconfirmations. When the predicted result fails to transpire, one can always *say* that something interfered. But saying does not make it true. And as epistemologists are always reminding us, saying, even when it is true, does not constitute knowledge, or even reasonable belief. We need a good reason for claiming that something is an interference. When we do not have any idea whether a nominated factor is an interference or not, then we equally have no idea how to classify the case. Our intended test is no test at all.
- It follows that one needs a great deal of information about what might and might not interfere with a process before we can carry out serious tests on the process and that in turn implies that we need already to have a great deal of information about the process itself. That just means that science is difficult, as we already knew, and that it is hard to get started in a vacuum of knowledge.

## 5. Conclusion

What makes everything in the disposition family belong there is, if anything, two-sidedness. Within this family ‘capacity’ seems especially like a power word. Nevertheless, I think that is the wrong way around to look at it. What marks out all capacities as capacities is not primarily that they *enable* systems to do things, but rather that they *can be stopped* – they can be interfered with. That does not rule them out of science, thus, whether something is an interference in a given situation is a matter of fact; and it is a fact we can

know about – though not in any mechanical way. But it is at any rate a big mistake to think that science could or should be mechanical.

### References

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

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

Chalmers, A. (2002), 'True Fundamental Laws in a Dappled, Patchwork World,' Popper Centenary Conference lecture, ms, Flinders University: Adelaide.

Earman, J. Roberts, J. and Sheldon S. '“Ceteris Paribus” Lost', *Erkenntnis*, 57, 3, pp. 281-301.

Ellis, B. (2001), *Scientific Essentialism*, Cambridge: Cambridge University Press.

Hampshire, S. (1972), *Freedom of Mind and Other Essays*, Oxford: Oxford University Press.

Hoover, K. (2001), *Causality in Macroeconomics*, Cambridge: Cambridge University Press.

Joseph, G. (1980) 'The Many Sciences and the One World', *Journal of Philosophy*, volume 77, 1980, pp. 773-791.

Menzies, P. (2002), 'Capacities, Nature's and Pluralism: A New Metaphysics for Science?', *Philosophical Books*.

Mill, J.S. (1843), *A System of Logic*, Book IV, London: Longman, Green.

Ryle, G. (1949), *The Concept of Mind*, London: Barnes and Noble.

Watkins, E. (2005), *Kant and The Metaphysics of Causality*, Cambridge: Cambridge University Press.

## Essay IV

### **Are RCTs the gold standard?\***

#### Abstract

The claims of RCTs to be the gold standard rest on the fact that the ideal RCT is a *deductive* method: if the assumptions of the test are met, a positive result *implies* the appropriate causal conclusion. This is a feature that RCTs share with a variety of other methods, which thus have equal claim to being a gold standard. This paper describes some of these other deductive methods and also some useful non-deductive methods, including the hypothetico-deductive method. It argues that with all deductive methods, the benefit that the conclusions follow deductively in the ideal case comes with a great cost: narrowness of scope. This is an instance of the familiar trade-off between internal and external validity. RCTs have high internal validity but the formal methodology puts severe constraints on the assumptions a target population must meet to justify exporting a conclusion from the test population to the target. The paper reviews one such set of assumptions to show the kind of knowledge required. The overall conclusion is that to draw causal inferences about a target population, which method is best depends case-by-case on what background knowledge we have or can come to obtain. There is no gold standard.

#### 1. Introduction

The answer to the title question, I shall argue, is ‘no’. There is no gold standard; no universally best method. Gold methods are whatever methods will provide a) the information you need, b) reliably, c) from what you can do and from what you can know on the occasion. Often Randomised Controlled Trials (RCTs) are very bad at this and other methods very good. What method best provides the information you want reliably will differ from case to case, depending primarily on what you already know or can come to know.

Since I have no expertise in psychiatry, I shall discuss methods in general use in the human sciences without trying to approach special problems of psychiatry. The paper will have six parts:

- I. Clinchers v Vouchers: A distinction and its implications
- II. A Straddler: The hypothetico-deductive method
- III. Examples of methods that clinch conclusions

---

\* I would like to thank participants of the BIOS ‘Searching for Gold Standards Conference’ June 2006 for their comments.

- IV. RCTs: Ideal RCTs, real RCTs and the scope of an RCT
- V. The vanity of rigor in RCTs
- VI. Closing remarks

Bits of part IV will rely on some formal results that I will present informally. I hope to convey a sense of the kind of information that is required to justify the claims of RCTs to be a gold standard as a basis for caution and for comparison with other methods that have an equal claim to this status (because they are what I shall call ‘clinchers’).

## 2. Clinchers v Vouchers: A distinction and its implications

Methods for warranting causal claims fall into two broad categories:

1. Those that *clinch* the conclusion but are *narrow* in their range of application, for example RCTs, derivation from theory or certain econometric methods.
2. Those that merely *vouch for* the conclusion but are *broad* in their range of application, for example qualitative comparative analysis, or looking for quantity and variety of evidence.

What is characteristic of methods in the first category is that they are deductive: *if* all the assumptions for their correct application are met, then if evidence claims of the appropriate form are true, so too will the conclusions be true. But these methods are concomitantly narrow in scope. The assumptions necessary for their successful application will have to be extremely restrictive and they can take only a very specialized type of evidence as input and special forms of conclusion as output. That is because it takes strong premises to deduce interesting conclusions and strong premises tend not to be widely true.

Methods in the second category are more wide-ranging but it cannot be proved that the conclusion is assured by the evidence, either because the method cannot be laid out in a way that lends itself to such a proof or because, by lights of the method itself, the evidence is symptomatic of the conclusion but not sufficient for it. What then is it to *vouch for*? That is hard to say since the relation between evidence and conclusion in these cases is not deductive and there are no general good practicable ‘logics’ of non-deductive confirmation, especially ones that make sense for the great variety of methods we use to provide warrant.

The fact that RCTs are a deductive method underwrites their claims to be the gold standard. But RCTs suffer, as do all deductive methods, from narrowness of scope. Their results are formally valid for the group enrolled in the study, but only for that group. The method itself does not underwrite any strong claims for external validity, that is for extending whatever results

are supposed to be established in the test population to other 'target' populations. This is important to keep in clear sight in comparing RCTs with other methods.

Compare then the costs and benefits of the two categories. Clinchers are deductive: *if* they are correctly applied *and* their assumptions are met, then *if* our evidence claims are true, so too will be our conclusions -- a huge benefit. But there is an equally huge cost. These methods are concomitantly narrow in scope. The assumptions necessary for their successful application a) tend to be extremely restrictive, b) can only take a very specialized type of evidence as input, and c) have only special forms of conclusion as output. In consequence we face a familiar kind of trade-off: We can ask for methods that clinch their conclusions but the conclusions are likely to be very limited in their range of application.

### 3. A Straddler: The hypothetico-deductive method

The hypothetico-deductive method is a straddler. Used one way – the way Karl Popper advocated – it is purely deductive and so is in the same category as the RCT. The method works, as all methods do, by presupposing a variety of auxiliary assumptions, otherwise nothing really follows from the hypothesis of interest.

Popper:

Hypothesis  $\rightarrow$  outcome

$\neg$ outcome

Therefore,  $\neg$ hypothesis

This is a clincher.

Positivists:

Hypothesis  $\rightarrow$  outcome

outcome

probability of the hypothesis increases (ceteris paribus)

This is a voucher.

Popper argued that the only correct use of the hypothetico-deductive method is as a clincher, to deduce that hypotheses are false. The argument accepted by the Positivists, he pointed out, is a deductive fallacy – the fallacy of affirming the consequent. And deductive logic, he maintained, is all the logic there is. This is borne out by centuries of failed efforts to establish some reasonable relatively uncontroversial theory of inductive confirmation. On the other hand, philosophers of physics maintain that the hypothetico-deductive method is the method by which physics theories are established.

Nevertheless, medical science – and most of current evidence-based policy rhetoric – will not allow it.

Perhaps an example related to topics of interest to psychiatry will help. Consider the widespread correlation between low economic status and poor health and look at two opposing accounts of how it arises. (For a discussion and references see Cartwright (2007).) Epidemiologist Michael Marmot from University College London argues that the causal story looks like this:

Marmot:

Low status → ‘stress’ → too much ‘fight or flight’ response → poor health

In contrast, Princeton University economist Angus Deaton suggests this:

Deaton:

Poor health → loss of work → low income → low status

Deaton confirms his hypothesis in the National Longitudinal Mortality Study (NLMS) data. He reasons: If the income-mortality correlation is due primarily to loss of income from poor health, then it should weaken dramatically in the retired population where health will not affect income. It should also be weaker among women than men, because the former have weaker attachment to the labour force over this period. In both cases these predictions are borne out by the data. Even more, split the data between diseases that something can be done about and those that nothing can be done about. Then income is correlated with mortality from both – just as it would be if causality runs from health to income. Also education is weaker or uncorrelated for the ones that nothing can be done about. Deaton argues that it is hard to see how this would follow if income and education are both markers for a single concept of socio-economic status that is causal for health.

Thus Deaton’s hypothesis implies a number of specific results that are borne out in NLMS data and would not be expected on dominant alternative hypotheses. So the hypothesis seems to receive positive confirmation, at least if we share the Positivists’ intuition. More carefully, it seems to receive some confirmation for the population sampled for the NLMS data. But what about other populations, i.e. what about *external validity*? The arguments I have just described that seem contra Popper to provide some evidence for Deaton’s hypothesis in the population sampled do nothing as they stand to support any claims about alternative populations. More premises and more and different arguments are needed to do that. So here we are reminded how badly even a non-clinching method can suffer from problems of external validity.

#### 4. Examples of methods that clinch conclusions

I list just a few other kinds of methods that work deductively.

1. Econometric methods
2. Galilean experiments
3. Probabilistic/Granger causality
4. Derivation from established theory
5. Tracing the causal process
6. Ideal RCTs

These are clinchers: It can be proved that if the auxiliary assumptions are true, the methods are applied correctly and the outcomes are true and have the right form, then the hypothesis must be true. Even though I do not have the space to discuss them here, I mention them in order to stress that when it comes to clinchers – to methods from which the hypothesis can be rigorously derived from the evidence – RCTs are not the only game in town. There are lots of methods that can clinch conclusions.

It is important to keep in mind one caution, however. To buy the benefits of a clinching method we must be able to ensure that it is highly probable that *all* the requisite premises are obtained. That's because of the *weakest link* principle for deductive reasoning. The probability of the conclusion can be no higher than that of the weakest premise.

- Suppose you have 10 premises, 9 of them almost certain, one dicey. Your conclusion is highly insecure, not 90% probable.
- In a deductive argument  $P(\text{conclusion}) \leq P(\text{conjunction of premises})$

I belabour this because of the benefits of clinching methods – clinchers are rigorous. It is transparent *why* the results are evidence: Given the background assumptions the hypothesis follows deductively from the results. And it is transparent *when* the results are evidence: When the background assumptions are met. This contrasts with ethnographic methods and expert judgment, for example. These can provide extremely reliable evidence. But there is no specific non-trivial list of assumptions that tell when they have done so. But if you want credit for this benefit of a clinching method, you must be able to show that the *conjunction* of your premises has high probability *in the case at hand*.

## 5. Randomised Controlled Trials

### *Ideal RCTs*

I have claimed that ideal RCTs are clinchers. That of course depends on how they are defined. But there are perfectly natural definitions from which it can be proved that RCTs, as thus defined, allow causal claims about the population in the study to be deduced from probability differences between the treatment and control groups<sup>84</sup>. The one I have worked with extensively is the probabilistic theory of causality, formalized by Patrick Suppes (1970) but widely adopted throughout the human sciences, even if not consciously so under that title. Suppes's concept of probabilistic causality is similar to the concept of Granger causality (Granger, 1969) that is frequently used in econometrics.

The root idea of the probabilistic theory of causality is that if the probability of an 'outcome' O is greater with a putative cause T than without T once all 'confounders' are controlled for in some particular way, that is sufficient for the claim 'T causes O' in that particular setting of confounding factors. So, in a population where 'all other' causes of O are held fixed, any difference in probability of O with T present versus with T absent shows that T causes O in that population. The rationale supposes that differences in probability need a causal explanation and if all explanations relying on confounders are eliminated, then T causes O is the only explanation left. T must be causing O in at least some members of the population in order to account for the difference in probability. I should note that whether one wishes to adopt the theory in exactly this form, some such assumption is necessary to connect causes and probabilities if we are to suppose that the probabilistic observations in RCTs can yield causal conclusions.

The definition so far only tells us when we can assert that T causes O for populations that have some fixed arrangement of 'all other' causal factors. To get a more general conclusion we may accept as well that if T causes O in a subpopulation of a given population  $\phi$ , then T causes O in  $\phi$ . This is consistent with my suggestion in the last paragraph that on the probabilistic theory of causality when we say T causes O in a population we mean that T causes O in at least some members of that population.

The proof that positive results in an ideal RCT deductively imply that the treatment causes the outcome would go something like this: To test 'T causes O' in  $\phi$  via an RCT, we suppose that we study a test population  $\phi$  all of whose members are governed by the same causal structure, CS, for O and which is described by a probability distribution P. P is defined over the event

---

<sup>84</sup> Cf. Cartwright (1989), Holland and Rubin (1988) and Heckman (2001).

space  $\{O, T, K_1, K_2, \dots, K_n\}$ , where each  $K_i$  is a state description over ‘all other’ causes of  $O$  except  $T$ .<sup>85</sup> The  $K_i$  are thus maximally causally homogeneous subpopulations of  $\phi$ . Roughly,

- ‘ $K_i$  is a state description over other causes’ =  $K_i$  holds fixed all causes of  $O$  other than  $T$ .
- ‘Causal structure’ = the network of causal pathways by which  $O$  can be produced, with their related strengths of efficacy.

Then assume

1. *Probabilistic theory of causality.*  $T$  causes  $O$  in  $\phi$  if  $P(O/T \& K_i) > P(O/\neg T \& K_i)$  for some subpopulation  $K_i$  with  $P(K_i) > 0$ .
2. *Idealization.* In an ideal RCT for ‘ $T$  causes  $O$  in  $\phi$ ’, the  $K_i$  are distributed identically between the treatment and control groups.

From 1 and 2 it follows that ideal RCTs are clinchers. If  $P(O)$  in treatment group  $>$   $P(O)$  in the control group in an ideal RCT, then trivially by probability theory  $P(O/T \& K_i) > P(O/\neg T \& K_i)$  for some  $K_i$ . Therefore: if  $P(O)$  in treatment group  $>$   $P(O)$  in control group,  $T$  causes  $O$  in  $\phi$  relative to CS,P.

What is going on here? We suppose that increase in probability of  $O$  with  $T$  does not show that  $T$  causes  $O$  in an arbitrary population. But it does in a maximally causally homogeneous population. We of course are almost never in a position to identify what makes for a maximally homogeneous population, so how can we tell whether  $T$  increases the probability of  $O$  in some one of these? The RCT is a clever way to find out. The RCT tells us that in some one or another maximally causally homogeneous subpopulation of the population in the study,  $T$  does increase the probability of  $O$ . Given the probabilistic theory of causality that tells us that  $T$  causes  $O$  in that subpopulation. So, what is established in the ideal RCT according to the account based on probabilistic theory of causality is that  $T$  causes  $O$  in at least one maximally causally homogeneous subpopulation of  $\phi$ . We may say we have established ‘ $T$  causes  $O$  in  $\phi$ ’ and that is a fine way to talk, so long as we recall that this means that  $T$  causes  $O$  in some subpopulation of  $\phi$ .

It is important to notice that on this account ‘ $T$  causes  $O$  in  $\phi$ ’ is consistent with ‘ $T$  causes  $\neg O$  in  $\phi$ ’. This lines up with what we know of RCTs:

- RCTs deliver population-average results. A *positive* result shows that  $T$  causes  $O$  in at least one subpopulation. It could produce exactly opposite results in other subpopulations.

Formatted: Bullets and Numbering

<sup>85</sup> This must include ‘spontaneous generation’. More formally,  $K_i$  holds fixed one variable on each pathway that does not go through  $T$ , as judged by the causal structure CS.

- Positive results are conclusive but negative are not: Equal probability for O in the treatment and control groups does not show that T does not cause O in  $\phi$ . It shows that if T causes O in  $\phi$  (because it does so in some  $K_i \subseteq \phi$ ) it must also cause  $\neg O$  (because it does so in some other  $K_i \subseteq \phi$ ).

### *Real RCTs*

So, from positive results in an ideal RCT for ‘T causes O in  $\phi$ ’ we can deduce that the causal hypothesis is true. But we can be no more certain of our causal conclusion than we are of our premises, to wit, that the RCT is ideal and that the probability of O is indeed higher with T than without in the test population. What do we do to ensure the premises? Here are just some of the principal precautions we take: careful use of statistics to move from frequencies to probabilities, ‘random’ assignment to treatment and control groups, quadruple blinding, careful attention to drop-outs and non-compliance, and so on.

I mention them just to point out that the practical methodology must match and be matched with the kind of formal treatment I have outlined. RCT advocates claim that RCTs are extremely reliable if carried out properly. That claim can be justified by an account of the kind I have outlined. But then – what is justified is that positive results *as defined by the account*, in an ideal RCT *as defined by the account*, imply a causal conclusion *of the kind defined by the account*. The practical methodology then must be geared to ensuring that the premises required by the formal account are very likely to be true; and the conclusions drawn can only be of the kind admitted by the account. Of course the converse holds as well: A formal account that does not match well with our most careful, most well thought-out practical methodology should be viewed with at least a little suspicion.

### *The scope of an RCT*

Starting as I have from the probabilistic theory of causality there are two kinds of causal conclusions we might naturally try to export from an RCT to some target population  $\theta$ :

1. T causes O in  $\theta$ . That is, T causes O in at least some members of  $\theta$ .<sup>86</sup>
2. Some measure of ‘average improvement’ that holds in the experiment will hold in the target population. I shall consider the simple case of  $P(O/T) > P(O/\neg T)$ .

Both conclusions need strong auxiliary assumptions to be warranted, well beyond those supported by the structure of the RCT. For the first, the RCT

---

<sup>86</sup> In the ‘long run’ of course since all results are probabilistic.

shows that T causes O in at least some members of some fixed causally homogeneous subpopulations. So to draw conclusions that T causes O in at least some members of  $\theta$ , we need at least these kinds of assumptions:

- Auxiliary 1.a. At least one of the subpopulations (with its particular fixed arrangement of ‘other’ causal factors) in which T causes O in  $\phi$  is a subpopulation of  $\theta$ .
- Auxiliary 1.b. The causal structure and the probability measure is the same in that subpopulation of  $\theta$  as it is in that subpopulation of  $\phi$ .

For the second we need to show that the outcome is more probable with T than without in  $\theta$ .<sup>87</sup> The simplest guarantee for this is

- Auxiliary 2. The causal structure (CS) and the probability (P) are the same in  $\theta$  as in  $\phi$ .

There are an indefinite number of other ways that guarantee  $P(O/T) > P(O/\neg T)$  in  $\theta$  given it holds in  $\phi$ , depending on the exact strengths of efficacy and the exact probabilities involved. But this is the only rule that does not require explicit statement of the specific numbers, most (if not all) of which are unknown to us. To get a sense for this, just imagine a case where there are only two relevant subpopulations, in one of which T is strongly positive for O and in the other it is equally strongly negative. The results will be positive in the RCT if the first subpopulation is more probable than the second, but will be reversed in targets where the second outweighs the first even if the new population has the same causal structure as the test population. Clearly if the causal structure differs, matters will depend on just how, just as the net result will depend on just what the probabilities are if the probabilities of the relevant subpopulations differ.

The central question for external validity then is, ‘How do we come to be justified in the assumptions required for exporting a causal claim from the experimental to a target population?’ Here rigor gives out. This is not to say that we do not have procedures or that we do not proceed in an intelligent way. We could aim to draw the test population ‘randomly’ from the target. We know that this is almost never possible. Moreover, we must not be deluded about sampling methods: You cannot sample randomly without any idea what factors are to be equally represented – which is just the issue that drives us to RCTs to begin with. One thing we certainly can do is to try to take into account all possible sources of difference between the test and target populations that we can identify. This is just what we do in matched observational studies. When it comes to internal validity, however,

---

<sup>87</sup> Or as near enough as matters for our purposes. I shall here ignore these niceties and how to treat them in order to focus on the main point.

advocates of the exclusive use of RCTs do not take this to be good enough – matching studies are not allowed just because our judgements about possible sources of difference are fallible. Yet exactly the same kinds of ‘non-rigorous’ judgements are required if RCTs are to have any bearing outside the test population. For an RCT the reliability of the claims in the target population is only as good as our estimates that very demanding auxiliaries like those above are met. The question then is about the trade-off between internal and external validity.

*Lesson.* We experiment on a population of individuals whom we take to have the same *fixed causal structure* (albeit unknown) and *fixed probability measure* (albeit unknown). Our deductive conclusions depend on that very causal structure and probability. How do we know what individuals beyond those in our experiment this applies to? We have seen some typical auxiliary assumptions about target populations that allow us to export conclusions from the experimental population to a target population and we have seen that these assumptions are very demanding, demanding of information that is not supplied by the RCT and that is hard to come by. But our conclusions about the target can be no more certain than these auxiliary assumptions. The RCT, with its vaunted rigor, takes us only a very small part of the way we need to go for practical knowledge. This is what disposes me to warn about the vanity of rigor in RCTs.

## 6. The vanity of rigor in RCTs

The title is borrowed from my paper ‘The Vanity of Rigor in Economic Models’ (Cartwright (forthcoming)). In both cases we see identical problems: that of internal versus external validity. Economists make a huge investment to achieve rigor *inside* their models, that is to achieve internal validity. But how do they decide what lessons to draw about target situations outside from conclusions rigorously derived inside the model? That is, how do they establish external validity? We find: thought, discussion, debate; relatively secure knowledge; past practice; good bets. But not rules, check lists, detailed practicable procedures; nothing with the rigor demanded inside the models.

And RCTs? If we compare them with economic models on internal validity, economic models have the advantage: we can readily see when the results are internally valid in an economic model just by inspecting the derivation. This is clearly not so with RCTs. Consider the equal distribution of ‘other’ causal factors. Once we check the causes we know about, we have no further evidence that our precautions, our quadruple blinding and random assignment and so forth, indeed result in an equal enough distribution. And we know lots of things can go wrong. The best we can do is for people

expert at what could go wrong to have a very close look at what actually happens in the experiment.

It is important though that these are not people like me (or independent experimental-design firms) who know only about methodology, but rather people with *subject-specific knowledge* who can spot relevant differences that come up. But this introduces *expert judgement* into the assessment of internal validity, which RCT advocates tend to despise. Without expert judgement, however, the claims that the requisite assumptions for the RCT to be internally valid are met depend on fallible mechanical procedures. Expert judgements are naturally fallible too, but to rely on mechanics without experts to watch for where failures occur makes the entire proceeding unnecessarily dicey.

This brief mention of economic models versus RCTs highlights the conventional trade-off I recalled at the start between internal and external validity. Despite the claims of RCTs to be the gold standard, economic models have all the advantages when it comes to internal validity. As I remarked, we need just mathematics and logic to decide if the conclusions are internally valid, whereas RCTs need a number of demanding assumptions beyond valid reasoning. But it seems that RCTs have the advantage over economic models with respect to external validity. Surely no matter what the target population, people in experiments are more like people in the target population than people in models are. Even here there is a caution, however, for of course this claim depends on exactly what kind of knowledge about people in the target population we build into the construction of our experiments versus how much we build into our models, and how we do so.

## 7. Closing remarks

I close with some reminders for those who advocate RCTs as the gold standard:

The method of our most successful science – the h-d method – is not a clincher at all. (And we do have some biomedical theory!)

There are many other clinching methods. Which method provides the most secure conclusions in a given case depends entirely upon which kinds of premises we can be most secure about and the situation at hand.

An argument that certain procedures achieve a given result much of the time may not be a good argument that they do so on any one occasion.

External validity for RCTs is hard to justify. Other methods, less rigorous at the front end, on internal validity, can have far better warrant at the back end, on external validity. We must be careful about the trade-offs. There is no a priori reason to favour a method that is rigorous part of the way and very iffy thereafter over one that reverses the order or one that is less rigorous but fairly well reasoned throughout.

### References

Cartwright, N. (2007), *Hunting Causes and Using Them*, Cambridge; Cambridge University Press.

Granger, C. (1969), 'Investigating Causal Relations by Econometric Models and Cross-Special Methods', *Econometrica*, 37, 424-438.

Heckman, J. (2001), 'Econometrics, Counterfactuals and Causal Models', Keynote Address, International Statistical Institute. Seoul, Korea.

Holland, P. W. and Rubin, D. B. (1988), 'Causal Inference in Retrospective Studies', *Evaluation Review*, 12, 203-231.

Suppes, P. (1970), *Probabilistic Theory of Causality*, Atlantic Highlands, N.J.: Humanities Press.

## Essay V

### **Economic Models: No Capacities, No Inductions**

#### Abstract

This paper argues that even when models are taken to picture parallel worlds, they still serve as isolating tools, but there are special problems in seeing them as isolating tools. These are common problems that beset any model that functions as a thought experiment. These problems are especially pressing for economic models however, because of the paucity of economic principles economic models are rich in structural assumptions. Without these, no interesting conclusions can be drawn. This makes a pressing problem for exporting conclusions from the model to the world. One uncontroversial constraint on induction from special cases is to beware of extending conclusions to situations we know are different in relevant respects. In the case of economic models, it is clear by inspection that the unrealistic structural assumptions of the model are intensely relevant to the conclusion. Any inductive leap to a real situation seems a bad bet.

#### 1. Introduction

The topic of our symposium is ‘Economics Models: Isolating Tools or Credible Parallel Worlds?’ In this discussion the kinds of models I have in mind are ones that picture simple analogue economies. They take over from the real world specific features we are interested in, i.e., but the features are not pictured in any real world setting and the analogue economies of the models are fantastically sparse compared to real world economies. They have few agents and few features, and few underlying processes are at work. We sometimes say that they are hugely ‘idealized’. This might though suggest that there is some specific target situation from which they are idealized, or to which they are intended to be de-idealized. I think that most often there is no such specific target in view. Rather, with our simple analogue economies we aim to study a specific regularly appearing aspect of a specific kind of phenomenon, not a specific target situation. So I prefer to say the models are ‘simple’ not ‘simplified’, or better: ‘sparse’.

‘Isolating Tools’ are meant to discover how capacities operate. The *capacity* associated with a feature is a power that systems with that feature have to produce a specific result characteristic of the capacity. An important feature of capacities, as I use the term, is the three-fold distinction between the *obtaining* of the capacity, its *exercise* and the *manifest results*. Consider gravity, i.e., the capacity to attract massive objects an object of mass  $M$  attracts another object of mass  $m$  a distance  $r$  away with force  $GMm/r^2$ . A well established empirical law describes *when* this capacity obtains – it

obtains in any system that has mass  $M$ . The production of the gravitational force,  $F_g = GMm/r^2$ , constitutes the *exercise* of the capacity. Some capacities need triggering before they are exercised but not gravity. It is exercised whenever it is present. That of course does not fix the *manifest results* – the motions. What motions result when the capacity is exercised will depend on what other forces are at play and on what the initial positions and velocities are.

The laws describe what other features (e.g., having mass  $M$ ) are universally or probabilistically associated with the obtaining of the capacity are empirical, but the connection between the capacity and its exercise is analytic: so long as the capacity really obtains, it is bound to be exercised whenever appropriately triggered. This is what makes it so valuable to learn what the capacities empirically associated with different economic factors are. Models that are simple in just the right ways, I shall argue, can be good for exhibiting characteristic effects of a capacity by showing what the capacity does ‘on its own’, without effects of other ‘confounding’ factors.

A *credible world*, as I understand it, is a world that contains features that occur in the real world in arrangements consistent with constraints of certain real world institutional structures, behaving in ways dictated by principles that are at least sometimes true in those structures. Consider as an example, Thomas Schelling’s famous (1978) checkerboard model of segregation. Black and white checkers are distributed randomly on a board, with spaces left empty. The dynamics of the model allow the checkers to move in a certain way in accord with their preferences concerning the colour make-up of neighborhoods they live in. As two London colleagues of mine (Romans Pans and Nicolaas Vriend) report in a recent follow up study, the model shows how an integrated board “*could unravel*” (Pans and Vriend 2007, p. 1) into a segregated one, notwithstanding the fact that no one prefers to live in a segregated neighborhood. Indeed to the contrary, the only preference of the agents is not to live in a neighborhood where their own colour is heavily underrepresented. The checkerboard model describes, I believe, a *credible world*, as I have characterized it, albeit an incredibly special one with essentially no probability of ever being actual.

In the debate, ‘Economics Models: Isolating Tools or Credible Parallel Worlds?’ I am supposed to be on the ‘isolating tools’ side. In a sense that is true, for I think that even for Robert Sugden’s credible worlds account, models must be studying something about capacities. For all of his exemplars focus on one or at most a handful of factors operating ‘in isolation’. This is precisely why I remarked about the Schelling model that the only preference operating for the checkers is their desire not to live in a neighborhood where they are badly outnumbered. Though we know that in the real world this would never be the only preference affecting choice of

neighborhoods in a move, nevertheless we suppose that we can learn something of importance about the real world from the model that studies the effects of this single preference. But to make the assumption that looking at how a factor behaves ‘in isolation’ (i.e., when it is the only cause of its type at work), whether in a model or in a real experiment, can help us understand non-experimental outcomes where a large number of causes are at play together, is to rely on the logic of capacities.

On the other hand, I am going to argue, we can not learn much about capacities from analogue-economy models. This could be true even of the analogous kinds of models in physics and even of real experiments. But it is especially troubling in economics for two reasons I shall describe. The first I have raised before (Cartwright 2007 and 1999): it is the paucity of acceptable economic principles to import into the models. The second is the special way in which economic capacities work.

Early work by me and others about what I call ‘capacities’ in economics was heavily based on Mill on tendency laws, both from Mill’s *System of Logic* and from his *Definition of Political Economy*: I think especially of my 1989 *Nature’s Capacities and their Measurement* and Hausman’s 1991, *The Inexact but Separate Science of Economics*. I now think that at the time I paid too much attention to these theoretical works of Mill’s and not enough to his more practical discussion of tendencies, where exactly my newer problem is central, especially in his work on the subjugation of women.

## 2. Analogue-economy models and Galilean experiments

How are analogue-economy models supposed to teach us about capacities? Answer: by mimicking Galilean experiments. By ‘Galilean experiment’ I mean one which isolates the cause under study so that it operates ‘without impediment’. What happens in the experiment then is the exercise of that capacity and of that capacity alone.

In many cases we can think of models as thought experiments designed to capture the requisite features of a real Galilean experiment. The cause under study and only that cause is written into the model: the cause operates ‘on its own’ or ‘without impediment’. In a real experiment Nature produces the effects of this cause in accord with systematic principles she adopts for the situation. In the model we produce the effects by deduction from the principles we adopt in the model. The result in real experiments is the exercise of the capacity associated with the cause as dictated by Nature’s principles; what results in the model is the exercise of the capacity as dictated by our principles. The results in the model replicate those of Nature if our principles are close enough to hers.

*Example, physics: the orbit of the earth due to the sun.*

Using Newton's laws we calculate the orbit of a small body in the vicinity of a larger, where the initial velocity of the smaller is orthogonal to the line connecting them. We conclude that the capacity of the larger to attract the smaller moves the smaller in an elliptical orbit around the larger. We read this as a capacity conclusion: we expect the elliptical orbit to contribute to the final trajectory in any real case but not to duplicate it since in real cases other factors will affect the trajectory as well.

*Example, economics: the effect of skill loss during unemployment on future unemployment.*

Using the principle that agents maximize their utility in the face of the decisions of others, Pissarides (2000) calculates that for an entrepreneur whose utility consists in profits and workers who care about wages and leisure, loss of skill by unemployed workers will perpetuate high levels of unemployment. We read this as a capacity claim: we do not expect that this necessarily happens in real cases since other factors affect unemployment levels as well.

Real experiments give wrong results if we do not succeed in eliminating all confounding factors. This shouldn't happen in the thought experiment where confounding factors can only appear if we put them in. On the other hand, in the real experiment the result is bound to be what Nature dictates whereas that will only be true in the thought experiment if we have our principles right. This is a trade-off between real and thought experiments.

What I want to talk about today is a problem that can beset real and thought experiments alike and in both physics and in economics. But it is a particular plague for thought experiments in economics, I shall argue, so much so that it regularly undermines the use of models to establish capacity claims. That is the problem of *overconstraint*.

### 3. Overconstraint

Economics has very few uncontroversial principles at its disposal. Its models must do a lot with a little. In both Pissarides' skill-loss model and Schelling's segregation model the only principle at work is that agents maximize their utility in the face of the options and the decisions of others. Contrast physics where a rich network of laws is available to import into models. Look at the sun/earth example where the combination of Newtonian laws is *enough*. By themselves they imply that the smaller body with initial velocity orthogonal to the line between the two bodies will execute an ellipse.

The Pissarides model is different. The desire for profit or for wages and leisure results in nothing without the addition of a great deal of structure.

Notably we need some kind of matching technology to put workers into jobs. But also there are just two generations of workers, all workers become unemployed at the end of the first period, etc. etc. And these assumptions play a role in the derivations that Pissarides makes. In a real experiment there will of course always be an infinitude of such additional factors; but in a real experiment they do not play a role. They obtain but they exert no influence on the outcome. The opposite is the case in most economic models. We can tell by inspection that they *are* relevant to the derivation of the results.

In consequence the results of these models are over constrained. So long as the models satisfy all the conditions of a Galilean experiment, the results that obtain in the model will be results that would be produced in *a* Galilean experiment. But they are not the general result that would be produced in *any* Galilean experiment. We see a genuine exercise of the capacity but a very special case of it. In the Pissarides model skill loss causes unemployment levels to stay up. It *could be* that the general result is that they either stay up or go down.

This seems to be just what happens in the Schelling model. Mere preferences about colour mix in neighborhoods cause absolutely nothing by themselves. Some structure for the neighborhoods is required -- like the checkerboard -- to execute moves as well as some dynamics to allow action on the preferences.

Economists like Pissarides or Schelling and successive thinkers are of course not blind to the possible effects of what I (but not they) call 'over constraining' assumptions, and this is a case in point. Pancs and Vriend investigate what happens with different moving technologies, different neighborhood structures and different specifications of preferences. They show that on a checkerboard segregation results across a variety of other differences so long as agents favor their own ghetto over others, but segregation disappears when agents have preferences like an inverted V: they prefer totally integrated neighborhoods and their preference decrease equally as level of integration decreases in either direction. But the same is not true on a torus, given the neighborhood structure Pancs and Vriend construct there: strict preference for integration leads to 'remarkably extreme segregation in finite time'. (p 36)

It is important to stress that in all cases we are seeing the exercise of the preferences but the manifest results depend intimately on 'extraneous' factors -- factors beyond those that define a Galilean experiment. The problem is that we cannot eliminate these; in contrast to the model for the sun and the earth, without the over constraining factors no manifest results follow at all. But with them, the results differ from case to case.

#### 4. What is special about economics?

The need for overconstraining factors is widespread in economic thought experiments but not in those of physics. I have already explained one reason why: the paucity of acceptable principles in economics. If you want to deduce conclusions, you need premises. Ideally besides the specific description of the cause whose capacities we study, the only premises in use should be general principles and assumptions that guarantee that the experiment is indeed Galilean. But we do not have many principles in economics to include, so we make do with overconstraining structural assumptions. But then we can read out only special-case conclusions, not general claims about the manifest results of the capacity.

A second reason is the peculiar nature of the capacities at work in economics. The first is a reason I have stressed for a long time. The second is one I have overlooked, in part due to Mill himself. Mill likens economics to mechanics. There are a variety of *causes* -- like gravity, electromagnetic attraction and repulsion, etc. --and each cause has its own *tendency law* describing what capacity will be exercised when it obtains: the law of gravity, Coulomb's law and the like. Then there is a *law of composition* that dictates what manifest results occur when capacities associated with a number of different causes are all exercised at once.

The trouble for economics with the mechanics analogy is that it supposes that everything that affects the manifest result in question is a cause of the same kind: each is an independent cause that has its own tendency law whose result will get treated in the law of composition. This seems to work in mechanics, where we make the assumption that everything besides initial position and velocity that affects a motion is a force and all the forces have their own tendency laws and all their consequences are handled by the law of composition. In this case the notion of a Galilean experiment makes perfect sense: in the experiment the force under study is to be introduced and all other forces are to be eliminated.

This idea falls apart in typical economics cases. Preferences to live in neighborhoods of given colour mixes can have fixed capacities. Trivially these preferences are exercised when they are acted on. But the manifest results do not depend only on other causes with their own tendency laws dictating their own consequences that can be treated in a law of composition, and this is so no matter how sophisticated we get about laws of composition. In reality the manifest results depend on structural circumstances and not just on the exercise of other capacities, even in principle. So it is not a surprise that the same should be true in the model.

I think this point is behind Anna Alexandrova's criticisms of the attempt to

draw capacity conclusions from ideal models but I haven't seen it so clearly before now. It is also apparent in Mill's own work on psychology and his quarrels with August Comte, about which I have learned a great deal from Vincent Guillin. Women, Mill argues, naturally have a capacity for independent and imaginative thought. But the capacity will not display itself in the 'expected' manifestations unless it is nurtured, trained and allowed to display itself freely, as it is with middle class Englishmen.

In the Schelling models, I argued that total segregation is just as much a manifest result of the preference for integration as is total integration. Mill argues the same with respect to the subjugation of women. His opponents claimed that women through the ages have lacked curiosity, independence of thought, intellectual drive, etc. If so, Mill claimed, that is how their natural capacities play out in the uninviting structures in which they were set and is in no way indicative of what can result in more felicitous circumstances. Women cannot be independent of mind without the natural capacity. But our failure to exhibit independent thought -- even a (contrary to fact) universal failure ever to do so -- would not show that the capacity is lacking.

### 5. Credible Worlds

Models' results, I have argued, are frequently overconstrained by structural assumptions, and often necessarily so. I have raised this as an objection to the view that models can act as Galilean thought experiments, thought experiments to investigate the 'natural' manifestations of capacities, that is, the manifestations that result purely from the exercise of that capacity alone. But the problem is equally pressing for the use that Sugden wants to put his account of models as images of credible worlds. How do we learn about the real world from the models under the two views? I say that we hypothesize *on other grounds* that the factor in the model has a (relatively) fixed capacity. That means that what we see when the factor acts on its own is what it will 'contribute' (given the relevant law of composition) when it acts in consort with other factors. Then we use the model to assess what this contribution is. So on this account the license for moving from the results in the model to some claims about the world rests on two assumptions: that the factor in question has a (relatively) fixed capacity and that the model has the right characteristics to count as a Galilean thought experiment that can reveal the natural manifestation -- the 'contribution' -- of that capacity. My worry that in very many cases purely structural constraints, not justified by the demands for a Galilean experiment, matter to the outcome threatens the second assumption.

Sugden proceeds differently. As I indicated, I think he is still looking only for 'contributions' since he surely does not think that what results in models that isolate a single factor or single complex of factors is what will happen when

lots of other causally potent factors are at work as well. But the logic does not start with the assumption – which must be defended from elsewhere – that the factor in question has a (relatively) stable contribution. Rather he imagines a cautious induction from one clear case. The model describes a perfectly credible world: it could be real. Then we do an induction from that either to what contribution the factor makes in general or to what contribution it will make in a targeted application. The license for moving from the model to other cases is not the assumption that the factor has a capacity but rather induction from a carefully thought about exemplar.

The inductive base is slim but let us lay this aside. (Perhaps this is like what Peter Galison describes as ‘the golden event’ in physics; as soon as we see it we rightly feel confident without the need of a huge number of further instances.) The license to make the inductive leap is threatened in exactly the same way that the inference to the ‘natural’ manifestation of the capacity is. One thing we know about induction, which Till Grüne-Yanoff stresses in this symposium, is that we should not do it when we have good reason to think there are differences that matter between the inductive base and the target. And that is just what I have been worrying about here. In many models we *know* by inspection that assumptions about the structure matter. Not only do they play a necessary role in the derivations we actually do in the models, in many cases – as in the Pancs and Vriend discussion of Schelling segregation models – we know that with different structures we get different results. Induction then is a very bad idea.

But what about robustness results? Economists often make great efforts to show that ‘the same’ results obtain under different assumptions. This is a complicated issue that deserves a great deal more study than it receives. From a logical point of view it seems to me that it does not make sense in the cases at issue – cases where the extra overconstraining assumptions are necessary in the proofs on offer. In these cases robustness looks like this:

$C \& X \rightarrow R$

$C \& Y \rightarrow R$

$C \& Z \rightarrow R$

Therefore  $C \& \alpha \rightarrow R$ , for all  $\alpha$ ? Some  $\alpha$ ? Some  $\alpha$  in some specific but perhaps unspecified range? ...?

What kind of logic is this?

From the point of view of some ‘inductive’ logic, perhaps it looks less problematic. 17-year old boys with red hair are interested in sex, 17-year old boys with brown are interested in sex, 17-year old boys with black hair are interested in sex. Therefore, more or less all 17-year old boys are interested in sex. Or: therefore, ginger-haired boys are interested in sex. Maybe this

inference is not too suspect. But that's because we think hair colour is irrelevant to the result. Exactly the opposite is the case in our examples. I repeat, in analogue-economy models we can tell by inspection that the overconstraining factors are relevant to the result. This is the feature that I have urged to be fairly typical of model derivations: structural factors beyond those demanded to ensure a Galilean experiment are necessary to derive the results. This is the case whether or not the results are always 'the same', as in successful robustness explorations, or highly different, as in cases like the Pancs and Vriend study of the Schelling model that I have been stressing here.

So...without robustness results it is hard to see how Sugden can be entitled to an induction since we can *see* that the extra assumptions matter. But even with a great many robustness results, the logic is questionable. And we should keep firmly in mind that, though each world figuring in the robustness investigation may be credible, they are all hugely different from almost all real cases, and in each case different in ways we know matter, unlike the relevance of hair colour to the interest in sex that develops among 17-year old boys.

## 6. Conclusion

Analogue-economy models may picture Galilean thought experiments or they may describe credible worlds. In either case we have a problem in taking lessons from the model to the world. The problem is the venerable one of unrealistic assumptions, exacerbated in economics by the fact that the paucity of economic principles with serious empirical content makes it difficult to do without detailed assumptions. But the worry is not just that the assumptions are unrealistic; rather, they are unrealistic in just the wrong way. In the case of Galilean thought experiments they are overconstraining; so the models reveal not the true contribution but a far narrower one that can be extremely misleading. In the case of credible worlds the assumptions provide good reason not to make the very kinds of inductions Sugden hopes for; we know just by inspection of the derivation that there are features that matter to what result obtains in the model world that will not be shared by real target applications.

Where does that leave us? Anna Alexandrova takes a minimalist view: the model shows that *there are* (or perhaps, can really be) situations where the factor investigated makes the contribution that occurs in the model. Clearly (if the principles of the model are correct) situations where all the relevant factors are the same as in the model are among these. Many other different situations may be as well, but typically this is not information one can read out from the model itself.

I close with a proposal for how we might do better that grows out of the concerns I have raised. We can, as Mary Morgan urges, probe the models and we often do. We change the models, experiment on them, see what results as assumptions are varied in relevant ways. Much of this probing gets described, as I mentioned, as ‘robustness testing’. This suggests that we are aiming to establish that the same contribution results across a variety of structural changes. But as I have urged, the logic of robustness reasoning does not seem to hold water. I suggest that instead we probe models as a means to understand *how* structure affects the outcomes. This leaves us with a new research problem in methodology. We can extrapolate from models to the world if probing the models leads us to understand how circumstances shape the contributions that economic factors make. What considerations should guide this kind of probing and what forms will the general results take?

### References

Alexandrova, A. (2007), *Making Models Count*, *Unpublished*.

Alexandrova, A. (2006), “Connecting Economic Models to the Real World”, *Philosophy of the Social Sciences* 36: 173-192.

Cartwright, N. (2007), *Hunting Causes and Using Them* Cambridge University Press.

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

Cartwright, N. (1989), *Nature’s Capacities and their Measurement*. Oxford University Press.

Galison, P. (1987), *How Experiments End*. Chicago: Chicago University Press.

Grüne-Yanoff, T. (2006), “Learning from Economic Models”, *Philosophy of Science* 73 (Proceedings).

Guillin, V. (2006), *Auguste Comte and John Stuart Mill on Sexual Equality: Historical, Methodological and Philosophical Issues*. Ph.D. Dissertation. London, U.K.: University of London, London School of Economics and Political Science.

Hausman, D. M. (1992), *The Inexact but Separate Science of Economics*. Cambridge: Cambridge University Press.

Mill, J. S. (1848), *Principles of Political Economy, with some of their*

*Applications to Social Philosophy*. London: J. W. Parker.

Mill, J. S. (1850), *A System of Logic, Ratiocinative and Inductive; being a connected view of the principles of evidence and the methods of scientific investigation*. New York: Harper.

Morgan, M. (2001), 'Models, Stories and the Economic World', *Journal of Economic Methodology* 8:3, 361-384.

Pancs, R. and Vriend, N. J. (2007), 'Schelling's Spatial Proximity Model of Segregation Revisited', *Journal of Public Economics*. 91: 1-24.

Pissarides, C. A. (2000) *Equilibrium Unemployment Theory*. Cambridge, Massachusetts: MIT Press.

Schelling, T. (1978), *Micromotives and Macrobehavior*. New York: Norton.