When Experiments Need Models

Donal Khosrowi[1]

Abstract

This paper argues that an important type of experiment-target inference, extrapolating causal effects, requires models to be successful. Focusing on extrapolation in Evidence-Based Policy, it is argued that extrapolation should be understood not as an inference from an experiment to a target directly, but as a hybrid inference that involves experiments and models. A general framework, METI, is proposed to capture this role of models, and several benefits are outlined: (1) METI highlights epistemically significant interactions between experiments and models, (2) reconciles some differences among existing accounts of experiment-target relationships, and (3) facilitates critical appraisal of inferential practices from experiments.

1. Introduction

Experiments and models are routinely used as surrogate systems in the social sciences (Mäki 2009): they are studied independently of target systems of interest, allow us to isolate phenomena and exercise discriminate control over

[1]Leibniz University Hannover, Germany

Corresponding Author:
Donal Khosrowi, Leibniz University Hannover, Im Moore 21, Hannover 30167, Germany.
Email: donal.khosrowi@philos.uni-hannover.de

them, and they facilitate inferences about these phenomena as they occur in distinct targets.

The existing literature has detailed how experiments and models, respectively, relate to targets, permit learning about them, and how their epistemic capabilities can differ. Models (and simulations drawing on them) are thought to relate to targets by means of *abstract* relationships such as formal similarity, resemblance, (partial) isomorphism, etc., and to permit reliable inferences when such relationships are adequately realized (e.g., Da Costa and French 2003; Frigg and Nguyen 2020; Giere 2010; Khosrowi 2020; Mäki 2005; Weisberg 2013). Experiments, by contrast, are often thought to promote learning about targets in virtue of exhibiting *concrete* relationships, such as material analogy, ontological equivalence, representativeness, or specimenhood (e.g., Currie and Levy 2020; Guala 2002, 2005; Morgan 2003, 2005), and there has been extensive debate concerning whether experiments are epistemically privileged in virtue of such concrete relationships (see Parker 2009; Winsberg 2009; and Parke 2014, who resist such conclusions).

An issue that has received somewhat less attention is how experiments and models can be *integrated* to facilitate learning about the world.[1] In this paper, I focus on the practice of extrapolating causal effects from experiments, which is increasingly common in various social sciences, including particularly in Evidence-Based Policy and economics. I argue that extrapolation requires (good) models to be successful. To extrapolate a causal effect from a study population to a novel target, one needs to support that populations are sufficiently similar, or account for how they differ. This is where models play important but underappreciated roles (see Cartwright and Stegenga 2011). Good causal models of both populations are needed for (1) guiding the search for similarities and differences, (2) making these accessible to investigators, and (3) deriving predictions about the effects of interest.

---

[1] I will not distinguish between experiments and models in a principled way, nor assume that a clear-cut intensional distinction is possible (see Morgan 2003; 2005; Guala 2002; 2005; Mäki 2005; Parker 2009; Winsberg 2009; Parke 2014 for discussions). First, the tools I consider here, that is, RCTs and structural/graphical causal models, are robustly identified as experiments and models over several such distinctions. Second, while it seems possible to conceive of these tools differently, for example, to think of randomized controlled trials as material models (like animal models in biology), this would perhaps change the main claim developed here to "some models need (good) further models to permit good inferences," but leaves the nuances elaborated and the general capabilities of the framework developed untouched. So it seems that finding ways to make epistemically pertinent distinctions between experiments and models is largely orthogonal to my project.

These crucial roles for models suggest that it is unhelpful to understand extrapolation as an inference proceeding directly from an experiment to a target. Rather, I argue it is better understood as a hybrid inference, called *Model-Mediated Experiment-Target Inference* (METI), which involves both experiments and models.

METI is detailed as a general framework to elucidate several important aspects of experiment-target inference. First, it highlights epistemically significant interactions between experiments and models: together, they enable conclusions not accessible with either tool alone. Second, METI emphasizes that experiment-target *inference* proceeds through thick layers of *abstract* experiment-target relationships. This departs from existing accounts like Morgan's (2003, 2005) and Guala's (2002; 2005), which characterize experiment-target inference as anchored in concrete relationships obtaining directly between experimental and target systems. Accommodating related arguments by Parker (2009), METI maintains that both concrete *and* abstract relationships play important roles, and neither takes priority over the other. Third, METI helps critically appraise inferential practices: extrapolation always requires models, but the models relied on in practice are likely to remain implicit, difficult to scrutinize, and inadequate for underwriting good inferences.

I proceed as follows. *Section 2* characterizes extrapolation in the context of Evidence-Based Policy, elaborates the challenges encountered there, and details how extrapolation requires models. *Section 3* focuses on two strategies for extrapolation, explaining how both involve models but differ in how explicit these are made. *Section 4* elaborates METI as a framework for understanding (some kinds of) experiment-target inference, highlights several desirable features of this framework, and considers questions about its relation to existing accounts as well as its scope. *Section 5* responds to an objection. *Section 6* concludes.

2. Extrapolation in Evidence-Based Policy

Evidence-Based Policy (EBP) is the movement according to which public policy should be based on high-quality evidence of 'what works'. Specifically, using experimental evidence from randomized controlled trials (RCTs) and other credible study designs, EBP seeks to clarify whether policies (e.g., universal basic income) are effective in doing what they are supposed to do (e.g., improving individual's welfare). Such evidence, in turn, is systematically reviewed, summarized, and collated in *evidence libraries* that offer decision-makers up-to-date evidence on interventions addressing a wide range of policy issues (Duflo 2004; Parkhurst 2017 ch.2).

In pursuing these goals, EBP faces a substantial obstacle: extrapolating causal effects measured in a study population to an eventual policy target. Extrapolation is challenging because populations often differ in causally important respects (Vivalt 2020): an intervention distributing free bed-nets to decrease malaria infection rates might be highly effective in one place, but fail in another because individuals experience incentives to use the nets for fishing instead (McLean et al. 2014). More generally, the causal mechanisms underwriting the effectiveness of an intervention in one place might be disrupted in another, institutions and individuals' dispositions to respond to interventions can differ, crucial background conditions needed for an effect might be absent, and so forth. So, in making an inference to a new population, extrapolation requires conscientiously seeking out information pertaining to causally relevant similarities and differences, as failure to do so runs the risk of mistaken conclusions.

Extrapolation can take various forms. Some problems are easier to handle than others, such as when study populations are representatively sampled from a target. In many cases, however, evidence is brought to bear on questions about populations that can importantly differ from a study population.[2] In the following, I assume that we face these more severe extrapolation problems.

In these cases, we proceed from a causal effect of X on an outcome Y estimated in A and endeavour to draw a conclusion about the X to Y-effect in a novel target B, where A and B can differ in various ways. A useful way of organizing these differences is to distinguish three layers of causal analysis (Khosrowi 2019). Differences can occur concerning: (1) the structure of causal mechanisms, for example, whether there is a causal relationship between certain variables or not; (2) the finer-grained details of individual causal relationships between variables, such as whether a relationship is best described linearly or non-linearly, or what values important structural parameters shaping these relationships take; and (3) the values of so-called *modifying variables*[3], that is, variables that can meddle with the sign and magnitude of an effect.

---

[2]Of course, whether populations *do* in fact differ, and how such differences bear on an effect in a target, is generally unknown to investigators when endeavouring to make an inference.

[3]I use 'modifying variables' to refer to both *moderating* and *mediating variables* (Baron and Kenny 1986). W is a *moderating variable* if different values W of induce different marginal effects of X on Y. Z is a *mediating variable* if it transmits the effects of X on Y, such as in the causal mechanism $X \rightarrow Z \rightarrow Y$. While mediating variables are traditionally not considered to *modify* an X-Y-effect since they merely transmit variation from X to Y, there are cases where they can do so, such as when Y is a non-linear function of Z, so Z's initial value matters for how much a given change in X changes Y.

Successful extrapolation requires that we can either assume that populations are similar at each of these three layers of causal analysis, or, if they differ, that we can express how these differences bear on the effect of interest in a target.[4] Either way, extrapolation is challenging since we need to make extensive causal assumptions about both populations and support these assumptions empirically. Let me outline how models can help make these assumptions explicit, which is an important precondition for attempts at supporting them.

2.1. Extrapolation Requires Models

Extrapolation requires models in two senses. The first is that in making an inference to a new environment one needs to make *some* assumptions about how populations are related causally (Banerjee and Duflo 2009:160). Even the most unsophisticated form of extrapolation, that is, *naïve extrapolation* (cf. Steel 2009 on *simple induction*), which predicts that an effect will be the same in a target as in a study, will need to make some implicit causal assumptions—in particular, that the causal makeup of both populations is such that an effect is the same in B as in A (cf. Cartwright 2013a). These assumptions, in turn, can be understood to constitute a class of *implicit causal models* of both populations. This class includes models asserting that the two populations are exactly causally alike in all respects, but also models asserting that the two populations differ, but in such a way that these differences do not matter. While a naïve extrapolator is not committed to any *specific* models among this class, they are at least committed to believing that *some* models from this class adequately characterize relevant causal features of both populations. Otherwise, they would not be justified in expecting an effect to be the same in B as in A.[5]

Implicit causal models are typically unhelpful (cf. Cartwright 2013a): the details of their assumptions remain vague and cannot be scrutinized; it remains unclear what empirical support might be needed to justify their assumptions, whether such support is available, and how it should be sought; and they cannot be easily investigated and manipulated to exploit their derivational resources. What is more, the implicit models relied upon by naïve

---

[4]I understand *successful* extrapolation here as requiring at least that an effect is correctly predicted (to within some margin) and that the prediction is adequately justified.
[5]An *exceedingly* naïve extrapolator might elude the committment to believe in *any* model by simply taking a leap of faith without seeking justification. I do not consider this approach here.

extrapolation are also particularly unhelpful, since populations can routinely differ in important ways (Cartwright 2013b; Fuller 2019; Reiss 2019; Steel 2009; Vivalt 2020). So not only is naïve extrapolation relying on models likely to lack adequate support, resist scrutiny, and elude further analysis, but is also likely to rely on models that are inadequate, thereby precluding successful extrapolation.

Conversely, as others have stressed (see Cartwright and Stegenga 2011), successful extrapolation is importantly facilitated by models that are explicit and well-supported. These virtues go hand in hand: the more explicit and detailed causal models are, including their ability to resolve different layers of causal analysis (e.g., those outlined earlier), the easier it is to tell which particular assumptions they encode, which assumptions are in need of support, and whether there is enough support in their favor. Recognizing these important benefits provides us with a second sense in which extrapolation requires models: one that arises not out of necessity but one that is normative in character and up to us to cater to.[6] Successful extrapolation requires not just *some* model, but *good* models: models that are explicit and well-supported.[7]

Unfortunately, there are reasons to think that extrapolation in EBP often proceeds against the background of bad, implicit models. While some authors have recently more fully acknowledged the intricacies involved in extrapolation (e.g., Peters, Jain, and Gaarder 2019), methodological guidelines issued by central EBP institutions remain unhelpfully thin on issues of extrapolation (e.g., Guyatt et al. 2011; Schünemann et al. 2019). They suggest that it is essential to consider whether populations are sufficiently similar, but provide little guidance beyond the level of such slogans. So, unless extrapolators move considerably beyond what the guidelines ask, extrapolation risks proceeding in a methodological vacuum and the models underlying these

---

[6]One might say that the first sense in which extrapolation requires models is normative in character, too, since it presupposes that extrapolation *should* be justified somehow (rather than relying on a leap of faith). I take this presupposition to be widely shared, however, so the normative character of extrapolation requiring *some* models does not need particular emphasis.

[7]The points made here are similar to those made by Cartwright (2013a), who argues that (good) extrapolation is best understood in terms of valid and sound *effectiveness arguments*. Focusing on models seems preferable to me because they offer cognitive and epistemic advantages that extend beyond the capabilities of arguments, for example, they are easier to manipulate and are analytically more tractable, as will be highlighted in *Section 3*. I am open, however, to recruiting models for the purpose of devising sophisticated effectiveness arguments.

inferences may remain implicit, poorly justified, and often inadequate causal accounts of the populations of interest.

To help elaborate more fully the important role that good, explicit models can play, let me turn to consider two approaches to extrapolation: statistical and causal graph-based approaches. While both involve models, they differ in how explicit these models are made, which bears on their ability to resolve important causal detail and promote the process of justifying extrapolation.

3. Two Approaches to Extrapolation

Among the various approaches to extrapolation that have been proposed and discussed (e.g., Cartwright 2012, 2013a, 2013b; Cartwright and Hardie 2012; Guala 2010; Muller 2013, 2014, 2015; Steel 2009, 2010; van Eersel, Koppenol-Gonzalez, and Reiss 2019), statistical and causal graph-based approaches stand out because they offer resources to overcome a wide variety of problems of extrapolation. Specifically, unlike other approaches, they offer formal recipes for how to adjust an effect estimated in one population to accommodate differences between populations. Let me introduce a working example to illustrate how both approaches work, before highlighting their differences.

Assume we are interested in the effect of a microfinance intervention intended to improve household welfare of the rural poor. The intervention seeks to help people lacking access to credit obtain the necessary capital for starting small businesses, which, in turn, is hoped to increase household income and ultimately welfare. Assume further that we have measured the effect of our intervention in a study population A: the intervention has proven to be highly effective at increasing welfare there. Let us also assume that the microfinance effect (in particular, how successful any investments in small businesses are in yielding increases to household income, and ultimately welfare) is influenced by two modifying variables in A. First, individuals' age is a negative modifying variable: younger individuals have better intuitions about what could be successful business ideas. Second, entrepreneurial ability (in the broadest sense) is a positive modifying variable: individuals who have a better grasp of how to run a small business are better at successfully converting business investments into household income. Finally, let us assume that age and entrepreneurial ability are positively correlated in A. Older individuals, other things being equal, are more likely to have received pertinent education and to have had previous opportunities in starting small businesses and learning from these experiences, which increases their entrepreneurial ability. With this example in place, let me proceed to outline how

statistical and graph-based approaches would handle an extrapolation of the microfinance effect.

Statistical approaches have been proposed in the microeconometrics literature (Crump et al. 2008; Hotz, Imbens, and Mortimer 2005; see Khosrowi 2019; Muller 2013, 2014, 2015 for discussions) and other proposals have since followed their general rationale (van Eersel, Koppenol-Gonzalez, and Reiss 2019). In enabling extrapolation even when populations differ relevantly, statistical approaches focus on differences in the distribution of modifying variables. Let us focus on the role of age in our microfinance example: assume we know that the microfinance effect depends negatively on age in *A*, and that the mean of the age distribution is higher in *B*. Based on this, we might think that the microfinance effect will be lower in *B* than in *A*. To tell us *how much* lower, statistical approaches proceed by measuring the effect conditionally on the age distribution in *A* as a so-called *conditional average treatment effect* (*CATE*). In essence, a *CATE* amounts to a statistical model of how the effect of interest measured *varies over*, and in that sense *depends on*, age. With this model in place, statistical approaches proceed to forming a prediction about *B* by reweighting the effect measured in *A* according to the age distribution in *B*.

In enabling such extrapolations, statistical approaches require a wide range of assumptions (see e.g., Khosrowi 2019; Muller 2013, 2014, 2015), including causal assumptions about *A* and *B*. Among other things, they require that populations only differ concerning how modifying variables are distributed, but not concerning *whether* a modifying variable that plays a role in *A* also influences an effect in *B*, or how, exactly, it does so. For instance, age might not be a modifying variable at all in *B* since individuals have similarly promising business ideas across the age distribution, or age might be involved in *B*, but differently than in *A*—say when high rather than low age coincides with more successful business ideas because older individuals have a better idea of what needs are currently uncatered to by businesses in their community. In adjusting for differences between populations, however, statistical approaches need to rule out such fundamental differences. This implies a host of substantive causal assumptions at the layers of causal analysis outlined earlier (see Khosrowi 2019). These assumptions are rarely explicitly recognized, including by proponents of statistical approaches. One reason for this omission could be that statistical approaches do what it says on the box: they focus on *statistical* models and not on the causal underpinnings of the effects of interest. This can easily become a problem, however, as populations may well often differ at those levels of causal analysis where statistical approaches must assume essential similarities.
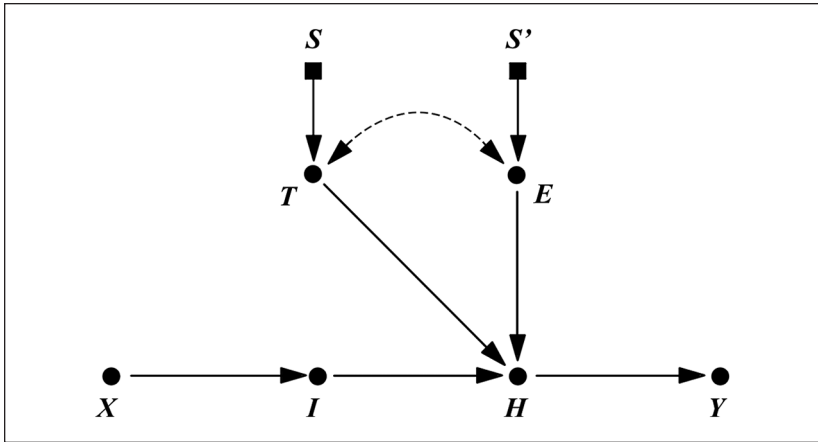
Figure 1. Selection diagram encoding causal assumptions and relevant differences.

By contrast, graph-based approaches (Bareinboim and Pearl 2012, 2016) proposed in the computer science literature offer a richer set of resources to make such assumptions explicit. They involve three main ingredients. The first are structural causal models (SCMs), that is, a set of structural equations that encode the causal relationships assumed to hold between variables. Second, the relationships stipulated by SCMs are captured by corresponding graphical causal models, called directed acyclic graphs (DAGs). In extrapolation, the graph-based approach involves building a special kind of DAG, a so-called *selection diagram*, which represents experimental and target populations in a single graph and uses *selection nodes* to encode differences between populations, such as when the distribution of a variable or the functional relationship between two variables differs. Figure 1 illustrates a selection diagram for our microfinance example.

The main causal pathway proceeds from microfinance (X), through investment in small businesses (I) and household income (H), to household welfare (Y). In addition, there are two further arrows pointing into H: they capture the modifying variables age (T) and entrepreneurial ability (E). I, T, and E interact with each other in producing H. Because T and E are modifying variables (which we know at least for the experimental population), different levels of T and E each induce different marginal effects of a given increment of investment I on household income H: higher age induces smaller effects, and higher entrepreneurial ability induces larger effects of I on H, other things being equal. In addition to this basic structure, the dashed arrow (confounding arc) between T and E indicates that there are

unmeasured confounding variables that induce a correlation between them.[8] For instance, school curricula and social structures might determine at which age individuals acquire pertinent knowledge or relevant experience that bear on their ability to convert investments I into household income H. Finally, to accommodate how A and B might differ, there are two selection nodes S and S' pointing into T and E respectively. They indicate that the distributions of T and E can differ between populations.

In facilitating causal inference, the graph-based approach relies on a third important element: an *inference engine*. Specifically, it offers a powerful formal calculus that helps us derive so-called *transport formulae*, that is, formal expressions that answer a causal query (such as: what is the effect of intervening on X on the distribution of Y in B) in terms of interventional and observational quantities from both populations (e.g., the causal effect measured in A and the distributions of T and E measured in B). I will not dive deeper into how transport formulae are derived (see Bareinboim and Pearl 2012, 2016), including for our example. For our present purposes, it is enough to note that the graph-based approach involves three elements: (1) structural causal models of the populations of interest, (2) corresponding graphical causal models, including a joint causal model of both populations encoding relevant differences (selection diagram), and (3) an inference engine that helps us answer queries about the target based on (1) and (2).

The outline provided above helps us recognize a key difference between statistical and graph-based approaches: the extent to which they make the causal models needed for extrapolation explicit. While statistical approaches model explicitly how an effect depends, statistically, on certain modifying variables, they keep crucial causal assumptions, and hence a substantial part of the models needed for an inference, implicit. The graph-based approach, by contrast, is firmly committed to the idea of making models explicit (as far as existing causal knowledge permits) in the form of an SCM, DAGs, and, in particular, through selection diagrams, that is, joint causal models encoding similarities and differences between populations.

Of course, an explicit model is not automatically a good model, as it may involve assumptions that are not adequately supported, or might indeed be wrong. Yet, making models explicit is an important step in facilitating successful extrapolation as it offers three important benefits.

First, having explicit models makes it easier to assess which assumptions, exactly, an extrapolation involves; whether they appear plausible; whether they enjoy sufficient support; and what one should do to substantiate them.

---

[8]Further confounding arcs are omitted for simplicity.

Consider our microfinance example: in helping us detail the causal structure governing the welfare outcome in both populations, a joint causal model helps us recognize which kinds of similarities are essential for an extrapolation. It makes clear that the ways in which T and E individually meddle with the effect of I on H must be the same in both populations. These assumptions, in turn, require empirical support: for instance, we must somehow get an empirical handle on whether age is indeed a negative modifying variable in B or not. A joint causal model is helpful for estimating T 's individual bearing on the effect of interest in the target because it can instruct us, for instance, to break the correlation between T and E by choosing an appropriate conditioning set.

A second important advantage of making models explicit is that it helps make fine-grained details of the causal and correlational structures of both populations salient to investigators, details that the statistical approach likely fails to recognize. Recall that our microfinance example assumes that, in A, T is a negative modifying variable, E a positive one, and both are positively correlated. If this correlation is strong enough, the statistical approach faces a problem: if a *CATE* is estimated conditionally *only* on T, it can appear as if T is a *positive* modifying variable, whereas, causally speaking, it really is a negative one. Likewise, a *CATE* measured conditionally *only* on E would underestimate E 's positive bearing on the effect of interest. To be sure, there is nothing, in principle, that would stop us from considering both variables when measuring a *CATE* and thereby obtaining the correct expectation of the effect in the target. However, this still requires that we can assume that T and E are involved in meddling with the microfinance effect in the same way in B as in A. The advantage of encoding these assumptions in an explicit causal model is not only that it can help us recognize which assumptions are needed, but also that it helps us extrapolate when these assumptions fail because populations indeed differ. For instance, if the way in which a correlation between T and E is induced is different in B, for example, because younger individuals tend to receive pertinent education that increases their entrepreneurial ability there, then even if we take both T and E into account when adjusting our prediction based on the *CATE* estimated in A, we would still get the effect in B wrong. An explicit causal model, by contrast, helps us distinguish between how variables are involved causally and how they are distributed, rather than mixing both together in an estimate of how an effect depends on these variables in a purely statistical sense.

Finally, a third important advantage of having a well-supported joint causal model of both populations is that such a model can also offer derivational resources for answering a range of further causal queries, such as what co-interventions are needed to help achieve a specific effect. For instance, if

a target differs in age distribution (higher mean) and entrepreneurial ability (lower mean) but is otherwise causally similar, we might consider complementing our microfinance program with a co-intervention that improves individuals' entrepreneurial ability (e.g., a skills training program).

In sum, through making models explicit, the graph-based approach offers crucial resources for turning the models that are invariably involved in extrapolation into explicit and good models, that is, models that (1) adequately capture the causal makeup of both populations, including important differences between them, (2) facilitate the process of providing support for licensing our inferences, and (3) can be used for a variety of further epistemic and practical purposes.

With these ideas in place, let me detail in a more systematic way how we can understand extrapolation as a hybrid inference involving experiments and models.

4. Experiment-Model-Target Inference

In elucidating what successful learning from experiments requires, the existing literature has tended to focus on the kinds of relationships that obtain between experimental and target systems directly. For instance, in the context of making epistemologically pertinent distinctions between experiments and simulations, Morgan (2005) argues that experiments exhibit distinctive epistemic powers in virtue of exploiting *ontological equivalences* between study and target entities of interest. Similarly, Guala (2002, 2005) argues that experiments permit successful inferences about targets to the extent that experimental and target systems are *materially analogous*, for example, because the entities figuring in them are of the same kind, the same material causes operate in both systems, or they exhibit analogies at some other material level of description.

What these proposals have in common is that they focus on a relationship R that obtains directly between two empirical systems and their respective causal makeup, that is, the entities figuring in these systems, their causal features, and specific instances of similarity and difference between these features. Importantly, R is a *concrete* relationship characterized at an *ontic* level: it is 'out there', its nature is up to the world to decide, and it dictates what form causal phenomena or effects take in each population and whether these are similar or not.[9] Importantly, all of this is the case regardless of

---

[9]Of couse, similarity and difference are a matter of judgment and not brute fact. I am interested here in those parts of the causal makeup of A and B that these judgments supervene on.
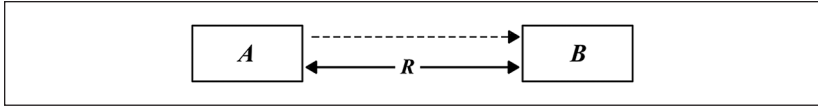
Figure 2. Direct experiment-target inference justified by appeal to R.

whether we are interested in learning about these phenomena/effects or not, or indeed learning about R itself. Figure 2 summarizes this schematically.

Here R is the relationship obtaining between the causal makeup of A and B, and the success of an inference from A to B (dashed arrow) is thought to depend crucially on R.

R is undoubtedly important in many cases where experiments are used to study broader phenomena of interest, especially in natural and life sciences. Particularly in those cases where we are interested in whether an effect/phenomenon established in an experiment will obtain in the same (or a similar) way in a target, R will need to take the right form to make this so (unless we are counting on lucky accidents), and our inferences will hence depend on R. For instance, a lab experiment investigating the reactive behaviors of a chemical substance can be thought to be informative about the behaviors of other instances of the substance outside the lab in virtue of R taking the right form, for example, that the substance involved in the experiment is materially similar/analogous to the substances that constitute the broader target of interest or that the same material causes relevant to its behaviors are present in the target, too. Similarly, in social sciences we are sometimes fortunate enough to study a representative sample from a target population, where arguments concerning the representativeness of the experimental population can be used to ground that R takes a suitable form and hence permits more straightforward inferences.

However, in focusing on R, this schema neglects important features of other, less straightforward kinds of experiment-target inference, such as the kinds of extrapolation discussed here. For one, establishing that R takes a suitable form is often significantly more difficult in extrapolation, such as when populations are distinct rather than the experimental population being sampled from the target, or because appealing to material analogies is not helpful when crucial causal similarities are not guaranteed by entities belonging to the same kind (see Strevens 2007). What is more, extrapolation is often not concerned with merely asking whether an effect will be the same in a target as in a study setting, but aims at drawing a conclusion about a target even when populations might differ relevantly, and by taking such differences into account.

As the preceding discussion made clear, such inferences are unlikely to be successful without explicit models that encode similarities and differences between populations. Because of this, experiment-target inference is not usefully characterized as an inference from an experimental system to a target system *directly*, and underwritten by reference to R. While R still governs whether or not effects will be similar between populations, R itself is not the relationship that matters most in extrapolation. Rather, extrapolation requires that we can construct adequate and well-supported causal models M_A and M_B of experimental and target populations respectively and a joint model M^* (e.g., a selection diagram) that encodes similarities and differences between them. These models, in turn, introduce a second layer to extrapolation: an *epistemic* layer at which our inference proceeds. Here, in addition to R, there is a second relationship of crucial importance: R^*. R^* is an epistemically pertinent (and fallible) abstraction on, and idealization of, the true causal makeup of both populations and their relationship R. Importantly, R^* is acquired by means of representation: the models M_A and M_B and the joint causal model M^* that encodes R^* are abstract and idealized representations of A and B.

Since R^* is needed to make similarities and differences accessible to investigators, the success of extrapolation hinges not only on R (which determines whether/how an effect departs from that in the experiment) but also, and more importantly, on R^*, that is, whether we manage to adequately capture similarities and differences with our joint causal model and whether, by using this information, we manage to correctly predict an effect. Importantly, R^* can encode similarities (and perhaps differences) over and above those 'contained' in R. For instance, individuals in two populations can respond similarly to one and the same intervention but for different underlying reasons: an implicit social norm might govern behaviors in one population and an explicit institution (e.g., a law) might do so in the other. Here, even though there are no concrete similarities in the underlying mechanisms governing behaviors (e.g., in terms of entities and activities), there are properly abstract and formal similarities in terms of behaviors, and these can be captured by R^*, and used for inferential purposes.

In light of the above, it seems that extrapolation is not well understood as an inference from an experiment to a target directly, but rather as a hybrid inference, which is mediated by causal models of both populations and a joint causal model encoding similarities and differences between them, and facilitated by an inference engine operating on the inputs from these models. Call this *Model-Mediated Experiment-Target Inference* (METI).

METI departs significantly from more straightforward experiment-target inference underwritten by reference to R, as Figure 3 illustrates.
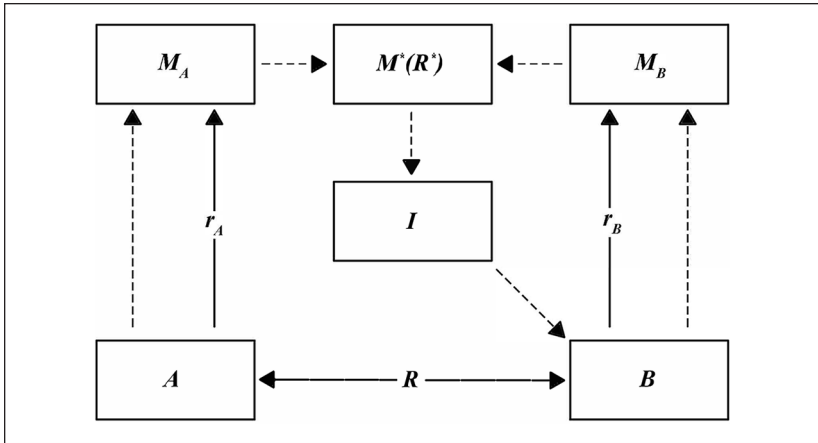
Figure 3. METI as a hybrid, mediated inference from A to B.

Here, we find the populations A and B; the concrete relationship R that obtains between them; the two causal models M_A and M_B, which each stand in a representational relationship r_A and r_B to A and B respectively; and a joint causal model M^* which encodes the abstract relationship R^*. The 'flow' of inference is depicted in dashed arrows. It proceeds 'upwards' from both populations toward the models M_A and M_B representing them, through the joint model M^*, and it is mediated by an inference engine I in forming a prediction about B. METI hence carves out the role of models in more detail: models sit, inferentially, between an experiment and a conclusion, and, just like the experimental result from A, they form an essential part of the resources employed by an inference.

Importantly, METI characterizes the resources involved in extrapolation as an epistemic whole whose ingredients interact in ways that have not been recognized by existing accounts. An experimental result, by itself, might inspire hypotheses about what an effect in a target could be, but unless it is joined by a causal model that helps clarify relevant similarities and differences, and how these bear on the effect to be expected in a target, it remains largely uninformative about that effect. Likewise, even a sophisticated joint causal model, by itself, is unlikely to permit conclusions about an effect in the target with the same confidence as when it is complemented by a well-identified causal effect obtained from an experiment. It can tell us, for instance, that *whatever* an effect is in A, it will be similar in B (or depart systematically according to certain differences), but it cannot tell us what, exactly, that

effect will be in a target unless we have confidently established the magnitude of the effect, for example, by means of an experiment. So, experiments and models each play important roles in METI, with neither being privileged, and together enable conclusions that would not be feasible with the same confidence from either ingredient alone.

Beyond elucidating the role of models in experiment-target inference, METI can also be useful for critically appraising such practices. As emphasized earlier, a key methodological shortcoming of EBP's methodological guidelines is that they say little on extrapolation and thereby tolerate that models often remain implicit. METI gives us some normative leverage for refining these guidelines, for example, to demand that models be explicit, amenable to scrutiny, and well-supported.

It is also important, however, to emphasize that METI is limited in its ameliorative scope: it tells us *that* good models are needed for successful inference, but not how to build such models. There are indeed important challenges to be overcome in building the kinds of explicit and well-supported models that can facilitate extrapolation. One is simply that constructing such models is rarely straightforward and requires substantial epistemic resources, which is perhaps why even proponents of model-heavy approaches, such as Bareinboim and Pearl, say little on how causal models are constructed, and simply assume that an investigator is in possession of sufficient empirical knowledge to build them (see Hyttinen et al. 2015). But we are not completely empty-handed when it comes to positive proposals: advocates of realist evaluation (Astbury and Leeuw 2010; Pawson 2013) as well as philosophers and methodologists (Cartwright 2020; Cartwright and Hardie 2012), have made important progress in furnishing details for how to construct so-called *program theories* or *logic models* that systematically and iteratively refine our understanding of the causal mechanisms by which policy interventions are effective and the conditions under which these interventions are likely to work best. These proposals constitute promising avenues for catering to METI's demand for good models.

There is also a second, more principled challenge to be acknowledged when emphasizing the role of models in experiment-target inference: the *extrapolator's circle* (Steel 2009; see also Khosrowi 2019). In a nutshell, the extrapolator's circle cautions that the causal knowledge about a target required for an extrapolation must not be so comprehensive that we could answer our causal query on the basis of that knowledge alone. So maintaining that extrapolation requires explicit and well-supported causal models, as METI does, must stop well short of demanding models strong enough to obviate the need to extrapolate at all, as that would be self-effacing.

Yet, while the extrapolator's circle is indeed an important challenge, METI is not *especially* vulnerable to it, nor does it remain mute on the issues to be addressed in overcoming it. For one, it is important to recognize that the extrapolator's circle is a challenge for *any* sincere attempt at extrapolation, with or without the help of models. Moreover, it seems that aiming to extrapolate by means of explicit causal models can indeed be helpful for evading the extrapolator's circle. A statistical approach to extrapolation, for instance, will often need to support similarity assumptions *en-bloc*. For lack of encoding finer-grained details about the causal makeup of two populations, it requires that they exhibit similarities wholesale at certain levels of causal analysis. This is problematic because such wide-ranging assumptions generally require more extensive support, for example, about mechanistic similarities from start to finish, rather than at particular stages.

By contrast, following METI's call to make models explicit can help us spell out in a more fine-grained way which assumptions are needed, which of these we can safely bet on and which ones are in need of further support, and whether providing such support would demand too much causal knowledge about a target. This is perhaps why existing proposals for overcoming the extrapolator's circle (Steel 2009, 2010) already take a model-based approach toward delineating what assumptions need to be made and elaborating which of these we should focus on to avoid trivializing our inferences.

So while METI emphasizes that good models are needed to facilitate extrapolation, it does not claim that models are a silver bullet or come for free. It acknowledges that there remain substantial challenges in the way of constructing models that are powerful enough to facilitate extrapolation while not requiring extensive causal knowledge about a target. And it also emphasizes that explicit models can play important roles in overcoming these challenges, for example, by helping us recognize which inferences are prone to raise concerns about the extrapolator's circle in the first place.

With these clarifications made, let me briefly touch upon two further issues to complete my detailing of METI: how it relates to existing accounts of experiment-target relationships, and whether its scope extends beyond extrapolation.

First, METI can help us improve on existing views of experiment-target relationships such as those offered by Morgan (2003, 2005) and Guala (2002, 2005). These views have so far not fully distinguished between ontic and epistemic layers of experiment-target inference. Focusing on the ontic layer of concrete experiment-target relationships R, like Morgan and Guala do, is helpful for elucidating more straightforward cases of experiment-target inference. Yet, extrapolation hinges not only on R, but also, and more importantly, on whether we successfully manage to represent R's relevant features

as well as further, abstract similarities through R^* at an epistemic level, and whether we can successfully use these representations to answer causal queries about the world. This hence accommodates some of the concerns expressed by Parker (2009) and Parke (2014), who have resisted Morgan's and Guala's emphasis on distinctive epistemic benefits yielded by concrete relationships obtaining directly between experimental and target populations, such as ontological equivalence and material analogy, by detailing how abstract, formal relationships of relevant similarity are often doing the important work of facilitating inferences from experiments. METI can help further reconcile these views. It acknowledges that R plays an important role in determining causal phenomena/effects in both populations and whether they are similar or not. But it also accommodates Parker and Parke's emphasis on abstract relationships of similarity (and difference) by clarifying the role of R^* and highlighting that extrapolation proceeds through thick layers of representation, abstraction, and idealization, before causal information is available to investigators at an epistemic level. So in METI, the success of extrapolation can hinge, at the same time, on concrete relationships obtaining directly between populations and on abstract relationships obtaining only at the epistemic level where inference proceeds.

What about the scope of METI? I have proposed METI as a characterization of extrapolative inference from experiments in EBP and social science more generally. But we might ask whether METI can also be helpful in elucidating other kinds of experiment-target inference. Here, it seems that much depends on the kind of inference at issue. We might think, for instance, that even the most straightforward cases of experiment-target inference, for example, those facilitated by justificatory strategies involving sampling-type or material analogy-type arguments, could be understood to involve models, too (at least implicitly). But it is questionable whether making these models explicit could provide epistemic benefits, so not much seems to be gained by thinking about experiment-target inference in terms of METI. Between these easier cases and more involved ones, like those encountered in extrapolation, there is however a whole spectrum of inferences that differ in various important respects, including in (1) the aims they pursue (e.g., inferring whether phenomena/effects are the same in a target vs. predicting precise quantities), (2) the severity of the challenges faced (in what ways populations/systems might differ and to what extent, and what resources are available to clarify this), (3) the availability of justificatory strategies that appeal directly to R (e.g., representative sampling-type or material analogy-type justifications), etc. At the most general level, it seems that METI will be more appealing as a way to understand these inferences the more ambitious the aims are, the more severe the challenges involved, and the less any justificatory shortcuts

are available. Delineating the extension of this class of inferences is beyond the scope of this paper, but it is hoped that future work can help not only refine METI but also make progress in exploring where it can be usefully applied.

   With this overview of METI in place, let us briefly consider an important objection against METI's core claim, that is, that some kinds of experiment-target inference need *models*.

5. Models proper?

While it is clear that good inferences from experiments need plenty of additional resources, including causal assumptions enjoying adequate support, one might nevertheless have reservations about whether this is enough to conclude that what is needed are models proper. The assumptions involved in extrapolation might be understood as causal models, but perhaps these causal models should not be counted as *scientific* models in the general sense, and no broader conclusions about interactions between models and experiments should be drawn beyond those specifically pertaining to causal models in the context of extrapolation.

   Let me defend here in more detail why the models involved in extrapolation, at least if built conscientiously, should be considered capital-M scientific models in the most general sense. This is because they exhibit a range of essential characteristics that figure prominently in existing accounts of scientific models:

(1) They are *artifacts* that are constructed and used for pursuing particular epistemic purposes, that is, extrapolating causal effects, and they figure as *surrogates* for real-world systems (cf. Mäki 2009), that is, they allow us to answer causal queries about a target without studying that target directly.

(2) They are formal/mathematical and/or diagrammatical *representations* of the causal makeup of real-world systems. In virtue of this, they raise standard issues of scientific representation (Frigg and Nguyen 2020; Giere 2010; Khosrowi 2020; Suárez 2003; Teller 2001; van Fraassen 2008; Weisberg 2013), that is, to promote certain epistemic purposes, they need to be suitably related (e.g., sufficiently similar, structurally isomorphic, etc.) to their targets in epistemically relevant respects.

(3) They involve *abstractions.* In order to remain analytically tractable, causal models omit inessential details, such as when an extraneous causal relationship understood to be inessential for an effect is

omitted from a model, or when a relationship between two variables is represented as analytically primitive even though we believe it is underwritten by a finer-grained causal structure in the real world.

(4) They involve *idealizations*. For instance, a parametric structural equation model might assume linear functional forms throughout if it is understood that this is a reasonable approximation for the epistemic purpose at hand.

(5) Context matters. Fidelity desiderata and contextual factors co-determine whether a particular model is adequate for a purpose (Parker 2020). For instance, we would expect a model to resolve greater causal detail if we are interested in precise quantitative predictions rather than qualitative predictions. Similarly, the stakes involved in acting on a prediction, and other contextual factors related to epistemic risk (Biddle and Kukla 2017), may bear on the fidelity desiderata imposed on both model construction and justification.

(6) To be useful, the models at work in extrapolation need to be *empirically informed* with measurements of certain variables and structural parameters relevant to an effect, much like structural models in economics and econometrics need to be populated with measurements and calibrated before being put to use successfully (see e.g., Boumans 2002).

(7) They are *autonomous epistemic instruments* (Morrison and Morgan 1999) that can be used for epistemic and practical purposes beyond extrapolation, for example, by facilitating understanding of causal phenomena (see e.g., de Donato Rodríguez and Zamora Bonilla 2009), such as how effects vary between populations and individuals; by guiding efforts to unpack finer-grained details of a given causal structure; and, as argued earlier, by inspiring novel interventions, such as when helping us realize that a co-intervention might be useful to achieve a certain desired effect.

In sum, the above characteristics suggest that we should take the causal models involved in extrapolation seriously as capital-M scientific models proper.

6. Conclusions

I have argued that an important class of experiment-target inference is not well understood as proceeding from an experimental system to a target directly, but rather as a hybrid inference, METI, which involves experiments and models. I have focused on the practice of extrapolation in EBP

and argued that extrapolation requires models for two reasons: first, because we need to make at least some implicit causal assumptions, and second, because good, explicit models play essential roles in facilitating successful extrapolation: they guide the search for relevant similarities and differences, encode such information at an epistemic level, and help with deriving predictions about the effects of ultimate interest. Recognizing the importance of models has three important benefits: (1) it brings to the fore epistemically significant but heretofore unrecognized interactions between experiments and models as part of epistemic wholes enabling conclusions that would remain inaccessible otherwise, (2) it helps critically appraise inferential practices involving experiments and offers normative leverage to ask that the models involved be made explicit, and (3) it reconciles some differences among existing views on experiment-target relationships by clarifying that both concrete and abstract relationships are important for successful inference.

In light of these benefits, it seems interesting to explore whether METI can also be informative for understanding and critically appraising other kinds of experiment-target inference beyond extrapolation, an issue that has not been explored here. It is hoped, however, that the first sketch of METI developed here can figure as an interesting starting point to further advance the larger philosophical project of elucidating how experiments and models, on their own, or when joined together, facilitate learning about the world.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Astbury, Brad, and Frand Leeuw. 2010. "Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation." *American Journal of Evaluation* 31 (3): 363-81.

Banerjee, Abhijit V, and Esther Duflo. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics* 1 (1): 151-78.

Bareinboim, Elias, and Judea Pearl. 2012. "Transportability of Causal Effects: Completeness Results." Proceedings of the Twenty-Sixth Conference on Artificial Intelligence (AAAI-12), Menlo Park, CA.

Bareinboim, Elias, and Judea Pearl. 2016. "Causal Inference and the Data-Fusion Problem." *Proceedings of the National Academy of Sciences of the United States of America* 113: 7345-52.

Baron, Ruben M, and Kenny, David A. 1986. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic and Statistical Considerations". *Journal of Personality and Social Psychology* 51: 1173-82.

Biddle, Justin, and Quill R. Kukla. 2017. "The Geography of Epistemic Risk." In *Exploring Inductive Risk: Case Studies of Values in Science*, edited by Elliott, K., and T. Richards, 215-37. Oxford: Oxford University Press.

Boumans, Marcel. 2002. "Calibration of models in experiments". In *Model-Based Reasoning: Science, Technology, Values*, edited by L. Magnani, and N. Nersessian, 75-93. Boston: Springer.

Cartwright, Nancy. D. 2012. "Presidential Address: Will This Policy Work for You? Predicting Effectiveness Better: How Philosophy Helps." *Philosophy of Science* 79 (5): 973-89.

Cartwright, Nancy. D. 2013a. "Evidence, Argument and Prediction." In *EPSA11 Perspectives and Foundational Problems in Philosophy of Science*, The European Philosophy of Science Association Proceedings, edited by V. Karakostas, and D. Dieks, 3-17. Cham: Springer International Publishing Switzerland.

Cartwright, Nancy. D. 2013b. "Knowing What We Are Talking About: Why Evidence Doesn't Always Travel." *Evidence and Policy: A Journal of Research, Debate and Practice* 9 (1): 97-112.

Cartwright, Nancy. D. 2020. "Middle-range Theory - (Teoría Del Rango Medio): Without It What Could Anyone Do?" *Theoria: An International Journal for Theory, History and Foundations of Science* 35 (3): 269-323.

Cartwright, Nancy. D., and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing it Better*. Oxford: Oxford University Press.

Currie, Adrian, and Arnon Levy. 2019. "Why Experiments Matter." *Inquiry* 62: 1066-90.

Crump, Richard K., Joseph V. Hotz, Guido W. Imbens, and Oscar A. Mitnik, 2008. "Nonparametric tests for treatment effect heterogeneity." *The Review of Economics and Statistics* 90 (3): 389-405.

Da Costa, Newton C. A., and Steven French. 2003. *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*. Oxford: Oxford University Press.

de Donato Rodríguez, X., and J. Zamora Bonilla. 2009. "Credibility, Idealisation, and Model Building: An Inferential Approach." *Erkenntnis* 70: 101–18.

Duflo, Esther. 2004. "Scaling Up and Evaluation." In *Annual World Bank Conference on Development Economics, 2004: Accelerating Development*, edited by F. Bourguignon and B. Pleskovic, 341-69. Washington, D.C.: World Bank; Oxford and New York: Oxford University Press.

Frigg, Roman, and James Ngyuen. 2020. *Modelling Nature: An Opinionated Introduction to Scientific Representation. Synthese Library - Studies in Epistemology, Logic, Methodology, and Philosophy of Science*. Cham: Springer.

Fuller, John. 2019. "The Confounding Question of Confounding Causes in Randomized Trials." *The British Journal for the Philosophy of Science* 70 (3): 901–26.

Giere, Ronald N. 2010. "An Agent-Based Conception of Models and Scientific Representation." *Synthese* 172: 269-81.

Guala, Francesco. 2002. "Models, Simulations, and Experiments." In *Model-Based Reasoning: Science, Technology, Values*, edited by L. Magnani and N. Nersessian, 59-74. New York: Kluwer.

Guala, Francesco 2005. *The Methodology of Experimental Economics*. Cambridge: Cambridge University Press.

Guala, Francesco 2010. "Extrapolation, Analogy, and Comparative Process Tracing." *Philosophy of Science* 77 (5): 1070-82.

Guyatt, Gordon H., Andrew D. Oxman, Regina Kunz, James Woodcock, Jan Brozek, Mark Helfand, Pablo Alonso-Ciello, et al. 2011. "Grade Guidelines: 8. Rating the Quality of Evidence - Indirectness." *Journal of Clinical Epidemiology* 64 (12): 1301-10.

Hotz, V. Joseph, Guido W. Imbens, and Julie H. Mortimer. 2005. "Predicting the Efficacy of Future Training Programs Using Past Experiences at Other Locations." *Journal of Econometrics* 125: 241-70.

Hyttinen, Antti, Frederick Eberhardt, and Matti Järvisalo. 2015. "Do-calculus when the true graph is unknown." In *UAI'15 Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, edited by M. Meila, and T. Heskes, 395-404. Arlington (VA): AUAI Press.

Khosrowi, Donal. 2019. "Extrapolation of Causal Effects—Hopes, Assumptions, and the Extrapolator's Circle." *Journal of Economic Methodology* 26 (1): 45-58.

Khosrowi, Donal. 2020. "Getting Serious about Shared Features." *The British Journal for the Philosophy of Science* 71 (2): 523–46.

McLean, Kate A., Aisha Byanaku, Augustine Kubikonse, Vincent Tshowe, Said Katensi, and Amy G. Lehman. 2014. "Fishing with Bed Nets on Lake Tanganyika: A Randomized Survey." *Malaria Journal* 13: 395.

Mäki, Uskali. 2009. "Missing the World. Models as Isolations and Credible Surrogate Systems." *Erkenntnis* 70: 29-43.

Mäki, Uskali. 2005. "Models are Experiments, Experiments are Models." *Journal of Economic Methodology* 12 (2): 303-15.

Morgan, Mary S. 2003. "Experiments Without Material Intervention: Model Experiments, Virtual Experiments and Virtually Experiments". In *The Philosophy of Scientific Experimentation*, edited by H. Radder, 216-35. Pittsburgh: University of Pittsburgh Press.

Morgan, Mary S. 2005. "Experiments Versus Models: New Phenomena, Inference, and Surprise." *Journal of Economic Methodology* 12 (2): 317-29.

Morrison, Margaret., and Mary S. Morgan. 1999. "Models as mediating instruments". In *Models as Mediators*, edited by M.S. Morgan and M. Morrison, 10-37. Cambridge: Cambridge University Press.

Muller, Seán M. 2013. "External Validity, Causal Interaction and Randomised Trials: The Case of Economics", Unpublished manuscript.

Muller, Seán M. 2014. "Randomised Trials for Policy: A Review of the External Validity of Treatment Effects." Southern Africa Labour and Development Research Unit Working Paper 127, University of Cape Town.

Muller, Seán M. 2015. "Interaction and External Validity: Obstacles to the Policy Relevance of Randomized Evaluations." *World Bank Economic Review* 29(1): S217-25.

Parke, Emily C. 2014. "Experiments, Simulations, and Epistemic Privilege". *Philosophy of Science* 81 (4): 516–36.

Parker, Wendy. 2009. "Does Matter Really Matter? Computer Simulations, Experiments, and Materiality". *Synthese* 169 (3): 483-96.

Parker, Wendy. 2020. "Model Evaluation: An Adequacy-for-Purpose View". *Philosophy of Science* 87 (3): 457-77.

Parkhurst, Justin. 2017. *The Politics of Evidence: From Evidence-Based Policy to the Good Governance of Evidence*. Abingdon, Oxon: Routledge.

Pawson, Ray. 2013. *The Science of Evaluation: A Realist Manifesto*. London: SAGE Publications.

Peters, Jörg, Monica Jain, and Marie Gaarder. 2019. "External Validity: Policy Demand is There but Research Needs to Boost Supply". 3ie blog post. Retrieved June 2020 from: https://www.3ieimpact.org/blogs/external-validity-policy-demand-there-research-needs-boost-supply

Reiss, Julian. 2019. "Against External Validity." *Synthese* 196 (8): 3103-21.

Schünemann, Holger J., Julian P. T. Higgins, Gunn E. Vist, Paul Glasziou, Elie A. Akl., Nicole Skoetz, and Gordon H. Guyatt. 2019. "Chapter 14: Completing 'summary of findings' tables and grading the certainty of the evidence." In *Cochrane Handbook for Systematic Reviews of Interventions. 2nd Edition*, edited by J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, et al., 375-99. Chichester (UK): John Wiley & Sons.

Steel, Daniel. 2009. *Across the Boundaries: Extrapolation in Biology and Social Science*. Oxford: Oxford University Press.

Steel, Daniel. 2010. "A New Approach to Argument by Analogy: Extrapolation and Chain Graphs." *Philosophy of Science* 77 (5): 1058-69.

Strevens, Michael. 2007. "Why Represent Causal Relations?" In *Causal Learning: Psychology, Philosophy, Computation*, edited by A. Gopnik and L. Schulz, 245-60. New York: Oxford University Press.

Suárez, Mauricio. 2003. "Scientific Representation: Against Similarity and Isomorphism." *International Studies in the Philosophy of Science* 17: 225-44.

Teller, Paul. 2001. "Twilight of the Perfect Model Model." *Erkenntnis* 55:393-415.

van Eersel, Gerdien G., Gabriela V. Koppenol-Gonzalez, and Julian Reiss. 2019. "Extrapolation of Experimental Results through Analogical Reasoning from Latent Classes." *Philosophy of Science* 86(2): 219-35.

Van Fraassen, Bas. 2008. *Scientific Representation: Paradoxes of Perspective*. Oxford: Oxford University Press.

Vivalt, Eva. 2020. "How Much Can We Generalize from Impact Evaluations?" *Journal of the European Economics Association* 18 (6): 3045-89.

Weisberg, Michael. 2013. *Simulation and Similarity*. Oxford: Oxford University Press.

Winsberg, Eric. 2009. "A Tale of Two Methods". *Synthese* 169 (3): 575-92.

Author Biography

Donal Khosrowi is a postdoctoral researcher at Leibniz University Hannover, Germany. His research interests and recent publications focus on causal inference in social science (especially in economics and Evidence-Based Policy), scientific representation with models, and values in science issues, especially surrounding the production and use of evidence for policy.