What is mechanistic evidence, and why do we need it for evidence-based policy?

Caterina Marchionni[1] and Samuli Reijula[2]

[1]Practical Philosophy, University of Helsinki
[2]Philosophy, University of Tampere

forthcoming 2018 in
Studies in History and Philosophy of Science Part A

Abstract

It has recently been argued that successful evidence-based policy should rely on two kinds of evidence: statistical and mechanistic. The former is held to be evidence that a policy brings about the desired outcome, and the latter concerns how it does so. Although agreeing with the spirit of this proposal, we argue that the underlying conception of mechanistic evidence as evidence that is different in kind from correlational, difference-making or statistical evidence, does not correctly capture the role that information about mechanisms should play in evidence-based policy. We offer an alternative account of mechanistic evidence as information concerning the causal pathway connecting the policy intervention to its outcome. Not only can this be analyzed as evidence of difference-making, it is also to be found at any level and is obtainable by a broad range of methods, both experimental and observational. Using behavioral policy as an illustration, we draw the implications of this revised understanding of mechanistic evidence for debates concerning policy extrapolation, evidence hierarchies, and evidence integration.

Keywords— Evidence-based policy, Behavioral policy, Mechanistic evidence, Extrapolation, Evidence integration

1. Introduction[1]

The evidence-based policy (EBP) movement urges policymakers to select policies on the basis of the best available evidence that they work. EBP utilizes evidence-ranking schemes to evaluate the quality of evidence in support of a given policy, which typically prioritize meta-analyses and randomized controlled trials (henceforth RCTs) over other evidence-generating methods. Many early applications of EBP were extensions of evidence-based medicine in health policy, but the

[1]Both authors contributed equally to the paper.

approach has since been applied to issues concerning social and public policy, and most recently in behavioral policy.

Philosophers and practitioners alike have recently questioned some of the central tenets of EBP. Philosophers of science criticize the status of RCTs as the gold standard of evidence (e.g., Worrall 2002), and argue that other kinds of evidence alongside the theory-free data provided by RCTs are needed to justify policy interventions (Cartwright and Hardie 2012; Clarke et al. 2013; Clarke et al. 2014). Similarly, social scientists suggest that attempting to establish internal validity by means of randomized controlled trials is only the first step in designing effective policy (Sampson, Winship, and Knight 2013).

What else, exactly, is needed in addition to evidence obtained from RCTs? According to one influential proposal, evidence of mechanisms – which evidence-based policy typically places at the bottom of the evidence hierarchies – is as important as evidence from RCTs and other statistical studies. There are several reasons for this. The most important one, on which we focus here, stems from the fact that evidence-based policy nearly always involves predictions about the effectiveness of an intervention in populations other than those in which it has been tested. Such extrapolative inferences, it is argued, cannot be based exclusively on the statistical evidence produced by methods higher up in the hierarchies (e.g., Clarke et al. 2014; Grüne-Yanoff 2016; Waldner 2012).

We agree with the spirit of this proposal insofar as it highlights the need for a broader evidence base when extrapolating policies. Yet, the notion of mechanistic evidence is problematic. Not only is it understood differently in different bodies of literature, it also tends to be packed together with questionable ideas about causation and mechanisms. We show that once such a notion of mechanistic evidence is stripped of misleading connotations, what is left is a generic notion of causal detail that, alone, is of little use to evidence-based policy.

In its stead, we put forward a componential difference-making account of mechanistic evidence for policy extrapolation (the CDM account for short). CDM evidence concerns the causal pathway that connects a policy to an outcome, which is what makes it, intuitively, an account of mechanistic evidence. Such evidence can be analyzed in terms of difference-making relations (between components falling on the causal pathway), and unlike the many existing views, our account does not presuppose that there are differences in kind between statistical and mechanistic evidence. The CDM account builds on two general observations about policy and evidence. First, policy design requires predicting the effectiveness of an intervention in populations other than those in which the intervention has been tested. Second, those predictions must rely on counterfactual information about the conditions under which the relation between the policy and the outcome remains invariant.

Three implications relevant to EBP follow from the CDM account. First, evidence about mechanisms is epistemically valuable for policy insofar as it contributes to refining invariance judgments. The CDM account therefore provides a clear rationale for determining when and why collecting evidence about mechanisms is useful for policy. Collecting evidence for policy is often slow and costly, hence a criterion is needed for sorting out which pieces of evidence to look for. Invariance provides such a criterion. Second, what in our account makes a piece of information mechanistic evidence is independent of the method by which it is produced. Therefore, recognizing

the importance of mechanistic evidence for EBP does not have any direct implications concerning the place of case studies, laboratory experiments and so on in evidence hierarchies. This insight defuses a common argument against evidence hierarchies claiming that they downplay the importance of mechanistic evidence. There might be other reasons for being critical of such hierarchies, but the importance of mechanistic evidence, when the latter is properly understood, is not one of them. Third, the CDM account implies that the challenge of evidential integration in EBP is twofold. Not only does it involve ranking and combining types of evidence about the *same* causal relationship (evidence amalgamation), it also requires bringing together evidence about *different* aspects of the causal scenario. The latter task is one in which the representational tools of policy graphs hold much promise. Hence, our account highlights a further policy-relevant epistemic challenge that evidence hierarchies in general and meta-analyses in particular are not meant to deal with.

Section 2 introduces the debate on mechanistic evidence in evidence-based policy, and Section 3 shows how a similar set of issues surfaces in current discussions on behavioral policy. We start to clarify the notion of mechanistic evidence in Section 4, in terms of what it is not, and in Section 5 we put forward our positive proposal which connects mechanistic evidence to the epistemologically more fundamental concept of invariance. Section 6 shows how our approach, when combined with policy graphs, helps to address the challenge of evidence integration in evidence-based policy. We set out the implications of our approach for behavioral policy in Section 7, and conclude the paper in Section 8.

2. Evidence hierarchies and mechanistic evidence

Evidence-ranking schemes promoted in evidence-based medicine (EBM), and also endorsed in EBP, place meta-analyses and RCTs at the top followed by the category of quasi-experimental and non-experimental statistical studies. Basic science, (referring to laboratory research, expert consensus and case studies) is placed at the bottom. Several arguments have been advanced against such evidence hierarchies (for a summary see, for example, Stegenga 2014). A recent influential line of critique is that the hierarchies unduly regard evidence from laboratory experiments and case studies as inferior to that obtained from statistical studies (whether experimental or observational). Russo and Williamson (2007), for example, referring to the health sciences, claim that evidence hierarchies fail to take into account the fact that many statistical studies (including RCTs) provide only "evidence of probabilities." They refer to this type of evidence interchangeably as "difference-making evidence" and "statistical evidence", arguing that evidence of mechanisms or "mechanistic evidence", which is primarily obtained through studies such as laboratory experiments and case studies, is also needed.

Clarke et al. (2014) make a similar point, and also discuss evidence-based policy more generally.[2]

[2]The connection between the different branches of evidence-based medicine, public health, social policy and behavioral policy runs through the general theme of the *epistemology of evidence*. For example, similar evidence hierarchies are used in all the aforementioned fields. Philosophers use the same vocabulary and distinctions such as statistical and mechanistic evidence in the context of several EBM and EBP fields.

The argument, in a nutshell, is that whereas statistical studies provide evidence that the policy variable, X, makes a difference to the policy outcome, Y, mechanistic evidence gives information about either the existence or the nature of a causal mechanism connecting the two; in other words, about the entities and activities mediating the X–Y relationship. Both types of evidence, it is argued, are required to establish causal claims, to design and interpret statistical trials, and to extrapolate experimental findings. It is the latter claim, namely that policy extrapolation requires both types of evidence, which is the main focus of our paper.

The Russo-Williamson distinction between statistical and mechanistic evidence was originally introduced in the context of their epistemic theory of causality. The necessity of both types of evidence for establishing causal claims is expressed in the Russo-Williamson thesis: in order to establish a causal claim, evidence that X and Y are probabilistically related and evidence that there is a mechanism between them are both necessary (for refinements and applications, see e.g. Dragulinescu 2017; Joffe 2013; Moneta and Russo 2014; for critical views, e.g. Campaner 2011; Claveau 2012).[3] Clarke et al. (2014) claim that the same distinction and the need for mechanistic evidence in EBM and EBP also arises from considerations that are independent from the Russo-Williamson thesis. Let us review the argument in some detail.

First, Clarke et al. (2014) claim that mechanistic evidence is often crucial for the design and interpretation of data obtained from an RCT. For example, they point to the need for accurate diagnostic criteria when selecting the population on which to perform a medical trial. Such diagnostic criteria tend to be derived from knowledge of the relevant (e.g., physiological and psychological) mechanisms however. In general, theoretical knowledge of the phenomena in question provides experimenters with ideas about which dependencies to study, and theoretical assumptions might be needed to make sense of experimental findings. We agree with this argument and will not say much more about it in the rest of the paper.

Second, Clarke's et al.'s other claim, which is more directly relevant to the debate on EBP, concerns the need for mechanistic evidence in extrapolating policy effectiveness from one population to another, and from a population to the individual level. In the philosophical literature, a well-known example is the failure to export the successful *Tamil Nadu Integrated Nutrition Program* (TINP) to Bangladesh. TINP aimed at diminishing malnutrition in children by (among other things) educating pregnant women about nutritional practices that promote their children's health (Cartwright 2012; Clarke et al. 2014). One of the reasons for the success of TINP and the failure of a similar program in Bangladesh is that whereas in India the mother is in charge of deciding what the child eats, in Bangladesh it is often the mother-in-law who plays such a role. This explains why the educational intervention, which only targeted mothers, was not successful in Bangladesh. The lesson to draw from this example is this: given that the different social structures in the two countries explain the failure to export the successful policy program, knowing the mechanism whereby TINP

[3]Note that the attributes "statistical" and "difference-making" are sometimes used interchangeably to refer to evidence of correlations. As will become clear below, however, unlike Russo and Williamson (2007) and Grüne-Yanoff (2016), we prefer to keep statistical evidence separate from difference-making evidence. Hence, from now on we use the term statistical evidence to refer to evidence of correlations.

worked would have helped in exporting it to Bangladesh. According to the critics, recognizing the importance of mechanistic evidence implies either that the evidence hierarchies should be dispensed with altogether or that they should be revised so as to include separate grading schemes for statistical and for mechanistic evidence.

3. Mechanistic evidence in behavioral policy

Behavioral policy, also known as behaviorally informed policy, is gaining increasing popularity among policymakers. It is premised on the idea that interventions in public policy should be based on a psychologically realistic picture of human behavior and its causes (Lourenço et al. 2016; Oliver 2013; Thaler and Sunstein 2008). Promoters of behavioral policy, and of EBP more generally, are also committed to the practice of testing behavioral policies through randomized experiments before large-scale implementation (Halpern 2016).

Behavioral policy often amounts to translational research on human decision-making; in other words to transporting results from basic decision-making research (e.g., in behavioral economics, cognitive psychology, and social psychology) to policy contexts. At the same time, however, theory and behavioral policy are only loosely connected. This is particularly obvious in practice in that psychological theory tends to function merely as a source of ideas for new interventions. Several theoretical hypotheses are considered in parallel, with no existing evidence pointing to which one best captures the policy situation in question. This is also reflected in the policy language: "what works," behavioral "insights," and policy "levers" all convey the message that policy design need not be based on systematic theorizing. In light of how theory is understood in the psychological sciences, such an attitude is hardly surprising. As Camerer (1999) points out, many psychologists understand theory merely as a verbal construct that organizes an experimental regularity. Moreover, psychology is theoretically disunified and major disagreements persist within several fields relevant to behavioral policy, even about how psychological theorizing should be done (see Berg and Gigerenzer 2010). Hence, it is no surprise that the relatively theory-free attitude of evidence-based behavioral policy has been seen as a welcome development in the context of policy-making.

The tension between the need for theoretical knowledge of mechanisms on the one hand, and the anti-theoretical practices in the field on the other, makes behavioral policy an interesting case for studying the role of mechanistic evidence in policy. In philosophy, Till Grüne-Yanoff (2016) has recently taken up this challenge arguing that behavioral policies are not "really evidence-based" unless they are based on both evidence of mechanisms and statistical evidence. In the behavioral sciences, Gerd Gigerenzer and colleagues have challenged mainstream behavioral policy on the grounds that it often lacks process models, models that detail the psychological mechanisms underlying decision making (see for example Berg and Gigerenzer 2010; Gigerenzer, Hertwig, and Pachur 2011; Katsikopoulos 2014).

By way of an illustration, let us consider Grüne-Yanoff's (2016) example of behavioral interventions that exploit the default effect. The default effect refers to the experimentally robust phenomenon

that people tend to choose the option given them as the default. For example, participation rates in organ-donation programs in different countries are significantly affected by whether people must actively choose to participate, or if they must actively opt out in case they do not wish to do so (Johnson and Goldstein 2003). Changing the default has also been utilized as a way of increasing charitable giving (Behavioral Insights Team 2013), for example, and helping people to save more for retirement (Thaler and Benartzi 2004).

It is unclear, however, what explains the default effect. At least three alternative mechanisms have been put forward (Dinner et al. 2011; Grüne-Yanoff 2016). First, the *cognitive effort avoidance hypothesis* explains the choice of the default option as the result of people's tendency to minimize effort in cognitively costly decisions. This hypothesis rests on the assumption that people are uncertain about their preferences, and therefore making an active choice is costlier than just going with the default. Second, the *recommendation effect hypothesis* holds that people interpret the default as the recommended option, and unless they have reasons not to trust the policymaker, they make use of the information embodied in the situation by choosing the default option. Finally, *loss aversion* has been suggested as a possible mechanism behind the stickiness of defaults. According to this hypothesis, people weigh losses more strongly than gains. In the case of retirement options, the idea is that if the default option (the reference point) entails higher levels of contribution to retirement than the alternatives, then the loss in future financial security that choosing these alternatives would entail will be weighted more heavily than the gain in current consumption obtained by departing from the default.

According to Grüne-Yanoff (2016), this example highlights the necessity of evidence of mechanisms alongside evidence about the average causal effect of the default-setting intervention in a specific population.[4] In fact, depending on which of the three hypothesized mechanisms is at work, different predictions regarding the effectiveness of a given intervention follow.[5] For example, if in one population, let us call it the *source*, changing the default option worked through the mechanism of cognitive effort avoidance, then what matters for the exportability of the intervention to the different context in which we wish to intervene, the target, is whether the deliberation costs are sufficiently significant in the *target*. Similarly, depending on which mechanism is supposed to be at play, different factors will affect whether the intervention will produce the desired effect when it is scaled up, for example, and for how long the effect of the policy will persist. Let us assume, again, that in the source population the policy of changing the default worked through cognitive effort avoidance. Let us further suppose that, over time, the targeted individuals who stuck with the new default come to realize that, in fact, they prefer current consumption to future financial security. Cognitive effort avoidance works when individuals are uncertain about their preferences, but as

[4]Grüne-Yanoff (2016) also argues that mechanistic evidence is necessary for the assessment of the welfare implications of alternative policy interventions. In this paper we only consider the claim that mechanistic evidence is necessary for the purposes of extrapolation.

[5]We follow Cartwright (2012) in using *evidence of efficacy* to refer to evidence that the intervention produces the desired outcome in a given study population ("there"), and *evidence of effectiveness* to refer to evidence supporting the claim that the intervention will produce the desired outcome in the target population or context ("here").

soon as they experience the effects of the new saving regime on their lifestyle, they learn about (or develop) their own preferences. In the longer term, the aggregate effect of the policy diminishes.

We agree with Grüne-Yanoff's analysis of the role that information about mechanistic pathways should play in behavioral policy. However, our paths diverge in how we understand the notion of mechanistic evidence.

4. What mechanistic evidence is not

Mechanistic evidence is claimed to play an important, even necessary, role at various stages in the ex-ante estimation of a policy's effectiveness, but what exactly is mechanistic evidence? It is claimed that, in addition to being evidence about the mechanism that connects the treatment to the outcome, it is also evidence generated by methods that are typically placed at the bottom of evidence hierarchies in EBM and EBP (e.g., case studies, expert consensus, and laboratory experiments). It is also often assumed to be different in kind from statistical evidence. These three aspects of evidence do not necessarily need to go together, however.

In this section we first highlight the difference between mechanistic *knowledge*, *reasoning*, and *evidence*. We then argue that mechanistic evidence should not be defined by contrasting it to statistical evidence. We outline our positive proposal, the CDM account, in Section 5.

4.1. Mechanistic knowledge, mechanistic reasoning and mechanistic evidence

Mechanistic knowledge, mechanistic reasoning, and mechanistic evidence are sometimes used interchangeably in the literature. However, fostering an understanding of how mechanistic evidence contributes to successful policy requires distinguishing it from the two related, but different, notions.

Mechanistic knowledge refers to the existing body of knowledge concerning the mechanisms of phenomena in a given domain (Campaner 2011). In medicine, for example, it encompasses what is known about, say, the biological mechanisms of metabolism and reproduction. Theories about mechanisms fall on a continuum from well-confirmed ones to mere hypotheses. As mechanistic hypotheses, the alternative explanations of the default effect reviewed in the previous section are examples of mechanistic knowledge in behavioral policy.

In line with Howick, Glasziou, and Aronson (2010, p. 434), we characterize *mechanistic reasoning* as involving inferences from what is known about mechanisms (mechanistic knowledge) to claims that an intervention produces a relevant outcome, where the reasoning involves a chain of inferences that links the intervention to the outcome. In medicine, for example, such reasoning typically involves inferring clinical efficacy from basic science (La Caze 2011, p. 88). In the behavioral-policy example it refers to reasoning from what is known about the cognitive mechanisms that underlie the preference for defaults to conclusions concerning the efficacy of a default-setting intervention. Although well-confirmed theories have more evidential weight than mere hypotheses, the general problem with mechanistic reasoning is the frequent uncertainty about whether the

mechanisms will be relevant in the policy context at hand, and how they interact with other mechanisms at work (Howick 2011, pp. 934-937). Considerations such as these constitute part of the rationale behind the low ranking of mechanistic reasoning in current evidence hierarchies. Independently of its evidential status, however, knowledge of mechanisms may well play a heuristic role: it is a source of hypotheses about possibly relevant processes.

Mechanistic *evidence*, in turn, refers to evidence about which causal pathway is active in a particular context of application. In the default example it should be evidence about the pathway through which the default intervention works or has worked in a particular case. If, as we argue, mechanistic evidence can be generated through a variety of methods, and evidence hierarchies typically rank evidence based on the method by which it is produced, then mechanistic evidence is not a category that should appear in such hierarchies.

Distinguishing mechanistic evidence from mechanistic knowledge is especially relevant in social-science contexts, in which instead of relying on well-established, well-confirmed theories about the policy-relevant phenomenon, the tendency is to have multiple competing theories suggesting different mechanisms behind a given phenomenon. This is why a reasonable way of proceeding is to try out interventions based on robust laboratory phenomena, and then collect evidence in the source and in the target in support of the different theoretical hypotheses about mechanisms.

4.2. Statistical versus mechanistic evidence

Having distinguished between mechanistic knowledge, reasoning and evidence, let us focus on the latter as it is understood in the philosophical and social-scientific literature. As mentioned, it is sometimes defined in contrast to statistical evidence, which is understood as evidence produced in observational or experimental studies (primarily RCTs) by means of co-variational methods. On the face of it, the contrast seems clear. Several interpretations of the difference between statistical and mechanistic evidence turn out to be problematic, however. Let us consider the following interpretations, all found in the literature, although not necessarily endorsed by the same author:

i The two types of evidence are produced by distinct methods or sets of methods,

ii statistical evidence is quantitative, mechanistic evidence is qualitative,

iii statistical evidence is about a population, mechanistic evidence is about one particular unit or individual (i.e. within-case evidence), and

iv statistical evidence is about the macro level, whereas mechanistic evidence concerns the micro level.

Consider, first, the interpretation according to which statistical and mechanistic evidence are different in virtue of being produced via different methods: whereas statistical evidence is produced by randomized experiments and observational variance-based designs, mechanistic evidence is

produced by means of methods such as laboratory research and case studies (see, for example, Beach and Pedersen 2016; Russo and Williamson 2007). However, evidence-gathering methods cannot be divided into two categories that correspond to either mechanistic or statistical evidence. The same methods are frequently used to gain mechanistic evidence and statistical evidence (Illari 2011, p. 143).[6] Observational studies, experiments, and computer simulations can all be used to support claims of both kinds. For example, a behavioral-economics experiment in the lab may generate evidence about the existence of the default effect, as well as whether the effect is due to loss aversion or to one of the other mechanisms (Dinner et al. 2011).

Even randomized field experiments, which are often considered the paradigmatic source of statistical evidence, are sometimes designed to yield evidence of mechanisms. Whereas *policy experiments* test the causal relationship between the policy and the outcome, *mechanism experiments* test the causal relationship between a factor that, according to the theory, intervenes between the policy and the outcome. Ludwig, Kling, and Mullainathan (2011) use the well-known broken-windows theory in criminology to illustrate the difference between policy and mechanism experiments. The theory holds that the presence of minor crimes in a neighborhood also increases serious crime because the effects of minor crimes (such as breaking windows) function as visual cues to potential offenders that criminal behavior is rarely sanctioned. A policy experiment randomizes broken-windows policing across neighborhoods, whereas a mechanism experiment might randomize visual cues (see also Sampson, Winship, and Knight 2013). Hence, the mechanistic-statistical distinction does not align with the distinction between different kinds of methods.

Second, a related intuition is that mechanistic evidence, in contrast to statistical evidence, is qualitative. It has been suggested that whereas randomized controlled trials provide numerical estimates of the size of the treatment effect, mechanistic evidence shows how the interacting components of the mechanism "really" create such an effect. Waldner (2012), for example, argues that mechanistic explanations describe "invariant causal properties" that explain how causal effects are transmitted, and hence cannot be reduced to adding variables between the cause and its effect: how depressing the gas pedal makes the car move faster is explained with reference to combustion and the relationship between torque and force as the invariants that mediate the interaction, not by adding variables that intervene in the process. Presumably, underlying these kinds of claims is the idea that mechanistic evidence reveals the causally continuous process consisting of concrete events and activities.[7]

[6] It is not clear whether Clarke et al. (2014) take the distinction between mechanistic and statistical evidence as being based on kinds of methods. On the one hand, they approvingly quote Illari (2011), who argues for a separation between kinds of evidence and kinds of methods. On the other, given their aim to show that evidence hierarchies, which rank methods, are mistaken in ranking mechanistic evidence lower than statistical evidence, it would seem that the distinction between the two kinds of evidence is one of method or, at least the two kinds correlate with different methods.

[7] As another example of the portrayal of such a process, let us consider how statins reduce the accumulation of cholesterol in the arteries. Being structurally similar to a natural enzyme, statins bind to its site, thus competing with the natural substrate. The competition reduces the rate at which a second molecule is produced, eventually blocking the pathway for producing cholesterol in the liver, and leading to reduced cholesterol levels in the bloodstream.

In many cases this picture is highly misleading, however. Even in EBM there is often no access to direct evidence about productive continuity in the target systems. Instead, knowledge of mechanisms is inferred from detected patterns of correlation (Campaner and Galavotti 2012). The connotation of concreteness in the case of mechanistic evidence is particularly misleading in EBP: it is hardly useful to understand psychological and social mechanisms in terms of physically continuous processes consisting of entities and activities. Instead, aspects such as the practical reasoning of agents, affect, cognitive biases, and social norms tend to play an important role (Hedström and Ylikoski 2010). Given the relatively abstract nature of psychological and social mechanisms, uncovering intermediate causal steps between X and Y may well require the judicious combination of both quantitative and qualitative research tools and evidence.

The fact that mechanisms are frequently uncovered by means of statistical evidence also indicates that the distinction between statistical and mechanistic evidence does not coincide with the distinction between evidence about population-level averages or aggregates, and evidence about individuals (cf. Waldner 2012). Obtaining evidence of mechanisms is certainly a means of dealing with problems caused by treatment heterogeneity in the population, but this does not require revealing "non-statistical" facts about particular individuals. If the worry concerning statistical evidence is that the treatment might affect different sub-populations differently, there are statistical methods for dealing with the problem (see Section 5 below). Here again, as in the case of mechanism experiments, prior mechanistic knowledge and hypotheses about how the policy works are useful guides for identifying which variables for stratification would be most informative, for example (Deaton and Cartwright 2018).

Finally, it is misleading to suggest that statistical evidence establishes causal relations at the macro level, whereas mechanistic evidence establishes such relations at the micro level. A venerable tradition takes mechanisms to be underlying structures that operate on a level lower than the phenomenon to be explained. Hence, the mechanism indicating a causal relationship between, say, unemployment and criminality would involve individual human behaviors and interactions, whereas the mechanism indicating such a relationship between changing defaults and the resulting choice would appeal to the cognitive processes of decision-making. The idea of mechanisms being at a lower level raises a set of difficult questions, however. What is the correct level at which to describe mechanisms for behavioral policy? Although the level of cognitive processes and representations would appear to be a natural contender, why should this be so? It is possible, at least in principle, to dig deeper and to obtain evidence about the neural structures that realize such cognitive mechanisms. In line with Kincaid (1996; 2012), we believe there is no reason to think that mechanisms should always be at a lower level. Moreover, as will become clear below, evidence about psychological processes is not always necessary for judging policy effectiveness.

Having put these arguably problematic connotations of mechanisticness to one side, we seem to have little left of the notion of mechanistic evidence apart from a generic connection with the idea of adding causal detail (see Gerring 2008). Causal detail is typically understood in terms of information about the entities, activities, and the organization involved in the mechanism. However, intuitions notwithstanding, it is not obvious why knowing about entities and activities would be

especially useful for policy in the first place. We therefore now turn to the question of which causal details are relevant to extrapolating policies and why.

5. From mechanistic evidence to evidence about invariance

The main worry among those recommending that EBP should take mechanistic evidence into account arises from seeing evidence-based policy as necessarily involving predictions about the effectiveness of an intervention in populations other than those in which it has been tested. This includes questions about whether the policy will continue to be effective once it is scaled up, and what will happen in the long run. Answering these questions requires making judgments about the conditions under which the policy brings about the outcome. In other words, will the policy continue to bring about its effect if things change? The answer depends on counterfactual information about the factors that affect the relationship between the policy and the outcome. The more a relationship remains invariant under changes in the intervention or in the background conditions, or the more we know about the conditions under which it remains invariant, the easier it is to extrapolate it successfully.

Woodward's manipulationist-counterfactual account of causality captures these ideas in precise terms. In Woodward's terminology, difference-making is not a statistical notion, but an objective relation in the world. Variable X makes a difference to variable Y if and only if there is an ideal intervention on X which, if implemented, would alter Y's value or its probability distribution. Not all statistical evidence is evidence of difference-making: roughly, a statistical dependence between two variables reflects a genuine causal relation, and hence one variable makes a difference to another only when changing the former can bring about regular changes in the latter. This general idea can be formulated more precisely in terms of the concept of *invariance*. A change-relating generalization concerning the difference-making relationship between variables X and Y is invariant to the extent that it continues to hold across possible interventions on X. It is *robust* to the extent that it would continue to hold despite variation in the values of the background variables. Finally, its *scope* is determined by the number of actual contexts across which it remains invariant (Woodward 2003; Woodward 2006). Evidence of causation (invariance) is necessary but not sufficient for extrapolating policies: evidence about robustness and scope is also needed for establishing invariance across different contexts.

Regardless of whether one wishes to commit to Woodward's manipulationist-counterfactual account of causality, we take it as uncontroversial that policymakers are interested in causality because causes are good handles for manipulating their effects, and thus for bringing about or preventing outcomes of interest. Therefore, the evidence desired by a policymaker is information about the policy intervention X being an effective way of changing the outcome Y. This interpretation of the role of causal information also dovetails with the core insight of EBP: pilots, policy experiments as well as various quasi-experimental methods can be considered relatively direct means of obtaining such difference-making evidence.

It follows from the above that evidence concerning entities and activities is not in itself

epistemically valuable for making accurate judgments about invariance for policy-making purposes. On the other hand, mechanistic evidence is valuable for policy insofar as it adds to current knowledge about the stability of the X–Y dependence: knowing how the intervention works in a given context and indicating which changes to other causally relevant factors are likely to cancel out, break down, or even reverse its effect on the outcome, contributes to refining judgments about the robustness and scope of the X–Y dependency.

It would thus seem that mechanistic evidence should not be contrasted to difference-making evidence about the X–Y connection. If, as the difference-making account of causation has it, the causal relationships that make up the mechanism are interpreted as invariant dependencies, then evidence about these causal relationships is also difference-making evidence. According to our componential different-making (CDM) account of mechanistic evidence, the difference between the two concerns the relata: mechanistic evidence contributes to policy in providing *difference-making information about the components of the causal pathway between X and Y*.[8]

The CDM account resolves (many of) the tensions in the notion of mechanistic evidence discussed above (Section 3). For instance, quantitative and qualitative appear not as fundamentally different kinds of evidence, but as different ways of representing difference-making relations. Likewise, evidence from policy experiments and evidence from mechanism experiments are both evidence of difference-making. The difference between them concerns the relata: the former directly concerns X–Y dependency, whereas the latter indirectly feeds into the understanding of such dependency in providing information about the difference-making relationships in the causal structure mediating the X–Y relationship.

How and why, then, does mechanistic information help us to refine judgments about the invariance of the relationship between the policy lever and the outcome? Having accurate knowledge about the causal mechanism mediating the causal effect of X on Y facilitates the making of more detailed counterfactual predictions about how changes in X influence the value taken by Y in situations in which relevant background variables take different values (see Section 5 for details). Although this is mechanistic evidence, it is still difference-making evidence about the causal pathway, and not information of a different kind – such as about productive continuity in the mechanism.[9] In sum, our account explains why it is that knowing the mechanism is useful for extrapolation: opening the black box and seeing how the phenomenon works is useful insofar as such information contributes to refining invariance judgments. For example, evidence about the mediating steps between maternal education and children's improved health in India tells the policymaker about the factors, the state of which determines how intervening on X influences Y. Exporting a policy to a novel context relies on knowledge about the scope of the X–Y dependence, and on knowledge about the state of the

[8]See Claveau (2012) for a view that is similar to ours, and Dragulinescu (2017) for a different line of argument according to which the importance of difference-making information about mechanisms supports rather than compromises the Russo-Williamson thesis.

[9]The CDM account is compatible with different ontological accounts of mechanisms. This allows us to sidestep debates about what mechanisms are (for overviews of such debates see e.g. Andersen 2014; Craver and Tabery 2017; Hedström and Ylikoski 2010; Marchionni 2017).

factors determining the robustness of the causal relationship.

That said, there are limits to the kind and quantity of information about mechanisms that is necessary or even useful for policy. If one is sure that X–Y dependency is invariant with respect to changes in some of the causal details through which a particular intervention works, then evidence about those details being present in the target is unnecessary. Suppose, for example, that changing a default option in the source population increases saving rates via the recommendation effect. Not all the details of the continuous causal process (e.g., the bank in which people have their savings accounts, or the time of day when the information about the change in default is delivered) are relevant in terms of making invariance judgments. Hence, not all the details of the mechanism are useful for policy extrapolation, and the policy-relevant contribution of knowledge of mechanisms can be captured in terms of difference-making. Of course, which causal details are relevant is always context-dependent. Without a principled way of thinking about relevance, however, we are groping in the dark. Thinking in terms of invariance provides the necessary criterion for relevance, a principled way of identifying, among the thicket of causal details, the ones that are likely to be relevant to the effectiveness of a policy.

Both information about robustness and information about scope facilitate extrapolation, but in different ways. On the one hand, causal dependencies underlying policies tend to rely on non-robust (sub)mechanisms related to the practical reasoning of the actors involved, for example. These types of mechanism would be disrupted even by slight neurological changes. However, given that, in practice, there is little variation in such properties between people (large scope), the sensitivity of the dependence tends not to be a problem for policy extrapolation. On the other hand, knowledge of the robustness of a mechanism reduces the need to learn about the scope of a particular dependence, as there are good reasons to believe that it holds under various changes to the background conditions.

The emphasis on invariance distinguishes our proposal not only from approaches that treat mechanistic evidence as evidence that is distinct in kind from statistical and difference-making evidence, but also from Cartwright and Stegenga (2011)'s position: like us, they regard information about mechanisms as a means of finding out what other causal factors need to be present for a policy to be successful. The added value of relating mechanistic information to invariance assessment, as we do, is that it helps to sort out the kind of mechanistic information that is most valuable in the policy world of limited time and resources. For example, thinking in terms of invariance allows us to distinguish between two cases with which we might be confronted: one in which the policy works only when the supporting factors are exactly right, and the other in which the policy brings about the outcome *in spite of* (at least some) changes in the supporting factors.[10]

One might object that the CDM account fails to address a major shortcoming of statistical evidence, which cannot be remedied if mechanistic evidence is understood in the minimal sense presupposed by the account. As we note in Section 4, one of the problems with statistical trials is

[10]Cartwright and Efstathiou (2011) also discuss invariance, but worry that it is not sufficient to capture changes in the underlying causal structure. The notion of robustness introduced above is meant to capture stability with respect to such changes.

that they typically provide only estimates of the *average treatment effect*. When causal homogeneity with respect to the treatment cannot be assumed, such an average does not always convey useful information about individuals. In fact, in some cases individual-level causal effects might even be reversed, and this could make the planned policy ethically problematic. Is not this – the need to get past mere averages – a good reason to look for non-statistical mechanistic evidence?

We certainly agree that non-statistical methods (e.g., case studies, process tracing) have their uses for identifying causal heterogeneity between individuals, and they may generate hypotheses about the different causal pathways at work in the population. Suppose that changing defaults works well for some people and poorly for others, and that this is so because the policy works through different pathways depending on the varying sets of individuals' attributes. What we would like to know is which attributes are related to which pathway. However, no matter how hypotheses about the different causal pathways come to be formulated, showing how widespread the different mechanisms are in the population of interest still relies on statistical methods, and the causal model can be refined to accurately describe the situation. It thus seems that the problem is not really that statistical trials "only" give averages. Averages (and other measures of central tendency) are useful and economical summaries of distributions of properties in a population (and information must be compressed somehow if one is to say anything general about it). Dividing the population into causally (more) homogeneous subpopulations and calculating conditional average treatment effects corresponding to the different subgroups allows us to draw more precise causal conclusions in contexts characterized by causal heterogeneity.

6. Evidence integration beyond meta-analysis

According to the CDM account, a piece of mechanistic evidence provides difference-making information about one of the links on the causal pathway between the policy and the outcome. How are such pieces of information utilized in extrapolative inference for policy purposes? We now show that the need to take mechanistic evidence into account poses a challenge of evidence integration that is not typically addressed in the philosophical debates concerned with evidence hierarchies and evidence amalgamation (cf. Stegenga 2013). Causal diagrams, which are often associated with the manipulationist-counterfactual theory of causality, can be used as tools for integrating evidence for policy (e.g., Pearl 2009; Steel 2008; Woodward 2003). We refer to such diagrams as *policy graphs*.

CDM evidence concerns dependencies between the mediating and modulating variables, and it can be used only indirectly to derive conclusions about treatment effects. Instead of combining several pieces of (sometimes discordant) evidence in order to choose from among a set of competing theoretical claims, the challenge confronting us in this case is how to combine complementary pieces of evidence that shed light on different aspects of the same policy scenario. Policy graphs provide a systematic framework for representing the information we already possess as well as the information we should look for so as to improve our inferences.

We address a typical scenario in which a policymaker has (i) some prior knowledge of the

processes involved (i.e. hypotheses about the psychological and social mechanisms through which the policy might work), (ii) evidence about the causal effect of the policy on the outcome in the source, such as from a pilot experiment, and (iii) only partial knowledge of the relevant causal structure at the target site. Building a causal model based on these resources could guide the search for the relevant supporting factors in the target and for mechanistic "fingerprints", in other words easily detectable pieces of evidence that are not likely to be caused by alternative mechanisms. Assembling and revising the causal model on the basis of various pieces of evidence obtained from the source and the target could help the policymaker to predict intervention outcomes and thereby design better interventions.

Causal structures can be represented in terms of directed graphs, in which the nodes correspond to variables, and the directed edges to relationships of direct causation between them (Fig 1). If we further assume that it is the causal structure that gives rise to the probability distributions of the variables and connect the graph to the joint distribution by means of the causal Markov condition, we have a powerful tool for causal inference also when working with observational data.[11] Here we need not concern ourselves with the mathematical details of causal inference using directed (acyclical) graphs, however. What we will do instead is suggest how the representational resources associated with causal diagrams can be used to clarify the role of mechanistic evidence in policy extrapolation. Causal diagrams offer a useful way of representing the information we have about the difference-making relationships between the causal factors involved. Refined invariance judgments are based on that counterfactual information.

Fig. 1a represents three different causal scenarios. To return to the default effect example, variable X depicts the policy lever, the default, and Y the outcome variable, i.e. the saving rate. The connection is mediated by variable M_1, M_2 or M_3, which stand for the three hypothesized mechanisms. If (and only if) intervening on X brings about a change in the probability distribution of Y, then X is said to be a *total cause* of Y. More fine-grained causal claims can be made by employing the notion of *direct cause*. In the figure, X is a direct cause of M_i (represented by the presence of an arrow between the two) and M_i are putative direct causes of Y. In each case, the two edges from X to Y form a causal pathway.

A pathway consisting of merely one mediating variable is a minimal example of a representation of a mechanism between X and Y (see Woodward 2002; Woodward 2013). Empirical information concerning the causal dependencies between variables on that pathway captures the aspect of the notion of mechanistic evidence that is needed for the successful extrapolation of policies. To be clear, this is not to claim that mechanisms are nothing other than sets of mediating variables. The ontological question of what mechanisms are can be separated from the epistemic question concerning the level of detail at which a description of a mechanism is useful for policy purposes (see footnote 9). The CDM account captures the idea that difference-making information "screens

[11]For detailed expositions of DAGs see e.g. Kincaid (2012), Steel (2008), Steel (2013), and Claveau (2012) in philosophy, and Sampson, Winship, and Knight (2013), Ludwig, Kling, and Mullainathan (2011), Morgan and Winship (2015), and Pearl and Bareinboim (2014) in social science.

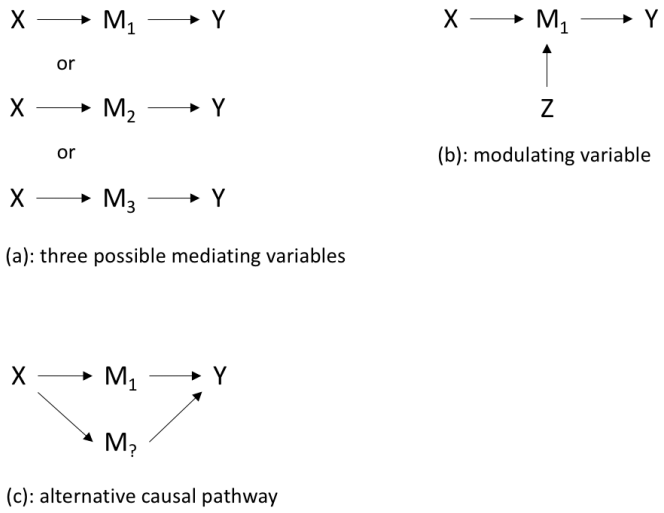

Figure 1: Examples of causal diagrams

off" other information about the causal pathway.

Mediating variables, in other words those falling on the X–Y pathway, are not the only kind of variables relevant to making invariance judgments. Typically, mechanisms are not insulated from their environment, but their functioning is modulated by variables that do not fall on the pathway but from which there is a directed path to the outcome variable (Z in Fig. 1b). An example of a modulator for the mechanism of cognitive effort avoidance would be time constraints: it might be the case, for example, that without perceived time constraints people spend time evaluating options instead of simply going for the default option. Because of such interaction effects, the state of a modulator variable can make all the difference as to whether an intervention is effective or not.

Therefore, knowing which one of the mechanisms underlies the successful intervention in the source alerts us to the possible presence of a relevant modulator.[12] Intuitively, the label "mechanistic" might seem to apply only to mediators, but according to the CDM account evidence concerning modulators should also be considered mechanistic. Both provide the kind of difference-making evidence about the causal detail needed for invariance judgments and extrapolation. In contrast, confounders (common causes of X and Y) and covariates (variables related to X or Y that do not alter the X–Y relation) are examples of third variables that are not mechanistic according to our definition.

Cases (a) and (b) in Fig. 1 describe simple causal structures in which the intervention affects the outcome only through a single causal pathway. Often this is not the case. A more complicated, yet

[12]Pearl and Bareinboim (2014) and Steel (2013) have developed formal rules for deciding the identifiability of causal effects in extrapolation in such situations under partial knowledge of the model and target populations.

still simple, example is one in which the intervention affects the outcome through two distinct causal pathways (Fig. 1c). Suppose, for example, that X only begins to affect Y through $M_?$ after a certain period of time has elapsed, as in the case in which changing the default might lead at least some people in the target population to gradually learn about their own preferences, and which in the long run might diminish the aggregate effect of the policy. A policy experiment might not be designed to capture this kind of long-term effect. Hence, as Grüne-Yanoff (2016) rightly argues, having the right causal model of how X brings about Y alerts us to these kinds of possibilities.

Properties of the mechanism observed in the source allow us to calibrate our assessment of how risky the extrapolative inference is: in other words, they have important consequences in terms of how cautious we should be in making inferences about the target based on the information we have about the source. On the one hand, if the source mechanism is robust, invariant across a wide range of changes, this presumably reduces the number of checks that must be made in the target. On the other hand, its fragility increases the burden on the policy maker to find evidence (mechanistic fingerprints) that the same mechanism is indeed at work in the target context.

Note, however, that asking whether the *same* mechanism works in the source and in the target may be a misleading approach to extrapolation: it must always be assumed that there are some causally relevant differences between the source and the target (Steel 2013). It is nevertheless possible to distinguish between two different scenarios. In the first one, the causal diagram constructed on the basis of the evidence obtained from the source population does not apply to the target population in any (non-gerrymandered) way. For example, in the source the psychological mechanism underlying the default effect is the avoidance of cognitive effort, whereas in the target the default option is chosen due to loss aversion. This is a clear instance of getting the mechanism wrong.

The mechanism in the second scenario is sensitive to modulating variables (selection variables à la Pearl and Bareinboim 2014), which take on different values in the source and in the target. Suppose, for example, that the time constraints that modulate the relationship between default and outcome behavior are not present to the same degree in the target: this would lead to a significant reduction in the effect of manipulating the default option. It may be that certain values of the modulator even make the causal effect disappear in the target. In such a case, is the mechanism in the target the same mechanism as in the source? This question has no clear answer. Although the same causal diagram could be used to describe the mechanisms in the source and in the target, it would seem odd to suggest that the mechanism is the same when, due to the modulator breaking the connection, there is no functioning causal pathway in the target population. Rather than trying to determine whether the same mechanisms are at work in the source and in the target, therefore, we should be asking whether the same causal model applies in both.

Causal diagrams have their shortcomings, of course. Although we believe that the simple ones introduced here suffice to clarify the CDM account, in causal inference the diagrams often have to be combined with more specific assumptions about the nature of the dependencies between the variables (expressed, for example, by structural equations). Furthermore, directed acyclical graphs cannot generally include non-causally-related (logically, mereologically etc.) variables (see Woodward 2015). These limitations reflect the fact that increased inferential power tends to come

at the cost of applicability. Moreover, causal diagrams do not play a direct role in the discovery of mechanistic hypotheses. Awareness of which variables to measure and include in the graph tends to come from theoretical knowledge about relevant mechanisms. Nevertheless, collecting and organizing information in a causal diagram facilitates policy-relevant inferences. Knowing that a causal pathway on which a policy relies consists of a set of relatively invariant edges, for example, should increase confidence in the possibility of successfully extrapolating it to other populations. Furthermore, if we understand that – and how – a causal pathway is sensitive to the state of a modulating variable we should know which supporting and undermining factors to measure in the target population. The model allows us to separate the causal detail that is relevant to the policy from the irrelevant parts of the causal fabric. Obtaining evidence is usually slow and costly, and diagrammatic representations are useful because they help in focusing evidence collection on the factors that matter most for extrapolation.

7. Implications for behavioral policy

We have argued that statistical and mechanistic evidence do not constitute two different kinds of evidence. We also suggest that all evidence fed into the causal policy model should ultimately function as *evidence about invariance*, i.e. evidence that supports and refines judgments about whether the X–Y dependency remains invariant under interventions on X and changes in other relevant causal variables (modulators). We now draw the implications of this view for debates concerning the respective roles of mechanistic knowledge and mechanistic evidence in evidence-based behavioral policy.

Behavioral policy, and EBP more generally, has typically tried to distance itself from theory. Reliance on RCTs, which are believed to need relatively little prior knowledge and few theoretical assumptions, is associated with an a- or even anti-theoretical attitude, originally intended to protect policy-making from subjective opinion and scientific folklore disguised as "theory". Theoretical knowledge is nevertheless useful both in the design and the interpretation of RCTs, and often in the extrapolation of results from one context to another. As Grüne-Yanoff (2016) shows, the fact that behavioral policy tends to rely solely on alternative theoretical hypotheses about the psychological and social mechanisms involved need not be an obstacle to a theory-friendly approach to policy-making. Different theories will typically predict different effect sizes for interventions and, most importantly, postulate different mediating and modulating variables that determine the conditions under which the X–Y link remains invariant. In the absence of knowledge about which of the possible mechanisms is in place in the target population, or of whether the population is heterogeneous with respect to the policy (different mechanisms in different individuals), considering different mechanistic hypotheses in parallel could indicate the kind of contingencies to prepare for and the mechanistic fingerprints to look for.

At the same time, recognition that theory is useful for extrapolation does not imply that full-fledged mechanistic or processual models are always necessary (pace Gigerenzer and colleagues), or

that consensus on the correct account of the whole cognitive architecture should be reached before a behavioral policy intervention is carried out. Incrementally adding information about mediators and modulators into a causal model may well be sufficient to improve external validity. If, as seen above, even small additions concerning mediators and modulators help extrapolation, then any result that psychologists can agree on could be used. Moreover, researchers could even exploit theoretical disagreements about competing mechanisms for the purposes of policy planning, as competing mechanisms suggest different mechanistic fingerprints to search for and modulating variables to measure. In sum, given that there is no need to choose sides about big theory (cf. Bond, 2009), behavioral policy need not go to the other extreme and be anti-theoretical in order to proceed.

We therefore agree with Grüne-Yanoff (2016) that evidence about mechanisms is useful for behavioral policy. However, this does not imply that "simply providing better difference-making evidence does not compensate for a lack of mechanistic evidence" (Grüne-Yanoff 2016, p. 481). At least in principle it should be possible to make accurate judgments about invariance without evidence concerning mediating mechanisms. For the sake of argument, let us consider a brute-force approach to studying invariance: applying a behavioral policy across a range of contexts and always finding that it works. Barring pure inductive skepticism, such evidence should make us more confident that the X–Y relationship is insensitive to differences across contexts. It could be, for example, that some cognitive biases are almost like "universal biological responses" in that they are predictably triggered in certain decision situations, and hence are quite robust across all such contexts.[13] Extrapolating policies that rely on such black-boxed but robust biases would then be based on invariance assessments established without direct reliance on mechanistic evidence. Strictly speaking, knowledge of mechanisms is not always necessary for extrapolation.

8. Conclusion

We have argued that extrapolating behavioral policies requires evidence about the invariance of the causal relationship between the policy lever and the outcome. Knowledge of whether and when a given causal relationship remains stable typically (but not necessarily) relies on evidence about the variables that mediate and those that modulate it, that is, on mechanistic evidence. In contrast to what is often assumed in the literature, such variables are to be found at any level, and evidence about them is, in principle, obtainable by any method. Evidence about different aspects of the policy scenario can be accumulated in causal diagrams, in other words models of mechanisms understood as structured representations of difference-making relations. A theoretical understanding of the causal processes involved is required to construct such models and to use them for the integration of evidence. Hypotheses about which variables should go into the graph and hence where to look for relevant evidence are generated on the basis of prior theoretical knowledge about mechanisms.

[13]The pervasiveness and robustness of cognitive biases has long been a matter of dispute in the empirical research on human decision-making (see Bond 2009).

Acknowledgments

This research has been supported by the University of Helsinki and the Academy of Finland. Previous versions of this paper have been presented at *INEM 2015* (Cape Town), the Philosophy of Science seminar at the University of Helsinki (Helsinki) 287388, 259403, 251192, the *First Bayreuth Workshop in Philosophy of Economics* (Bayreuth), *Current trends in the Philosophy of the Social Sciences III* (Helsinki), and *Models and Explanations in Economics* (Rostock). We would like to thank the participants of these events for their comments. We would especially like to thank Till Grüne-Yanoff and Jaakko Kuorikoski for reading through an earlier draft.

References

Andersen, H. (2014). "A field guide to mechanisms: Part i". *Philosophy Compass* 9.4, pp. 274–283.

Beach, D. and R. B. Pedersen (2016). "Selecting Appropriate Cases When Tracing Causal Mechanisms". en. *Sociological Methods & Research*, p. 0049124115622510.

Behavioral Insights Team (2013). *Applying behavioral insights to charitable giving*.

Berg, N. and G. Gigerenzer (2010). "As-if behavioral economics: Neoclassical economics in disguise?" *History of economic ideas*, pp. 133–165.

Bond, M. (2009). "Decision-making: Risk school". *Nature* 461.7268, pp. 1189–1192.

Camerer, C. (1999). "Behavioral economics: Reunifying psychology and economics". *Proceedings of the National Academy of Sciences of the United States of America* 96.19, pp. 10575–10577.

Campaner, R. (2011). "Understanding mechanisms in the health sciences". *Theoretical Medicine and Bioethics* 32.1, pp. 5–17.

Campaner, R. and M. C. Galavotti (2012). "Evidence and the Assessment of Causal Relations in the Health Sciences". *International Studies in the Philosophy of Science* 26.1, pp. 27–45.

Cartwright, N. (2012). "Presidential address: Will this policy work for you? Predicting effectiveness better: How philosophy helps". *Philosophy of Science* 79.5, pp. 973–989.

Cartwright, N. and S. Efstathiou (2011). "Hunting causes and using them: Is there no bridge from here to there?" *International Studies in the Philosophy of Science* 25.3, pp. 223–241.

Cartwright, N. and J. Hardie (2012). *Evidence-based policy: a practical guide to doing it better*. Oxford University Press.

Cartwright, N. and J. Stegenga (2011). "A theory of evidence for evidence-based policy". *Proc Br Acad* 171, pp. 289–319.

Clarke, B. et al. (2013). "The evidence that evidence-based medicine omits". *Preventive Medicine* 57.6, pp. 745–747.

— (2014). "Mechanisms and the Evidence Hierarchy". *Topoi* 33.2, pp. 339–360.

Claveau, F. (2012). "The Russo-Williamson Theses in the social sciences: Causal inference drawing on two types of evidence". *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences* 43.4, pp. 806–813.

Craver, C. and J. Tabery (2017). "Mechanisms in Science". *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Spring 2017. Metaphysics Research Lab, Stanford University.

Deaton, A. and N. Cartwright (2018). "Understanding and misunderstanding randomized controlled trials". *Social Science & Medicine* 210, pp. 2–21.

Dinner, I. et al. (2011). "Partitioning default effects: Why people choose not to choose". *Journal of Experimental Psychology: Applied* 17.4, pp. 332–341.

Dragulinescu, S. (2017). "Mechanisms and Difference-Making". *Acta Analytica* 32.1, pp. 29–54.

Gerring, J. (2008). "Review article: The mechanismic worldview: Thinking inside the box". *British Journal of Political Science* 38.1, pp. 161–179.

Gigerenzer, G., R. Hertwig, and T. Pachur (2011). *Heuristics : the foundations of adaptive behavior*. New York: Oxford University Press.

Grüne-Yanoff, T. (2016). "Why behavioural policy needs mechanistic evidence". *Economics and Philosophy* 32.3, pp. 463–483.

Halpern, D. (2016). *Inside the nudge unit: How small changes can make a big difference*. Random House.

Hedström, P. and P. Ylikoski (2010). "Causal mechanisms in the social sciences". *Annual Review of Sociology* 36, pp. 49–67.

Howick, J. (2011). "Exposing the vanities-and a qualified defense-of mechanistic reasoning in health care decision making". *Philosophy of Science* 78.5, pp. 926–940.

Howick, J., P. Glasziou, and J. Aronson (2010). "Evidence-based mechanistic reasoning". *Journal of the Royal Society of Medicine* 103.11, pp. 433–441.

Illari, P. M. (2011). "Mechanistic evidence: disambiguating the Russo–Williamson thesis". *International Studies in the Philosophy of Science* 25.2, pp. 139–157.

Joffe, M. (2013). "The Concept of Causation in Biology". *Erkenntnis* 78.SUPPL.2, pp. 179–197.

Johnson, E. and D. Goldstein (2003). "Do Defaults Save Lives?" *Science* 302.5649, pp. 1338–1339.

Katsikopoulos, K. (2014). "Bounded rationality: the two cultures". *Journal of Economic Methodology* 21.4, pp. 361–374.

Kincaid, H. (2012). "Mechanisms, causal modeling, and the limitations of traditional multiple regression". *Oxford Handbook of the Philosophy of the Social Sciences*, pp. 46–64.

Kincaid, H. (1996). *Philosophical foundations of the social sciences: Analyzing controversies in social research*. Cambridge University Press.

La Caze, A. (2011). "The role of basic science in evidence-based medicine". *Biology and Philosophy* 26.1, pp. 81–98.

Lourenço, J. S. et al. (2016). "Behavioural insights applied to policy: European report 2016". *Web: http://publications. jrc. ec. europa. eu/repository/bitstream/JRC100146/kjna27726enn_new. pdf. Zugegrrifen Zugegriffen: am* 5.05, p. 2017.

Ludwig, J., J. Kling, and S. Mullainathan (2011). "Mechanism experiments and policy evaluations". *Journal of Economic Perspectives* 25.3, pp. 17–38.

Marchionni, C. (2017). "Mechanisms in economics". *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, pp. 423–434.

Moneta, A. and F. Russo (2014). "Causal models and evidential pluralism in econometrics". *Journal of Economic Methodology* 21.1, pp. 54–76.

Morgan, S. L. and C. Winship (2015). *Counterfactuals and causal inference*. Cambridge University Press.

Oliver, A. (2013). *Behavioural public policy*. Cambridge University Press.

Pearl, J. and E. Bareinboim (2014). "External validity: From do-calculus to transportability across populations". *Statistical Science* 29.4, pp. 579–595.

Pearl, J. (2009). *Causality*. Cambridge university press.

Russo, F. and J. Williamson (2007). "Interpreting causality in the health sciences". *International Studies in the Philosophy of Science* 21.2, pp. 157–170.

Sampson, R., C. Winship, and C. Knight (2013). "Translating causal claims: Principles and strategies for policy-relevant criminology". *Criminology and Public Policy* 12.4, pp. 587–616.

Steel, D. (2008). *Across the Boundaries: Extrapolation in Biology and Social Science*. Oxford: Oxford University Press.

— (2013). "Mechanisms and extrapolation in the abortion-crime controversy". *Mechanism and Causality in Biology and Economics*, pp. 185–206.

Stegenga, J. (2013). "An impossibility theorem for amalgamating evidence". *Synthese* 190.12, pp. 2391–2411.

— (2014). "Down with the Hierarchies". *Topoi* 33.2, pp. 313–322.

Thaler, R. H. and S. Benartzi (2004). "Save more tomorrow™: Using behavioral economics to increase employee saving". *Journal of political Economy* 112.S1, S164–S187.

Thaler, R. and C. Sunstein (2008). *Nudge: improving decisions about health, wealth, and happiness*. eng. New Haven ; London: Yale University Press, 2008, ©2008.

Waldner, D. (2012). "Process tracing and causal mechanisms". *The Oxford Handbook of the Philosophy of Social Science*, pp. 65–84.

Woodward, J. (2002). "What is a Mechanism? A Counterfactual Account". *Proceedings of the Philosophy of Science Association* 2002.3, pp. 366–377.

— (2003). *Making things happen : a theory of causal explanation*. Oxford studies in philosophy of science. New York: Oxford University Press.

— (2006). "Sensitive and insensitive causation". *Philosophical Review* 115.1, pp. 1–50.

— (2013). "Mechanistic explanation: Its scope and limits". *Aristotelian Society Supplementary Volume*. Vol. 87. Wiley Online Library, pp. 39–65.

— (2015). "Interventionism and Causal Exclusion". *Philosophy and Phenomenological Research* 91.2, pp. 303–347.

Worrall, J. (2002). "What evidence in evidence-based medicine?" *Philosophy of Science* 69.3, S316–S330.