Manipulationism, *Ceteris Paribus* Laws, and the Bugbear of Background Knowledge

Robert Kowalenko
University of the Witwatersrand
Johannesburg, South Africa

This is an accepted manuscript of an article to be published by Taylor & Francis in *International Studies in the Philosophy of Science*

ABSTRACT According to manipulationist accounts of causal explanation, to explain an event is to show how it could be changed by intervening on its cause. The relevant change must be a 'serious possibility' claims Woodward 2003, distinct from mere logical or physical possibility—approximating something I call 'scientific possibility'. This idea creates significant difficulties: background knowledge is necessary for judgments of possibility. Yet the primary vehicles of explanation in manipulationism are 'invariant' generalisations, and these are not well adapted to encoding such knowledge, especially in the social sciences, as some of it is non-causal. *Ceteris paribus* (*CP*) laws or generalisations labour under no such difficulty. A survey of research methods such as case and comparative studies, randomised control trials, ethnography, and structural equation modeling, suggests that it would be more difficult and in some instances impossible to try to represent the output of each method in invariant generalisations; and that this is because in each method causal and non-causal background knowledge mesh in a way that cannot easily be accounted for in manipulationist terms. *Ceteris paribus*-generalisations being superior in this regard, a theory of explanation based on the latter is a better fit for social science.

Introduction

The problem of implicit background knowledge, like a bugbear, stalks many areas of the philosophy of science and epistemology. It troubles studies of confirmation,[1]

[1] Popper 1963, 1976, Hempel 1966, 1988, Mackie 1969, Watkins 1960, 1987.

knowledge,[2] the relationship between observation and scientific theory,[3] the formalisation of the latter,[4] causal modelling and inference,[5] expertise,[6] learning theory,[7] explanation,[8] and many more. Any attempt to drive it out of all of its hiding places simultaneously would be a Herculean task, so I will focus on the last item on this list, background knowledge in explanation—although some of what follows is likely to be relevant elsewhere as well. Implicit or tacit background knowledge plays an indispensable, albeit not always explicitly theorised or even fully acknowledged, role in a number of philosophical accounts of scientific explanation. C. G. Hempel, for example, faced with prima facie decisive counterexamples to his deductive-nomological (DN) model of explanation, famously brushed these off with the observation that explanations can satisfy the adequacy conditions of his model despite being 'practically useless' (Hempel 1965: 425ff). The issue of practical usefulness was not the business of philosophy of science in Hempel's eyes, but rather of 'pragmatics', so he did not concern himself with it much further.[9] Yet pragmatic analyses of explanation typically show that whether a given body of information is explanatory depends directly on the nature of the background knowledge of the audience, as well as its contextual interests (see e.g. van Fraassen 1980: ch. 5, Achinstein 1983).

Similarly, a primary difficulty for W. Salmon's Causal-Mechanical model of explanation (Salmon 1984) is that of explanatory relevance: it seems that without the right kind of background information we cannot insure that our causal explanations pick out only those causal processes/interactions that are explanatorily relevant to the explanandum (see Hitchcock 1995). The late Salmon thought that information about causal processes in conjunction with statistical relevance relationships would solve this problem (Salmon 1997: 476), but acknowledged that there is an epistemic worry: when considering the statistical relevance of a cause C and background conditions A to the effect B 'we must consider whether $A.C$ constitutes a homogenous reference class, or whether there are other factors D, E, F that are also relevant to the occurrence of B' (ibid.). Happily, 'such questions are open to empirical investigation'.[10] Thus we find that in a causal explanation of why C causes B we must appeal to a condition of homogeneity,

[2] Polanyi 1962, Peacocke 1998, Davies 2000.
[3] Duhem 1906, 1991, Quine 1951, Hanson 1958, Kuhn 1962, Strevens 2001.
[4] Kuipers 2001; Kamps 2005.
[5] Robins and Wasserman 1999, Cartwright 2007, van Gerven and Lucas 2007, Bonnefon, Da Silva Neves et al. 2008, Borboudakis and Tsamardinos 2012.
[6] Dreyfus and Dreyfus 1986, Collins and Evans 2007.
[7] Tenenbaum, Griffiths et al. 2006, van Gerven and Lucas 2007
[8] Hempel 1965, van Fraassen 1977, 1980, Achinstein 1983, Woodward 2003.
[9] Hempel also considered the difference between explanation and prediction to be purely pragmatic (Hempel 1965: 249).
[10] See Woodward 2014 for scepticism about whether Salmon's last proposal can work.

or a *ceteris paribus* condition, established empirically on the basis of our background knowledge about the relevance of *A* as well as a potentially indeterminate number of additional factors *D*, *E*, *F*... This background knowledge may or may not be causal itself, since statistical relevance factors may or may not be—a recurring theme in the present paper.

An important contemporary development of causal accounts of explanation is 'manipulationism.' Arguably the most worked out statement of the view is Woodward 2003, who proposes a notion of causal explanation according to which explanations are causal when they show how an outcome—the instantiation of a certain property or quantity expressible as a numerical magnitude—depends on the instantiation of other such properties or quantities.[11] The dependence in question is not merely logical or conceptual: causal explanations are meant to provide information about how we can change a given outcome by changing the property on which it depends (Woodward 2003: 10), thereby enabling the manipulation and control of nature. Since a causal relationship can obtain and an explanation based on the latter be given even in circumstances where actual manipulation of the cause is impossible, the relevant information is counterfactual, and the control provided by it often of a merely 'in principle' kind. Successful explanations, Woodward says, are associated with 'a hypothetical or counterfactual experiment that shows us that and how manipulation of the factors mentioned in the explanation [...] would be a way of manipulating or altering the phenomenon explained' (Woodward 2003: 11). Causal explanations, in other words, answer 'what-if-things-had-been-different questions' by highlighting 'the difference for the explanandum if the factors cited in the explanans had been different' (ibid.).

Manipulationism thus conceptualises causation as a special kind of counterfactual dependence, and cashes out causal explanation as a matter of the discovery and description of patterns of such dependence via 'invariant' generalisations.[12] One of Woodward's central theoretical concerns is to favourably contrast his account with regularity theories of causation and the DN-model of explanation: he highlights the relative deficiencies of the DN-model when compared to manipulationism, and emphasises problems with the explanatory use of laws of nature (conceived traditionally as exceptionless regularities) when compared to such use of invariant generalisations (Woodward 2003: 152-186 and 265-288). Contemporary refinements of the regularist account especially as applied to the social sciences in the form of the theory of hedged or *ceteris*

[11] Earlier manipulationist accounts are Collingwood 1940 and von Wright 1971; contemporary contributions include Hausman 1998, Hitchcock 2001, as well as Woodward and Hitchcock 2003 and Hitchcock and Woodward 2003, and Ylikoski and Kuorikoski 2010.
[12] Woodward in recent work uses 'interventionist' to refer to his theory of causation (Woodward 2009, 2014, 2015).

paribus generalisations come in for particular criticism. The approach is 'fundamentally flawed', Woodward claims, and 'an appreciation of these flaws […] bring[s] out the superiority of the [manipulationist] account' (Woodward 2003: 308).

The aim of this paper is to show that Woodward's programmatic claim that manipulationism 'fit[s] a wide range of scientific contexts, *especially in the social and behavioral sciences*' (Woodward 2003: 6, my emphasis) is mistaken, and by the same token to defend and advocate the explanatory use of *ceteris paribus* generalisations in those contexts. I will argue (Section 1) that Woodward's key notion of 'serious possibility' underemphasises the problem of the quantity and variety of background knowledge necessary for causal explanation; and that this reduces manipulationism's suitability as a theory of how we come to construct social scientific explanations on the basis of that knowledge. Section 2 shows why Woodward's objections to *ceteris paribus*-generalisations as a vehicle for scientific explanation fail, ironically despite the fact that the *ceteris paribus*-concept itself is at the very core of manipulationism. Section 3, finally, argues that the role of background knowledge in core social scientific methods—case- and comparative studies, randomised controlled trials, ethnographic studies, and structural equation modeling—suggests that a non-interventionist notion of explanation based on *ceteris paribus*-generalisations provides a better fit.

1. Of boulders and bugbears

Manipulationism as sketched above rests on two central concepts, 'intervention' and 'invariance'. An 'intervention' captures in non-anthropomorphic terms the 'conditions that would need to be met in an ideal experimental manipulation of the value of *X* performed for the purpose of determining whether *X* causes *Y*' (Woodward 2003: 14), where *X* and *Y* are variables that range over types and take on numerical values that denote the magnitude of a given quantity. In an intervention, the putative causal factor *X* is set to a specific value in such a way that *X* becomes causally and probabilistically independent of its previous cause(s); and the endogenous causal relationships that determined its value prior to the intervention are replaced wholly by an exogenous causal process, without however affecting *X*'s causal relationship with *Y*.[13] The change in *X* effected by the intervention must be such that if any change occurs in *Y* as a result of it, it is due to *Y*'s relationship to *X* and not to any other factor (Woodward 2003: 747). In other words, an intervention is, as Michael Strevens memorably put it, 'the result of the hand of God descending and directly tweaking the relevant factor', altering it and only it with no side effects (Strevens 2007: 234); or in Judea Pearl's metaphor, it is akin to

[13] Woodward credits among others Frisch 1995 [1938], Haavelmo 1944, Spirtes, Glymour et al. 1993, and Pearl 2000 with different variants of the concept.

'surgery' (Pearl 2000: 240, 347). 'Interventionism' stipulates that all genuine relations of cause and effect are such that the effect is susceptible to be changed through this sort of modification of its cause.

The concept of 'invariance' joined to that of an intervention yields manipulationism. A causal generalisation G that relates changes in the value of *X* to changes in the value of *Y* is *invariant* if G would continue to hold under an intervention on *X*, in the sense that G correctly describes how the value of *Y* would change as a result of this intervention (Woodward 2003: 239ff). 'Invariance under interventions' comes in degrees. At one end of the spectrum are fundamental physical laws that hold under all or almost all conditions and are invariant under a very wide range of interventions; at the other are (some) social science generalisations that are invariant under only a few interventions and hold only under a narrow and local set of conditions. (For example, regression equations correlating editorial newspaper endorsements and change in voting patterns, Woodward 2003: 207, 241). A generalisation's invariance is thus a function of the range of interventions under which it holds and the variation in the 'background conditions' over which it holds—by which Woodward intends changes in spatiotemporal location and other initial or boundary conditions that are not explicitly cited in G (Woodward 2003: 254). The wider the range of interventions and background changes under which G remains invariant, the better ('deeper') any explanation that deploys it.[14] Successful explanation according to manipulationism is a matter of displaying patterns of counterfactual dependence via generalisations that are as invariant as possible, the depth of the explanation they afford being a function of the range of 'what-if-things-had-been-different?' questions they answer (Woodward 2003: 193ff).

Woodward runs into a spot of trouble as he fleshes out the details of his account, however. In response to cases where manipulationism risks yielding counterintuitive outcomes, namely the boulder example due to Hall 2004 (see Woodward 2003: 79-81) and the similar cricket ball case due to McDermott 1995 (see Woodward 2003: 86-87), he acknowledges that 'conclusions about causal relationships are sensitive to one's choice of representation' (Woodward 2003: 80). In particular, our conclusion that *A* causes *B* will in many situations depend on which variables we choose for our representation of the causal system in which *A* and *B* are embedded, and this choice, in turn, can depend on what we consider to be a 'serious possibility' in that causal situation (Woodward 2003: 86). For instance, take the case of a mountain hiker surviving the fall of a boulder by ducking in time; had she not ducked, she would have died. Manipulationism risks yielding the result that the boulder's fall can be said to have caused her *survival*, and that its failure to fall would have caused her *death* (rather than the reverse), if we include an additional variable representing a second boulder in our representation

[14] For criticism of this conception of 'explanatory depth', see e.g. Strevens 2009; Imbert 2013.

of the situation, and arrange things in a just-so manner. Namely, a second boulder positioned in such a way that its fall at the right time would have killed the hiker were it not for the fall of the first—which the hiker happens to notice, causing her to duck and to thereby avoid *both* boulders. In such a situation, an intervention that stops the first boulder from falling would successfully manipulate whether or not the hiker survives the fall of the second, yielding the intuitively unappealing result above (see Woodward 2003: 79-80). Intuition might concede that the first boulder's fall was an indirect or 'contributing' cause of her survival, but it seems to balk at calling it its 'actual' cause, as Woodward is committed to do.[15]

Woodward dismisses the two-boulder scenario by pointing out that in order to arrive at the unwelcome conclusion we need to assume a causal system very different from the one initially intended. For there is no way that a falling boulder could cause or explain a hiker's survival in a system involving just that one boulder, and no way that it not falling could cause or explain her death (Woodward 2003: 81). Yet this observation obviously cannot be the end of the discussion. Manipulationism does not in principle disallow changes in causal structure when assessing the truth-value of causal counterfactuals. For instance, we legitimately use Coulomb's law to explain the strength and form of the electromagnetic field created by a live electrical wire, because the law holds in the system whose behaviour is being explained, a straight copper wire, as well as in countless other causal systems: wires made of other materials, other shapes, experimental setups, etc. (Woodward 2003: 241). In fact, change of causal

[15] See Woodward 2003: 79. Woodward defines a type variable X as a *direct cause* of type variable Y relative to a variable set V iff a possible intervention on X will change Y or its probability distribution while *all other variables* in V are held fixed at some value (Woodward 2003: 59, emphasis mine). Using this definition he says that X's taking value x is an *actual cause* of Y's taking value y, iff $X = x$ and $Y = y$ (i.e. the change actually takes place) and 'there is at least one route R from X to Y for which an intervention on X will change the value of Y or its probability distribution, given that *other direct causes* Z_i of Y that are not on this route have been fixed at their actual values' (where 'x' and 'y' range over particulars, and 'R' refers to a *directed path*, defined as an ordered set of variables $Z_1 \ldots Z_n$ such that X is a direct cause of Z_1, Z_n is a direct cause of Y, and for all $1 \leq i > n$ in between, Z_i is a direct cause of Z_{i+1}) (Woodward 2003: 77). He calls X a *contributing cause* of Y with respect to V iff 'there is a directed path from X to Y'(Woodward 2003: 57).
The rather complex definition of 'actual cause' successfully explains how transitivity of actual causation fails in the one-boulder-scenario. The fall of the boulder is an actual cause of the ducking, and the ducking an actual cause of survival, but the fall cannot be an actual cause of survival: given that the hiker ducked, an intervention on the boulder falling would make no difference to her survival; and there being (in that scenario) no other direct causes of survival on a route that bypasses her ducking, 'there is no well-defined operation of fixing such variables' in order to evaluate the influence of the boulder's fall via the route that does not bypass it (Woodward 2003: 80; see also Hitchcock 2001: 276ff). The two-boulder-scenario provides such a route, however, and needs to be ruled out. Some audiences, incidentally, declare no discomfort with calling the fall of the first boulder the *actual* cause of the hiker's survival in that scenario; they appear however to be in the minority.

structure is one of the dimensions along which causal generalisations must be invariant in order to be considered explanatory in the first place, so this cannot be why we must disregard the two-boulder-scenario when assessing counterfactual dependencies in the one-boulder-scenario.

Enter 'serious possibility'. Woodward suggests that in certain instances the relevant change in structure is just *too substantial* to accept the possibilities it generates as 'serious' (Woodward 2003: 86). This immediately generates the question: what degrees of change are consistent with 'serious possibility', and more generally, what are the factors that make a possibility 'serious'? Woodward offers a list of 'considerations' that he says are relevant to this question: (1) the probability of an event's occurrence 'given the actual obtaining background conditions [or] those that usually or commonly obtain in similar situations'; (2) moral requirements, expectations, and custom; (3) whether an outcome is controllable (or easily or cheaply controllable) by current technology; (4) an investigator's interests and purposes. Given the nature of the items in this list, he concedes that there is unlikely to be an algorithm for determining serious possibility (Woodward 2003: 88-89). That seems correct: (3) is historically contingent, and some elements of (2) and (4) would appear to be highly subjective. More importantly for my purposes, (2)-(4) do not address the question of the degree of counterfactual change in causal structure that is permissible.

(1) appears to do so at first blush, by speaking of background conditions that 'usually obtain in similar situations'. But we have no metric by which to objectively measure the relevant similarity, such that a causal scenario S_2 in which background conditions vary slightly, but foreground conditions are identical to S_1, would qualify as similar enough to be included in our causal evaluation of S_1; neither do we have a principled way to specify what ought to count as 'background' and 'foreground' in each case. Woodward certainly provides no formal criterion of similarity, but merely appeals to the intuition that the two-boulder-scenario is too different to be included as relevant in our causal calculus regarding the one-boulder-scenario. At the same time, he acknowledges that we need to limit the number of counterfactual situations we are allowed to consider when assessing causality and invariance: 'if we don't appeal to some notion like that of serious possibility, a manipulability theory will at the very least be led to causal judgments that are different from those that are ordinarily accepted' (Woodward 2003: 89).

Does 'serious possibility' amount to nothing more than an appeal to raw causal intuition, then? Woodward notes that while some aspects of the concept are clearly subjective, others are very much objective (Woodward 2003: 90). True enough, it is an objective fact, say, that mountain boulders qua macroscopic objects are not subject to quantum mechanical laws, therefore the likelihood of a boulder suddenly appearing at a new spatiotemporal location is near zero. But judgments of objective probability are of course sensitive to the scientific theory being used to make those judgments. Woodward gives short shrift not just to the question when two causal systems are sufficiently similar for us to be justified to transfer our judgments of probability from one to the other, he says nothing at all about the theory-relativity of such judgments. In the social domain, at which manipulationism is explicitly aimed, our knowledge of causal structures is often very limited, and so are our causal intuitions. In many cases we cannot base our estimates of the probability of an event or of the seriousness of a possibility on pre-theoretical causal knowledge or intuitions, because what we think seriously possible depends entirely on which relevant theory we accept. For example, our assessment of the probability of a social group's economic status being affected by the content of the group's religious beliefs depends on whether we subscribe to a Weberian vs. a classical Marxist sociological paradigm—just as our assessment of the likelihood of light deflection by massive objects depends on whether we operate in Newtonian mechanics or relativistic physics.

Sometimes, scientific judgments of possibility are indirectly supported by theories from other disciplines—political claims by economic theory or vice versa, psychological ones by biology, sociological ones by history, etc.—or depend on case-specific and non-causal background knowledge from those fields (as e.g. in Woodward's consideration (2)). In other cases what we think possible does not depend at all on what could occur according to our currently held scientific theories, because in thought experiments it may be indeterminate whether an event is possible under those theories (see El Skaf 2017: 20-21), but rather on newly hypothesised ones. Therefore, if it is to be scientifically useful at all the concept of 'serious possibility' can neither be collapsed into narrow nomological possibility, where this refers to the set of counterfactuals supported by currently known and accepted scientific laws, nor be identified too broadly with conceptual possibility. Many hypotheses, albeit strictly speaking conceptually possible, are not scientifically reasonable or acceptable. I suggest that when we speak of 'serious possibility' in the context of scientific explanation, we replace it with the term 'scientific possibility' in a *sui generis* sense specific to that context (for more see *infra*).

Woodward, as it happens, floats another defence of his use of the concept of 'serious possibility'. It is a *tu quoque:* most other theories of causation, he argues, share the flaw of centrally relying on a comparable idea, because 'the properties, variables, equations, and state spaces that the theorist uses to model or represent specific cases will either directly reflect judgments about which possibilities are to be taken seriously or will end up doing largely the same work as such judgments' (Woodward 2003: 89). For example, regularity theorists will have trouble explaining why regularities regarding omissions and absences do not amount to "real" regularities or to the instantiation of "real" properties without appealing to a concept very similar to that of serious possibility (ibid.); idem for David Lewis' counterfactual theory of causation. As we have seen, the problem of background knowledge (i.e. the often indeterminate set of domain-and non-domain specific, case- and non-case specific knowledge needed to assess the seriousness of a possibility, the similarity of a set of situations, or the reality of a property or regularity) is indeed the bugbear of not only most contemporary theories of causation, but also of theories of explanation, prediction, and scientific knowledge generally. Woodward is also quite correct that the ubiquity of this bugbear means that it cannot be a fatal objection to any particular theory that it appeals at some point to such knowledge.

Background knowledge can, however, play different roles in scientific inference and explanation; and in determining what is possible different types of background knowledge can be relied on to different degrees and in more or less well understood ways. In what follows I will argue[16] that the difficulty of accounting for the explanatory role of background knowledge is greater for manipulationism than for a suitably updated regularity theory that relies on *ceteris paribus* generalisations as vehicles of causal explanation; and that the latter is therefore to be preferred.

2. Manipulationism and regularities, *ceteris paribus*

Manipulationism, as we have seen, conceives of causal explanation via the concept of a ‘hypothetical or counterfactual experiment’, an intervention that shows how a change in the *explanans* would be a way of changing the *explanandum*. It requires interventions that are neither logically impossible nor ‘ill-defined for conceptual or metaphysical reasons’ (Woodward 2003: 128), though not necessarily physically achievable. For example, we can correctly judge that ‘the moon causes the tides’, because the relevant intervention is conceptually clear, despite the fact that it is probably impossible to physically change (the position of) the moon without in the course of such an intervention also impacting the tides (Woodward 2003: 131). Beyond changing X in a way that does not directly affect Y, Woodward further stipulates that interventions must be actual causes of X taking on some particular value x, in other words, that other possible direct causes of the value of X have been fixed at their actual values (Woodward 2003: 98; see also footnote nº 15). The concept of an ‘intervention’ therefore involves in its very definition the idea of other causes not on the path from the intervention variable to the variable intervened on, yet part of the variable set that describes the causal system, remaining unchanged (Woodward 2003: 98). This is identical to the notion of the appropriate set of *cetera* remaining *pares*.

Since ‘invariance’ is itself characterised as a property of generalisations under interventions thus defined, we can say that the ‘other things being equal’ idealisation per-

[16] In a manner Woodward anticipated, see Woodward 2003: 384n.

meates manipulationism from start to finish (cf. Henschen 2015). Why then would theories of *ceteris paribus*-generalisations that are explicitly based on the latter be 'fundamentally flawed' (Woodward 2003: 308), while invariant generalisations are not? Take a typical empirical generalisation of the special sciences, e.g. Okun's law, which states that a 2% increase in GDP growth is associated with a 1% decrease in the rate of unemployment. The "law" does not satisfy many of the criteria for lawfulness frequently considered standard in the philosophical literature. In particular, it is not clear whether the generalisation is universal, strict, as well as counterfactual-supporting, or necessary in a sense distinct from logical or conceptual/metaphysical necessity. Philosophers of the special sciences attempt to tackle two of these issues, the lack of universality and the proneness to exceptions of almost all generalisations in these domains, by arguing that candidate laws are generalisations hedged by a *ceteris paribus*-clause (henceforth, '*CP* laws'); a standard theory of *CP* laws is the 'completer account.'[17]

In its simplest version, a completer account stipulates that an otherwise law-like but non-strict generalisation of the form 'All *A*s are *B*s'—i.e. a generalisation of this form that is true and otherwise comparable to a genuine law of nature despite the fact that some *A*s are not *B*s—qualifies as a law if it is possible to specify a condition *C* such that when *C* is the case, all *A*s are *B*s (or such that $\forall x$ [(Ax & Cx) ⊃ Bx]). Condition *C*, so the idea, functions as a 'completer' of the non-strict generalisation by backing it up with a strict one. Woodward holds that there are three fundamental problems with completer accounts: (i) explanatory information must be information that is epistemically accessible, yet *C* is usually unknown or only partially known, because it refers to an indefinite or potentially infinite number of causal factors; hence *CP* laws cannot be used to successfully explain. (ii) there are few, if any, actual examples of truly exceptionless laws to be found in science, even in physics. (iii) there are many non-strict generalisations with completers that we would not naturally regard as laws, or as explanatory (Woodward 2003: 209).

None of (i)-(iii) are fatal, as I have previously argued (Kowalenko 2011, 2014 this journal). Take (i), epistemic inaccessibility: statements of the form '*CP*, All *A*s are *B*s' can be interpreted to mean the same as 'All *A*s in *C* are *B*', where this is a two part-claim: the *explicit* claim that a finite and determinate number of conditions $\{C_1 \ldots C_n\}$ need to obtain for the generalisation to be true, and the *implicit* claim that an opaque and potentially infinite set of further conditions $\{C_{n+1}, C_{n+2}, \ldots\}$ are nomically irrelevant to the generalisation, given the evidence in its support (Kowalenko 2014: 147; cf.

[17] Completer accounts explain how laws of nature could have exceptions without, however, necessarily resolving all other outstanding issues with 'laws' (see e.g. Fodor 1991: 22). They have been proposed in different forms by Hempel 1965 (arguably also Hempel 1988), Hausman 1988, 1992, Fodor 1991, Pietroski and Rey 1995, Carrier 1998, Cartwright 2002, Steinberg, Layne et al. 2012, Strevens 2012, 2014; Pemberton and Cartwright 2014; Kowalenko 2011, 2014, among others.

Strevens 2012, Schurz 2014). In other words, *CP* clauses when spelt out explicitly state the causal factors that are known, given the evidence, to interfere with the truth of the generalisation, while implying that other causal factors are not thus known (ibid.). Hence, in order to explain via a *CP* generalisation it is contra Woodward sufficient to non-trivially characterise the explicit (transparent, determinate, and epistemically accessible) part of the content of its *CP* clause, and unnecessary to perform the impossible feat of describing its implicit (opaque, indeterminate, and epistemically inaccessible) content (cf. Woodward 2003: 308). Depending on the scientific field, the explicit part of a given *CP* clause can be supplied by a variety of scientific methods commonly used to derive a generalisation from evidence, such as: the methodologies used in single and comparative case studies, qualitative ethnographic approaches, as well as quantitative methods such as Randomised Controlled Trials, structural equation modeling, multivariate regression and statistical curve-fitting (see Kincaid 1995: 63-84; Kowalenko 2007, 2014: 150; Steinberg, Layne et al. 2012; and *infra*).

(ii) no exception-less laws: the view that there are very few or no strict law-like empirical regularities has been forcefully argued by Cartwright 1983, 1999 for nature in general, and it has been a staple of (the philosophy of) the social sciences since at least J.S. Mill.[18] Recently the consensus has begun to fray somewhat: a number of authors suggest that there could be exception-less natural laws after all (Earman and Roberts 1999, Earman, Roberts et al. 2002, Gildenhuys 2010), even in the biological and social sciences (Elgin 2006; Kincaid 1995). This is of course also Hempel 1965, 1988's position. Even Cartwright 2002 and Pemberton and Cartwright 2014 now think that laws of nature can be seen as strict, on condition of an ontologically correct interpretation of their relata in terms of capacities or dispositions (cf. Mumford 2004). On my view, *ceteris paribus*-laws properly construed similarly are strict laws, although I do not require any particular ontological interpretation. Overall, I take the jury on this question to be very much out.

(iii) The claim that there are fatal counterexamples in the form of obviously spurious or vacuous *CP* generalisations with completers has been popular ever since Hausman suggested '*ceteris paribus*, all dogs have six legs' as a problem case (Hausman 1988: 309); and Fodor queried the difference between 'it'll fly *ceteris paribus*' and 'it'll fly unless it doesn't' (Fodor 1991: 22). The difficulty is evident: under entirely conceivable albeit somewhat fantastical conditions all dogs would indeed have six legs, so the generalisation does have a "completer" of sorts. Any occurrence of a non-six-legged dog could be chalked up to the completer's failure to materialise. And on pain of saying precisely what we actually mean by '*ceteris paribus*,' any airplane's or other potentially airborne object's failure to become such could be explained (away) by appealing to the

[18] See e.g. Mill 1843, 1882: Book 6, Chapter III; also Popper 1935, 1957, Nagel 1952, and Feigl and Broadbeck 1953: chapter 7.

fact that another more-or-less fantastical yet possible completer failed to be instantiated in the circumstances. The conclusion seems to impose itself that reality cannot in principle contradict *CP* statements, rendering the latter either vacuous or spurious. Yet, the charge is false when *CP* laws are understood correctly as referring to a finite set of factors determined by a particular scientific model or theory supported by empirical data. For none of the spurious examples in the literature are associated with a scientifically reasonable or plausible model able to show how the relevant generalisation is based on evidence interpreted by such a model.

Let us assume, for instance, that Fodor's 'it'll fly *ceteris paribus*' refers to an airplane prototype in the course of construction. Suppose, further, that we inferred it from a set of *CP* generalisations in aerodynamics, aeronautical engineering, and other relevant truths about planes of the same kind as the prototype. We know, in particular, that for the prototype to become airborne it needs to be built in such a way that lift, drag, weight, and thrust forces are at specific ratios. A structural equation model derived from experience that describes these relations could be used to provide an important part of the content of the completer in this instance (Drela 1999: 5). Suppose, finally, that the plane fails to fly despite these forces being at the correct ratios and other relevant conditions also being satisfied, and that someone attributes this to 'diabolic influence'—to use Hempel 1988's example of a spurious interferer. We would be entitled to say that the *CP* clause hedging our statement *implicitly* rules out such influence, because diabolic forces were not part of the model of the forces governing airplanes that we used to derive the prediction. They are not part of any scientifically reasonable model, in fact, and we would not consider the model's failure to make room for Satan's potential meddling a falsification of it and the generalisations we derive from it. We take only those factors to require explicit mention in the *ceteris paribus* condition that are causally relevant, given our best current scientific knowledge (Kowalenko 2014: 147).

Mutatis mutandis for most counterexamples in the literature that putatively establish the vacuity of *CP* generalisations (for other examples, see e.g. Kowalenko 2011: 448-449, 2014: 142). Yet, we evidently cannot say that a hypothetical causal factor such as 'diabolic force' is not a scientifically possible interferer in aerodynamics—in Woodward's terms, a serious possibility—without making use of theoretical background knowledge. Deeply steeped as we are in a physicalist worldview and a host of related background assumptions, we rule out such forces almost *a priori*: their inclusion would cause too much theoretical disruption in our physical theory as well as in our philosophical outlook.[19] Almost all causal inference and explanation ultimately relies on an

[19] Kuhn taught us, of course, that if the existence and constant operation of a hitherto unnoticed force were robustly confirmed in numerous subsequent observations across numerous different experiments, a theoretical and philosophical tipping point could be reached (all else being equal) that would force us to conclude that our initial model—and paradigm—must have been misspecified.

amorphous body of tacit background knowledge and judgment that is famously resistant to theoretical description and modelling (see e.g. Polanyi 1962: 91ff; Dreyfus and Dreyfus 1986; Kamps 2005; and *infra*). Nonetheless, I think it can be shown both that *CP* generalisations represent viable vehicles of explanation, causal and otherwise, especially in the social sciences; and that they are in crucial respects superior to invariant generalisations—in particular in the way they allow us to deal with the bugbear of background knowledge.[20]

3. *CP* generalisations in Social Scientific Method

CP generalisations are ubiquitous all-purpose vehicles for the representation of explanatory background knowledge in a way that invariant generalisations are not. This makes them vastly preferable for encoding or representing our causal knowledge, especially in the social sciences. To show this, I survey some of the most widely used methodologies of the social sciences—case studies, comparative small-N studies, randomised controlled trials (RCTs), ethnographic studies, as well as structural equation modeling—and argue in each case that: it would be less natural and more cumbersome to try to represent the output of the relevant method in the form of invariant generalisations, rather than *CP* ones; that in some instances, it would be strictly impossible; and finally, that this is because in each methodology causal and non-causal background knowledge mesh in a way that cannot easily be accounted for in manipulationist terms.

3.1 Case studies

I begin with case studies, the in-depth and multi-faceted investigation of a single social phenomenon, episode, or context. Case study methodology often invokes multiple qualitative or 'mixed' research methods such as participant observation, long interviews, archival and socio-historical research, action research, historical process tracing, content analysis, ethnographic decision tree modelling, etc., as well as purely quantitive

[20] Others have provided strong arguments for the view that manipulationism is at odds with the methods of the social sciences. Boumans 2003, for example, points out that manipulationism is ill equipped to account for inference from passive observation, a much more common practice in the social sciences than active manipulation; Cartwright 2007 says it gives a sufficient condition for causal inference only for ideal experimental settings far removed from the conditions under which it takes place in real life; Reiss 2007 notes that it assumes the existence in social contexts of robust causal relations that do not break down under intervention; and Russo 2011 claims it does not provide a suitable causal test in contexts in which interventions are not physically or ethically possible (see Russo 2014 for a summary). However, none of these authors contrast manipulationism with a theory of explanation based on *CP* laws specifically in terms of its treatment of social scientific background knowledge.

statistical analyses (see e.g. Yin 2009; Woodside 2010). Generalisation of the findings is an often explicit, though not undisputed, goal (see Platt 1988; Hammersley, Foster et al. 2000; Evers and Wu 2006; Steinberg 2015). Countering the argument that any attempt to infer universal propositions is inappropriate when just one instance has been observed, Yin points out that generalisation is not necessarily from a sample to a population, as it can also be from a set of data generated in one singular episode or context to a theory. His term for this type of inference is 'analytic generalisation,' others prefer 'theoretical generalisation' or 'theoretical analysis' (Yin 2009: 14ff; Mitchell 1983: 189). Like experiments, case studies can be used to test or expand theoretical propositions, and they are particularly useful when the studied phenomenon is not yet sufficiently clearly theorised or is likely to be the result of causal relationships that are too complex for quantitative survey, causal modelling, or experimental strategies (Yin 2009: 14ff).

Insofar as the theoretical propositions established by generalisation from a single case study are in fact *general*, however, they will in almost all actual examples have a wide range of exceptions. Generalisations from case studies are thus subject to a *ceteris paribus* condition: to preserve their truth value they have to be hedged by exception clauses, i.e. have the form of a *CP* generalisation or a set of them. *Invariant* generalisations, on the other hand, will often be very hard to derive from the data. For illustration, let's look at an early and now classic single case study in sociology, Jahoda, Lazarsfeld et al. 1933.[21] The authors studied the effects of mass unemployment on a 1930s mining community in Marienthal, Austria. Using activity and attitudinal criteria, they proposed a four-part typology of families coping with the consequences of unemployment of the main bread winner: the 'unbroken', 'resigned', 'in despair', and 'apathetic' family (Jahoda, Lazarsfeld et al. 1971: 45-56). Jahoda et al. found that several years after the closure of the main mine in town, out of a total of 478 families, 23% were unbroken, 70% resigned, and 7% either in despair or apathetic (Jahoda, Lazarsfeld et al. 1971: 56). They interpret the results as lending support to the conclusion that mass unemployment leads to resignation and apathy; rather than, say, to political radicalisation, an important concern in the 1930s.

This inference is clearly *ceteris paribus*, as the typology itself shows that for a large group of people (23%) unemployment did not such thing, suggesting that additional factors must have an impact on the causal link between the two variables.[22] Although the authors take an explicitly methodological holist approach to their study ('The

[21] Page references are to the first English edition Jahoda, Lazarsfeld et al. 1971.
[22] The criteria used to establish the typology were themselves exception-prone. For example, the authors considered absence of 'careful maintenance of the family budget' to be a characteristic of one category of families, while acknowledging that *all* unemployed families were found to display varying degrees of irrational spending (Jahoda, Lazarsfeld et al. 1971: 55).

object of this investigation was the unemployed community, not the unemployed individual. Character traits were given little attention', Jahoda, Lazarsfeld et al. 1971: 2), they acknowledge that the sample they studied may have been skewed by individual character traits: 'probably some of the most active and energetic families had already escaped […] by emigrating before we came to Marienthal' (Jahoda, Lazarsfeld et al. 1971: 57). Thus, despite their declared methodological orientation Jahoda et al. do in fact display interest in the question how 'an individual's life history affects his powers of resistance during unemployment' (Jahoda, Lazarsfeld et al. 1971: 89). They explicitly correlate resignation during the crisis with either (a) traces of earlier resignation, or (b) a previous life of particularly high ambitions/standard of living, while explaining cases of a member of the (b) group *not* experiencing resignation by their 'adaptability' (Jahoda, Lazarsfeld et al. 1971: 94-97). Meanwhile, families with "normal" previous lives that display no clearly distinguishing features are found among all four types, and the authors speculate similarly that here, too, 'character traits' were the decisive factor (Jahoda, Lazarsfeld et al. 1971: 98).

So we see a complex interplay between 'unemployment', 'life history,' and 'energy,' 'adaptability,' and an unspecified further range of 'character traits.' How could one operationalise these observations in manipulationist terms? Recall from the boulder example that causal explanation calls for answers to what-if-things-had-been-different questions, and that the latter require us to make judgments about the possibility of events based on comparisons between relevantly similar causal systems. Although we may have hunches about the types of factor that are relevant to the link between unemployment and resignation, not all of these are necessarily causal or understood in their causal efficacy, and we certainly do not know anything like the complete set V of causal variables relevant to a given Marienthal family in its social context. We therefore have trouble estimating degrees of similarity between relevant causal systems—say between families with a previous life history of high ambition and those with low ambition—and hence cannot say what the counterfactual dependencies are.

Generally speaking, it seems that true invariant generalisations cannot reliably be formulated without quite extensive prior causal background knowledge. Yet the background knowledge relevant to generalisations in the social sciences is often either not causal at all, or not easily representable in terms of other factors whose causal role is well understood (proxy variables, etc.). The problem is not quite the same for *CP* generalisations: a completer $\{C_1 \ldots C_n\}$ may contain causal or non-causal information or both simultaneously, for example indiscriminately list character traits, life histories, as well as economic data, etc. This is crucial, as it allows us to more comfortably encode our background knowledge regarding the reasons why Jahoda et al.'s conclusion has a high number of exceptions. The ostensibly causal claim that 'mass unemployment leads to resignation and apathy' is hence best seen as a causal *CP* claim. It is not hard to find other seminal causal claims in the social sciences based on case studies that for the same reasons are most naturally expressed in *CP* generalisations—Malinowski 2005 (1922), Liebow 2003 (1967), Allison 1971, Geertz 1971, come to mind. Case study methodology generally seems inhospitable to manipulationism.

3.2 Comparative small-N studies

Much the same can be said about comparative small-N studies. Here the researcher compares a handful of cases with the aim of setting up matching and contrasting cases in a way that would allow identification of primary variables of interest, and (perhaps) achieve the equivalent of controlling for exogenous factors. The comparative method is invaluable for concept formation and the formulation of general explanatory ideas, but it does not seem particularly suitable for a causal account of explanation (cf. Brady and Collier 2010: 10), especially a manipulationist one. For matching and contrasting cases is not likely to succeed at controlling exogenous variables in the way required by manipulationism, by holding fixed at their actual values all variables in the representation not currently on the *X*-to-*Y* path. Usually, the N is not nearly large enough to do this even statistically. Despite this limitation comparative studies like case studies are a *sine qua non* of much social science, perhaps because the operant notion of explanation in the fields employing this method is precisely not a manipulationist one, not even implicitly or ideally. For if we try to derive universal propositions from comparative method—through what King, Keohane et al. 1994: 56ff call 'descriptive inference'—these will often have to be hedged generalisations, not invariant ones.

Kitschelt 1986, for example, in a well-known study looks at the anti-nuclear power movements in France, Sweden, the United States, and Germany. He argues that notable differences in mobilisation strategy employed by these movements in pursuit of their goals as well as their differential impact can be explained by the same set of variables, namely a particular configuration of resources, institutional arrangement, and historical precedent, that together determine a given democratic nation's 'political opportunity structure'. Kitschelt does not hold that political opportunity structure fully determines the dynamics of social movements, only that it 'explains a good deal […] *if other determinants are held constant*' (Kitschelt 1986: 57-58, my emphasis). Although the author does not use the expression, his conclusion is most naturally summarised in terms of a *CP* generalisation—in fact, it must be. For even if we interpret the term 'determination' in the quote causally, Kitschelt 1986's claim cannot be construed in terms of an interventionist causal claim: the notion of an intervention is conceptually unclear in his context.

As in the moon and tides case, we cannot even in principle conceive of an intervention on opportunity structure that would not simultaneously influence a panoply of other country specific cultural and political factors that would, in turn, impinge on social movement dynamics. Generally, it is not only physically impossible to manipulate many important social causal factors in order to test their effects, such as, say, race or gender influence on wages, even a *fictitious* intervention on these types of variable might not be conceivable in a given social context without the intervention affecting the way they are related to their effects in that context. But we do want to say that race and gender are causes of wage disparity; idem for our standard ways to manipulate macroeconomic properties such as inflation or household income (money supply, interest rates, wages), and so on. As a result, constructing social scientific explanations in terms of invariant generalisations based on knowledge generated from comparative small-N studies often often does not work. No such problems, by contrast, for *CP* generalisations: given Kitschelt's explicit proviso above, it seems that they are in fact the *only* way in which a general conclusion of any sort could be drawn from a study such as his. I venture the claim that the same will be true of most comparative small-N studies.

3.3 Randomised Controlled Trials

Randomised Controlled Trials (RCTs) face problems of generalisation similar to those of case studies. RCTs are designed specifically to analyse the causal effect of a given causal intervention in contexts where we lack information about what further factors might influence the effect: a randomly selected group of individuals is exposed to a treatment—a cause—whose effects are under study, while a similarly selected control group is not. (Although they may receive a placebo, making it impossible for them to determine in which group they are). As a result of randomisation, which distributes even unknown confounders more or less identically in the treatment group and in the control, a well-designed RCT will provide an unbiased estimate of the true causal effect of a given treatment in a given context, i.e. it will be internally valid. RCTs are 'convenient ways to introduce experimenter-controlled variance [i.e. manipulation]—if you want to see what happens, then kick it and see, twist the lion's tail' (Deaton and Cartwright 2016: 58). Yet, a core difficulty, for example for the purposes of formulating clinical practice or predicting the efficiency of government policy, is that this is true only for the population from which subjects were selected: if our goal is to export the causal conclusions of the RCT from the experimental population to a larger and different target population, then we cannot be assured of the same conclusions in the target without presupposing the trial's external validity, or generalisability (see e.g. Deaton and Cartwright 2016).

Glennerster 2012 recognises this problem in the case of government policy evaluation (Glennerster 2012: R11). She argues that to improve our grasp of the context sensitivity of a given policy intervention, we need: a thorough theoretical understanding of the mechanism by which treatment and expected outcome are related; knowledge of the manner and process through which treatment was implemented; as well as evidence from non-randomised studies and qualitative data (Glennerster 2013: 395ff). Take, for example, school-based programs of deworming. We know that deworming leads to increased school attendance by pupils across many different countries, because the biomedical mechanism by which deworming pills reduce parasites and increase health is well understood, as is the causal link between overall health and school attendance. Medical background knowledge suggests that these links are likely to be stable across all human populations. By contrast, programs designed to improve the quality of government social services through information campaigns that aim to increase citizen advocacy can be expected to be highly sensitive to differences in institutional context, in particular to the degree of government responsiveness to local political pressure (Glennerster 2013: 398).

Glennerster therefore suggests that we crack open the "black box" of external validity by carefully distinguishing between general principles about human behaviour provided by theory, on the one hand, and the differences in local conditions and trial implementation that are responsible for frequent breakdowns of RCT generalisability, on the other (Glennerster 2013: 398, 2017). Notice that the black box cannot be cracked open if we limited ourselves to manipulationist methodology, however. We may have a satisfactory representation of the causal system in which an intervention on parasitic worms via administration of a drug leads to health and higher school attendance. But no such construct is available for the route by which an increase in the available information about the quality of government services causes increased advocacy, and via the latter their improvement; we would need a causal model of the entire political economy of a country. Glennerster is explicit that *qualitative* background knowledge is often indispensable to estimate the external validity of RCTs; in fact, she leaves it open whether the very theory grounding our understanding of how treatment and outcome are related must rely on *causal* mechanism at all: all we need is a 'theory of change' that specifies a '*logical chain* of how program inputs achieve changes in outcomes' (Glennerster 2013: 181ff; emphasis mine).

I suggest that the task of integrating what we can learn from a given RCT with qualitative knowledge from non-randomised studies and case studies is easier, if we conceive of the relevant conclusions as couched in *CP* generalisations. After all, the theoretical rationale for RCTs is a *ceteris paribus*-thought, too: an *ideal* RCT is, by definition, a RCT such that all factors that can produce (or eliminate) a causal dependence between the cause *C* and effect *E* are exactly the same in the treatment group and control, except for *C* (Cartwright 2009: 64). In other words, a trial in which, *everything else being equal*, *C* causes *E*. Therefore, the evidential import of RCTs is naturally described in terms of appropriately phrased *CP* generalisations. It is important, again, for my purposes, that causal and non-causal *CP* regularities mesh better (especially in terms of logical inference relations) than causal models mesh with non-causal models or with non-causal descriptive knowledge; thus we can expect *CP* generalisations to make the task of conjoining these different types of knowledge easier. To take a simple example: given that '*CP* All *A*s are *B*s' and that '*CP* *B*s cause *C*s', we can infer that '*CP* *A*s cause *C*s,' despite the fact that the latter generalisation is a causal one and the former is not, and notwithstanding that the content of the *CP* clause of each proposition is likely different.[23]

3.4 Ethnography

The case for the explanatory use of *CP* generalisations is easily made for purely qualitative studies. Take, for example, ethnography, a qualitative research method employing long-term 'participant observation', 'thick description', and other observational methods for data collection. Ethnography avoids non-qualitative approaches because ethnographers, anthropologists, and sociologists see the background knowledge required to make the (causal) assumptions underlying causal modeling as leading to an illicit imposition of an a priori structure on social inquiry (Hammersley 1992: 11-12, cited in Mitchell 2007: 56). Descriptive observational methods are favoured because they issue (it is hoped) in an inductive or abductive process by which theoretical models and classifications are derived from direct observations that describe the processes at work—"description is explanation!" being the slogan.

Ethnographers and cultural anthropologists have come to see social processes as primarily generated by a web of meanings that we have spun ourselves; its description ('thick description') requires the ethnographer to enter the imaginative universe within which the cultures's signs and behaviour have meaning, the ultimate aim being the 'enlargement of human discourse' via the translation of meaning from one culture to another (Geertz 1973: 13-14). Geertz interprets for example the popular practice of betting on cockfights among Balinese men as an *emblem* (a symbol) of social relationships in local society—in particular, kinship, village, and status relationships (Geertz 1971). Arguably (though not uncontroversially so), ethnography's concern like that of every science is ultimately with *universal* knowledge, and hence with generalisation (Firestone 1993, Williams 2000). It is immediately evident that the type of knowledge required for thick description is of a different kind than the knowledge required to, say, distinguish between the one-boulder and the two-boulder-scenario: it is as much semantic as it is causal. While one may realistically hope that some ethnographic insights at least can be formulated in hedged generalisations and that they can in *that* form be

[23] The question how the content of the *CP* clause in 'CP_3 *A*s cause *C*s' could be derived from the *CP* conditions in 'CP_1 All *A*s are *B*s' and 'CP_2 *B*s cause *C*s' is, however, complicated. It seems that it cannot be simply the conjunction 'CP_1 & CP_2', for it is as unlikely that relations between *CP* clauses are summative as that relations between models are (cf. Kowalenko 2014: 149).

integrated with causal social scientific knowledge, prospects of manipulationist explanation in ethnography appear to be zero.

3.5 Structural Equation Modelling

Last but not least, what about structural equation modeling? Glymour 2004 says that Woodward's manipulationism is in essence nothing but a decomposition and rearrangement of the theory of causal Bayes nets, with the concept of intervention as place of departure (Glymour 2004: 780, 784; see also Freedman 2007). Woodward himself appears to think that the best way to account for the widespread use of structural equation modeling in the testing of causal claims in economics is via an interventionist concept of causation and generalisations invariant under such interventions (Woodward 1999, Woodward 2003: 315ff). Structural equation models were probably what he primarily had in mind when he touted manipulationism as 'fitting' many social science contexts. Pearl 2000 gives as a simple example of such a model the 'functional causal model' constituted by the two econometric equations '$Q = b_1P + d_1I + u_1$' and '$P = b_2Q + d_2W + u_2$', 'where Q is the quantity of household demand for [a product], P is the unit price of [the product], I is household income, W is the wage rate for producing [the product], and u_1 and u_2 represent error terms - unmodeled factors that affect quantity and price, respectively' (Pearl 2000: 27, citing Goldberger 1992). Each equation is intended to represent an autonomous mechanism showing how household income and price are an immediate cause of demand, on the one hand, and wages and demand are an immediate cause of price, on the other (the model is cyclic); note that u_1 and u_2 represent the errors, or "disturbances," introduced into the causal relationship between Q, P and I, and P, Q and W, respectively, by factors omitted in the equations (Pearl 2000: 27).

I have previously argued that multivariate regression models can provide the content of the *CP* clauses that explicitly or implicitly hedge a number of generalisations in the special sciences (2014, this journal). The idea is that the content of a *CP* clause can be interpreted as given by the factors controlled for in the regression model used to interpret the evidence in support of the generalisation the clause hedges. Structural equation modeling being a second-generation multivariate technique, we can expect the same to be true here. The epistemic rationale of structural equation models like that of many other scientific methodologies is to identify a causal relation and isolate it (in the model) from the impact of potential interfering factors—to hold *cetera pares*. Hence, I argue that structural equation models can naturally be connected to sets of *CP* statements, e.g. in the case above to the claims '*CP P* and *I* are immediate causes of *Q*' and '*CP Q* and *W* are immediate causes of *P*', where the unmodeled factors represented by

u_1 and u_2 are "parked" in the respective *CP* clause.[24] Structural equation models are often visually represented by Direct Acyclic Graphs (DAGs) (acyclic models are usually preferred for capturing counterfactual dependence relations). A problem for DAGs and the associated models analogous to the problem of external validity in RCTs is their *transportability* (see e.g. Petersen 2011). I propose that this problem is better tackled if we think of structural equation models as associated with or implying *CP* generalisations.

For, a DAG will attempt to represent *all* variables measured and unmeasured that are thought to be significant to the relevant causal relationship along with all common causes, just as structural equations attempt to capture all modeled factors as well as the unmodeled, but known to be potentially relevant, factors in the disturbance term. This corresponds (when the attempt to spell them out is made) to the *explicit* content $\{X_{a1} \ldots X_{an}\}$, $\{X_{b1} \ldots X_{bn}\}$, etc. of the *CP* clauses CP_a, CP_b, etc. hedging the corresponding set of *CP* generalisations. As in the case of the disturbance term, varying degrees of scientific investigative effort will yield *ceteris paribus* conditions of varying complexity, but the crucial point here is that neither '*CP*' nor '*u*' should be interpreted as explicitly referring to infinities. The explicit meaning at least of a *CP* clause must be redeemable, and so must be the disturbance or error term. We can then view failures of transportability as due to discrepancies in the *implicit* content of the model, i.e. conditions $\{X_{an+1}, X_{an+2}, \ldots\}$, $\{X_{bn+1}, X_{bn+2}, \ldots\}$, etc. of the corresponding *CP* clauses, which are never spelt out because it is impossible to do so. These are the potentially infinitely many factors that by virtue of being excluded from the model are implicitly declared to be irrelevant for the truth or falsity (given the data) of the causal claim or generalisation (cf. 2014: 147, this journal). Interpreting structural equation models and DAGs as (associated with) a network of *CP* statements in this way presents advantages, because the latter can capture not just *causal* or probabilistic, but also *qualitative* conditional independence relations (see Boutilier, Brafman et al. 2004: 140). This gives us more power than manipulationism to theorise failures of transportability.

Conclusion

Woodward's *tu quoque* argument, being the fallacy that it is, should be rejected. It may be true that all theories of explanation are equally plagued by the role of possibility in explanation, but not all theories of explanation are equally germane to all scientific practices. I have been concerned to show that at least as far as the social sciences are

[24] See applications of this idea e.g. in decision and learning theory: Boutilier, Bacchus et al. 2001; Boutilier, Brafman et al. 2004; McGeachie and Doyle 2004.

concerned, a traditional regularity-approach to explanation in terms of *CP* generalisations is in many cases superior. Manipulationism, I have argued, presupposes causal counterfactual knowledge of a special kind—'serious possibility', as evinced in the intuitions of the boulder example—which is not useful in those areas where this knowledge is precisely what we lack, or where our conception of 'serious possibility' in Woodward's sense meshes in subtle ways with and presupposes non-causal knowledge. Not all conceivable 'counterfactual experiments' can be associated with successful causal explanation (as Woodward 2003: 11 acknowledges), because in social science contexts in particular we have no idea what would result from such experiments, despite clearly having *some* relevant knowledge—as e.g. in the case of Jahoda, Lazarsfeld et al. 1971's families.

This suggests that the appropriate background knowledge necessary for scientific explanation is not counterfactual knowledge of what is 'seriously possible' in manipulationist terms, but of what I called 'scientific possibility': a modal concept referring to a set of possibilities or possibles worlds larger than nomic possibility, yet smaller than conceptual/metaphysical possibility. Scientifically possible is that which it is reasonable/justifiable to hypothesise given a specific factual situation, evidence, theory and background knowledge. Using a number of examples from the social sciences, I argued that the manner in which these elements are welded together in each instance to draw particular causal conclusions or formulate generalisations cannot be adequately explained in manipulationism. For our concept of 'scientific possibility' is generated from sources that include both qualitative and quantitative, causal and non-causal information, and as such transcends the traditional assumptions of causal modeling. *CP* generalisations, I have argued, are the best vehicle for encoding such knowledge.

A methodology based on *CP* laws as the primary modus of explanation presents a more promising avenue for controlling the bugbear of background knowledge, for the completer account promises the much-needed flexibility to model how causal and non-causal knowledge can mesh and interact. In fact, the assumption that completers have an explicit, determinate, and in principle fully redeemable, as well as an implicit, indeterminate, and irredeemable component, is the best way I can see to theorise 'scientific possibility', background knowledge, and the way we use the latter to grasp the former. What we are implying when we chose not to refer to or describe a factor in our representation of the actually obtaining regularities in a given situation, is that it is irrelevant to that situation. This judgement is closely connected to our judgment of what is scientifically reasonable, i.e scientific possibility. After all, the completer of a *CP* clause in a scientific context is always the amalgamated output of a number of scientific methodologies, as we have seen, many of which are not based exclusively on causal or nomothetic knowledge, but require background understanding of an amorphous type. That we cannot simply identify scientific possibility with nomic possibility is in fact the very *raison d'être* of *CP* generalisations, but further elaboration of this idea will have to wait for another day.

Bibliography

Achinstein, P. (1983). *The Nature of Explanation*. Oxford, Oxford University Press.

Allison, G. T. (1971). Essence of Decision: Explaining the Cuban Missile Crisis. Boston, Little Brown.

Bonnefon, J.-F., R. Da Silva Neves, D. Dubois and H. Prade (2008). 'Predicting causality ascriptions from background knowledge: model and experimental validation.' *International Journal of Approximate Reasoning* **48**: 752-765.

Borboudakis, G. and I. Tsamardinos (2012). Incorporating Causal Prior Knowledge as Path-Constraints in Bayesian Networks and Maximal Ancestral Graphs. *Proceedings of the 29th International Confer- ence on Machine Learning*. Edinburgh, Scotland**:** 59.

Boumans, M. (2003). 'How to Design Galilean Fall Experiments in Economics.' *Philosophy of Science* **70**: 308-329.

Boutilier, C., F. Bacchus and R. I. Brafman (2001). *Ucp-Networks: A Directed Graphical Representation of Conditional Utilities*. Proceedings of Seventeenth Conference on Uncertainty in Artificial Intelligence, Seattle.

Boutilier, C., R. I. Brafman, C. Donshlak, H. H. Hoos and D. Pool (2004). 'CP-nets: A Tool for Representing and Reasoning with Conditional Ceteris Paribus Preference Statements.' *Journal of Artificial Intelligence Research* **21**: 135-191.

Brady, H. E. and D. Collier (2010). Rethinking Social Inquiry. Diverse Tools, Shared Standards. Lanham, Maryland, Rowman & Littlefield.

Carrier, M. (1998). 'In defense of psychological laws.' *International Studies in the Philosophy of Science* **12**(3): 217-232.

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

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

Cartwright, N. (2002). 'In Favour of Laws That Are Not Ceteris Paribus After All.' *Erkenntnis* **57**: 425-439.

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

Cartwright, N. (2009). 'What are randomised controlled trials good for?' *Philosophical Studies* **147**(1): 59-70.

Collingwood, R. G. (1940). *An Essay on Metaphysics*. London, Oxford University Press.

Collins, H. and R. Evans (2007). *Rethinking Expertise*. Chicago, University of Chicago Press.

Davies, M. (2000). Externalism and Armchair Knowledge. *New Essays on the A Priori*. P. A. Boghossian and C. Peacocke. Oxford, Clarendon**:** 384-427.

Deaton, A. and N. Cartwright (2016). Understanding and Misunderstanding Randomized Controlled Trials. *NBER Working Paper Series*. Cambridge, MA, National Bureau of Economic Research.

Drela, M. (1999). Integrated Simulation Model for Preliminary Aerodynamic, Structural, and Control-Law Design of Aircraft. *40th AIAA SDM Conference April 12-15, 1999*. St. Louis, MO, American Institute of Aeronautics and Astronautics.

Dreyfus, H. and S. E. Dreyfus (1986). Mind over Machine: the Power of Human Intution and Experties in the Era of the Computer. New York, Free Press.
Duhem, P. (1906, 1991). *The Aim and Structure of Physical Theory*. Princeton, Princeton University Press.
Earman, J. and J. T. Roberts (1999). 'Ceteris Paribus, There is No Problem of Provisos.' *Synthese* **118**: 439-478.
Earman, J., J. T. Roberts and S. Smith (2002). 'Ceteris Paribus Lost.' *Erkenntnis* **57**: 281-301.
El Skaf, R. (2017). 'What Notion of Possibility Should We Use in Assessming Scientific Thought Experiments?' *Lato Sensu: Revue de la Société de Philosophie des Sciences* **4**(1): 18-30.
Elgin, M. (2006). 'There may be strict empirical laws in biology, after all.' *Biology and Philosophy* **21**: 119-134.
Evers, C. W. and E. H. Wu (2006). 'On Generalising from Single Case Studies: Epistemological Reflections.' *Journal of Philosophy of Education* **40**(4): 511-526.
Feigl, H. and M. Broadbeck, Eds. (1953). *Readings in the Philosophy of Science*. New York, NY, Appleton-Century-Crofts.
Firestone, W. A. (1993). 'Alternative Argumets for Generalizing from Data as Applied to Qualitative Research.' *Educational Researcher* **22**(16-23).
Fodor (1991). 'You Can Fool Some People All of the Time, Ceteris Paribus; Heged Laws and Psychological Explanations.' *Mind* **100**(397): 19-34.
Frisch, R. (1995 [1938]). Autonomy of Economic Relations. *The Foundations of Econometric Analysis*. D. F. Hendry and M. S. Morgan. Cambridge, Cambridge University Press**:** 407-419.
Geertz, C. (1971). Deep Play: Notes on the Balinese Cockfight. *The Interpretation of Cultures*. C. Geertz. New York, Basic Books.
Geertz, C. (1973). Thick Description: Toward an Interpretative Theory of Culture. *The Interpretation of Cultures*. C. Geertz. New York, Basic Books.
Gildenhuys, P. (2010). 'Causal Equations without Ceteris Paribus Clauses.' *Philosophy of Science* **77**(4): 608-632.
Glennerster, R. (2012). 'The Power of Evidence: Improving the Effectiveness of Government by Investing in More Rigorous Evaluation.' *National Institute Economic Review* **January**(219): R4-R14.
Glennerster, R. (2013). Running Randomized Evaluations. A Practical Guide. Princeton, Princeton University Press.
Glennerster, R. (2017). From Evidence to Scale: Drawing Policy Lessons from Randomised Evaluations. *African Micro Economic Research Unit Seminar Series*. University of the Witwatersrand, unpublished.
Glymour, C. (2004). 'Critical Notice. James Woodward 'Making Things Happen: A Theory of Causal Explanation'.' *British Journal of Philosophy of Science* **55**: 779-790.
Goldberger, A. S. (1992). 'Models of substance.' *Brazilian Journal of Probability and Statistics* **6**(1-56).
Haavelmo, T. (1944). 'The Probability Approach in Econometrics.' *Econometrica* **12 (Supplement)**: 1-118.
Hall, N. (2004). Two Concepts of Causation. *Causation and Counterfactuals*. J. Collins, N. Hall and L. A. Paul. Cambridge, MA, MIT Press**:** 225-277.

Hammersley, M. (1992). What's Wrong With Ethnography? London, Routledge.
Hammersley, M., P. Foster and R. Gomm (2000). Case study and generalisation. *Case Study Method: Key Issues, Key Texts*. R. Gomm, M. Hammersley and P. Foster. London, SAGE Publications**:** 98-115.
Hanson, N. R. (1958). *Patterns of Discovery*. Cambridge, Cambridge University Press.
Hausman, D. M. (1988). *Ceteris Paribus Clauses and Causality in Economics*. PSA 1988: Proceedings of the 1988 Biennial Meeting of the Philosophy of Science Association, Philosophy of Science Association.
Hausman, D. M. (1992). The Inexact and Separate Science of Economics. Cambridge, Cambridge University Press.
Hausman, D. M. (1998). *Causal Asymmetries*. Cambridge, Cambridge University Press.
Hempel, C. G. (1965). Aspects of Scientific Explanation. New York, Free Press.
Hempel, C. G. (1966). *The Philosophy of Natural Science*. Upper Saddle River, New Jersey, Prentice Hall.
Hempel, C. G. (1988). 'Provisoes: A Problem concerning the Inferential Function of Scientific Theories.' *Erkenntnis* **28**: 147-164.
Henschen, T. (2015). 'Ceteris paribus conditions and the interventionist account of causality.' *Synthese* **192**(10): 3297-3311.
Hitchcock, C. (1995). 'Salmon on Explanatory Relevance.' *Philosophy of Science* **62**(2): 304-320.
Hitchcock, C. (2001). 'The Intransitivity of Causation Revealed in Equations and Graphs.' *Journal of Philosophy* **98**(6): 273-299.
Hitchcock, C. and J. Woodward (2003). 'Explanatory Generalizations, Part II: Plumbing Explanatory Depth.' *Noûs* **37**(2): 181-199.
Imbert, C. (2013). 'Relevance, Not Invariance, Explanatoriness, Not Manipulability: Discussion of Woodward's Views on Explanatory Relevance.' *Philosophy of Science* **2013**(5): 625-636.
Jahoda, M., P. F. Lazarsfeld and H. Zeisel (1933). Die Arbeitslosen von Marienthal. Ein soziographischer Versuch über die Wirkungen langandauernder Arbeitslosigkeit. Leipzig, Hirzel.
Jahoda, M., P. F. Lazarsfeld and H. Zeisel (1971). Marienthal. The Sociography of an Unemployed Community. Chicago and New York, Aldine Atherton.
Kamps, J. (2005). The Ubiquity of Background Knowledge. Cognitive Structures in Scientific Inquiry: Essays in Debate with Theo Kuipers. R. Festa, A. Aliseda and J. Peijnenburg. New York, Rodopi. **2:** 317-337.
Kincaid (1995). Philosophical Foundations of the Social Sciences. Cambridge, Cambridge University Press.
King, G., R. O. Keohane and S. Verba (1994). Designing Social Inquiry. Scientific Inference in Qualitative Research. Princeton, New Jersey, Princeton University Press.
Kitschelt, H. P. (1986). 'Political Opportunity Structures and Political Protest: Anti-Nuclear Movements in Four Democracies.' *British Journal of Political Science* **16**(1): 57-85.
Kowalenko, R. (2007). A Curve Fitting Approach to Ceteris Paribus-Laws [Manuscript]

Kowalenko, R. (2011). 'The epistemology of hedged laws.' *Studies in History and Philosophy of Science* **42**(445-452).
Kowalenko, R. (2014). 'Ceteris Paribus Laws: A Naturalistic Account.' *International Studies in the Philosophy of Science* **28**(2): 133-155.
Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago, The University of Chicago Press.
Kuipers, T. A. F. (2001). Structures in Science: Heuristic Patterns Based on Cognitive Structures. Dordrecht, Kluwer Academic Publishers.
Liebow, E. (2003 (1967)). Tally's Corner. A Study of Negro Streetcorner Men. Lanham, Rowman & Littlefield.
Mackie, J. L. (1969). 'The Relevance Criterion of Confirmation.' *British Journal of Philosophy of Science* **20**: 27-40.
Malinowski, B. (2005 (1922)). Argonauts of the Western Pacific. An Account of Native Enterprise and Adventure in the Archipelagoes of Melanesian New Quinea. London, Routledge.
McDermott, M. (1995). 'Redundant Causation.' *British Journal for the Philosophy of Science* **46**(4): 523-544.
McGeachie, M. and J. Doyle (2004). 'Utility Functions for Ceteris Paribus Preferences.' *Computational Intelligence* **20**(2): 158-217.
Mill, J. S. (1843, 1882). A System of Logic, Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation [8th ed.]. Toronto, University of Toronto Press Routledge & Kegan Paul.
Mitchell, J. C. (1983). 'Case and situation analysis.' *The Sociological Review* **31**(2): 187-2011.
Mitchell, J. P. (2007). Ethnography. *The SAGE Handbook of Social Science Methodology*. W. Outhwaite and S. P. Turner. London, SAGE Publications.
Mumford, S. (2004). *Laws in Nature*. London, Routledge.
Nagel, E. (1952). 'The Logic of Historical Analysis.' *The Scientific Monthly* **74**.
Peacocke, C. (1998). 'Implicit Conceptions, Understanding and Rationality.' *Philosophical Issues* **9**: 43-88.
Pearl, J. (2000). Causality: Models, Reasoning, and Inference. Cambridge, Cambridge University Press.
Pemberton, J. and N. Cartwright (2014). 'Ceteris Paribus Laws Need Machines to Generate Them.' *Erkenntnis* **79**: 1745-1758.
Petersen, M. L. (2011). 'Compount Treatments, Transportability, and the Structural Causal Model.' *Epidemiology* **22**(3): 378-381.
Pietroski, P. and G. Rey (1995). 'When Other Things Aren't Equal-Saving Ceteris Paribus Laws From Vacuity.' *British Journal of Philosophy of Science* **46**: 81-110.
Platt, J. (1988). What can case studies do? *Studies in Qualitative Methodology*. R. Burgess. London, JAI Press. **1:** 1-24.
Polanyi, M. (1962). Personal Knowledge. Towards a Post-Critical Philosophy. London, Routledge & Kegan Paul.
Popper, K. (1935). *Logik der Forschung*. Vienna, Springer Verlag.
Popper, K. (1957). *The Poverty of Historicism*. Boston, MA, The Beacon Press.
Popper, K. (1963). Conjectures and Refutations: the Growth of Scientific Knowledge. London, Routledge and Kegan Paul.

Popper, K. (1976). Background Knowledge and Scientific Growth. *Can Theories Be Refuted?* S. Harding. Dordrecht, Holland, D. Reidel Publishing Company**:** 113-116.

Quine, W. v. O. (1951). 'Two Dogmas of Empiricism.' *The Philosophical Review* **60**(1): 20-43.

Reiss, J. (2007). Error in Economics: Towards a More Evidence-based Methodology. London, Routledge.

Robins, J. M. and L. Wasserman (1999). On the Impossibility of Inferring Causation from Association without Background Knowledge. *Computation, Causation, and Discovery*. P. Glymour and G. Cooper. Cambridge, MA, The MIT Press**:** 305-321.

Russo, F. (2011). 'Correlational Data, Causal Hypotheses, and Validity.' *Journal for General Philosophy of Science* **42**: 85-107.

Russo, F. (2014). 'What Invariance is and How to Test for It.' *International Studies in the Philosophy of Science* **28**(2): 157-183.

Salmon, W. C. (1984). Scientific Explanation and the Causal Structure of the World. Princeton, New Jersey, Princeton University Press.

Salmon, W. C. (1997). 'Causality and Explanation: A Reply to Two Critiques.' *Philosophy of Science* **64**: 461-477.

Schurz, G. (2014). 'Ceteris Paribus and Ceteris Rectis Laws: Content and Causal Role.' *Erkenntnis* **79**: 1801-1817.

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

Steinberg, J. R., C. M. Layne and A. M. Steinberg (2012). 'Ceteris Paribus Causal Generalizations and Scientific Inquiry in Empirical Psychology.' *Journal of Theoretical and Philosophical Psychology* **32**(3): 180-190.

Steinberg, P. F. (2015). 'Can We Generalize from Case Studies?' *Global Environmental Politics* **15**(3): 152-175.

Strevens, M. (2001). 'The Bayesian Treatment of Auxiliary Hypotheses.' *British Journal for the Philosophy of Science* **52**: 515-537.

Strevens, M. (2007). 'Review of Woodward, Making Things Happen.' *Philosophy and Phenomenological Research* **24**(1): 233-249.

Strevens, M. (2009). *Depth: An Account of Scientific Explanation*. Boston, Massachusetts, Harvard University Press.

Strevens, M. (2012). 'Ceteris Paribus Hedges: Causal Voodoo That Works.' *Journal of Philosophy* **109**(11): 652-675.

Strevens, M. (2014). 'High-Level Exceptions Explained.' *Erkenntnis* **79**: 1819-1832.

Tenenbaum, J. B., T. L. Griffiths and C. Kemp (2006). 'Theory-based Bayesian models of inductive learning and reasoning.' *Trends in Cognitive Sciences* **10**(7): 309-318.

van Fraassen, B. C. (1977). 'The Pragmatics of Explanation.' *American Philosophical Quarterly* **14**(2): 143-150.

van Fraassen, B. C. (1980). *The Scientific Image*. London, Clarendon.

van Gerven, M. and P. J. F. Lucas (2007). The Role of Background Knowledge in Bayesian Classification. *Advances in Probabilistic Graphical Models*. P. J. F. Lucas, J. A. Gamez and A. Salmerón. Berlin, Springer**:** 377-396.

von Wright, G. H. (1971). *Explanation and Understanding*. London, Routledge & Kegan Paul.

Watkins, J. W. N. (1960). 'Confirmation without Background Knowledge.' *British Journal for the Philosophy of Science* **10**(40): 318-320.
Watkins, J. W. N. (1987). 'Lawson on the Raven Paradox and Background Knowledge.' *British Journal for the History of Science* **38**(4): 567-571.
Williams, M. (2000). 'Interpretivism and Generalisation.' *Sociology* **34**(2): 209-224.
Woodside, A. G. (2010). Case Study Research: Theory, Methods, Practice. Bingley, UK, Emerald Group Publishing.
Woodward, J. (1999). 'Causal Interpretation in Systems of Equations.' *Synthese* **121**: 199-247.
Woodward, J. (2003). Making Things Happen. A Theory of Causal Explanation. Oxford, Oxford University Press.
Woodward, J. (2009). Agency and Interventionist Theories. *The Oxford Handbook of Causation*. H. Beebee, C. Hitchcock and P. Menzies. Oxford, Oxford University Press**:** 234-262.
Woodward, J. (2014). 'A Functional Account of Causation; or, A Defense of the Legitimacy of Causal Thinking by Reference to the Only Standard That Matters—Usefulness (as Opposed to Metaphysics or Agreement with Intuitive Judgment).' *Philosophy of Science* **81**: 691-713.
Woodward, J. (2015). 'Interventionism and Causal Exclusion.' *Philosophy and Phenomenological Research* **91**(2): 303-347.
Woodward, J. and C. Hitchcock (2003). 'Explanatory Generalizations, Part I: A Counterfactual Account.' *Noûs* **37**: 1-24.
Yin, R. K. (2009). *Case study research design and methods*. Thousand Oaks, CA, Sage Publications.
Ylikoski, P. and J. Kuorikoski (2010). 'Dissecting Explanatory Power.' *Philosophical Studies* **148**: 201-219.