



Assessing the role of evidence of mechanisms in causal extrapolation

(Evaluando el rol de la evidencia de mecanismos en la extrapolación causal)

Saúl PÉREZ-GONZÁLEZ*, Valeriano IRANZO

University of Valencia

ABSTRACT: Extrapolation of causal claims from study populations to other populations of interest is a problematic issue. The standard approach in experimental research, which prioritises randomized controlled trials and statistical evidence, is not devoid of difficulties. Granted that, it has been defended that evidence of mechanisms is indispensable for causal extrapolation. We argue, contrarily, that this sort of evidence is not indispensable. Nonetheless, we also think that occasionally it may be helpful. In order to clarify its relevance, we introduce a distinction between a positive and a negative role of evidence of mechanisms. Our conclusion is that the former is highly questionable, but the latter may be a trustworthy resource for causal extrapolation.

KEYWORDS: Extrapolation; evidence of mechanisms; statistical evidence; mechanism; causality; evidence; external validity; randomized controlled trial.

RESUMEN: *La extrapolación de relaciones causales de poblaciones de estudio a otras poblaciones de interés es una cuestión problemática. El procedimiento estándar en investigación experimental, el cual prioriza los ensayos controlados aleatorizados y la evidencia estadística, no está carente de dificultades. Dada esta situación, se ha planteado que la evidencia de mecanismos es indispensable para la extrapolación causal. Nosotros argumentamos que, por el contrario, este tipo de evidencia no es indispensable. Sin embargo, pensamos que puede ser de ayuda en ciertas ocasiones. Para clarificar su relevancia, distinguimos entre el rol positivo y el rol negativo de la evidencia de mecanismos. Nuestra conclusión es que el primero es altamente cuestionable, pero el segundo puede ser un recurso fiable para la extrapolación.*

PALABRAS CLAVE: *Extrapolación; evidencia de mecanismos; evidencia estadística; mecanismo; causalidad; evidencia; validez externa; ensayo controlado aleatorizado.*

* **Correspondence to:** Saúl Pérez-González. Department of Philosophy, University of Valencia. Avd. Blasco Ibáñez, 30 (46010 Valencia, Spain) – saul.perez@uv.es – https://orcid.org/0000-0001-9594-9558

How to cite: Pérez-González, Saúl; Iranzo, Valeriano (2021). «Assessing the role of evidence of mechanisms in causal extrapolation»; *Theoria. An International Journal for Theory, History and Foundations of Science*, 36(2), 211-228. (https://doi.org/10.1387/theoria.21642).

Received: 2020-04-09; Final version: 2020-12-07.

ISSN 0495-4548 - eISSN 2171-679X / © 2021 UPV/EHU



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

1. Introduction

The relevance of mechanisms in science has been advocated by many scientists and philosophers of science (see Glennan, 2017; Glennan & Illari, 2018). This relevance is associated with several issues such as scientific explanation (Machamer *et al.*, 2000; Bechtel & Abrahamsen, 2005; Hedström, 2005; Glennan, 2017), scientific inquiry (Bechtel & Richardson, 1993; Darden, 2018), knowledge growth and organization (Hedström & Ylikoski, 2010; Glennan, 2017), integration of different fields (Craver & Alexandrova, 2008), and establishment of causal claims (Steel, 2004; Russo & Williamson, 2007; Glennan, 2017). Over the last decade, the role of evidence of mechanisms in extrapolating causal claims from study populations to other populations of interest has also been a matter of contention (Steel, 2008; Clarke *et al.*, 2013, 2014; Howick *et al.*, 2013a, 2013b; Parkkinen *et al.*, 2018; Marchionni & Reijula, 2019; Anjum *et al.*, 2020).

The aim of this paper is to analyse the role of evidence of mechanisms in extrapolation and to evaluate whether, as some authors have defended, it is indispensable for legitimately extrapolating a causal claim. We will argue that, although it is not indispensable, evidence of mechanisms can be a valuable resource for causal extrapolation. The structure of the paper is as follows. Section 2 examines and discusses the “indispensability thesis” about evidence of mechanisms. Firstly, the main shortcomings confronted by the standard statistical approach (i.e. the absence of universal responses and the context-variance) are introduced (subsection 2.1). Subsequently, after characterising evidence of mechanisms, it is claimed that it is not indispensable for addressing those difficulties. Evidence of mechanisms is not required for dealing with the heterogeneity of responses (subsection 2.2) and is also beset by contextual variability (subsection 2.3). After establishing the non-indispensability of evidence of mechanisms, section 3 discusses its appropriate role in causal extrapolation. For that purpose, a distinction between a positive and a negative role of evidence of mechanisms is introduced. It is argued that the negative role of evidence of mechanisms is reliable and valuable for causal extrapolation (subsection 3.1), while the actual relevance of the positive role is, at the very least, uncertain (subsection 3.2).

2. Statistical evidence and evidence of mechanisms

Extrapolation of causal claims is usually linked to “*external validity*” in literature on scientific methodology. It is contrasted with internal validity. The latter “refers to the approximate validity with which we infer that a relationship between two variables is causal or that the absence of a relationship implies the absence of cause”, while external validity “refers to the approximate validity with which we can infer that the presumed causal relationship can be generalized to and across alternate measures of the cause and effect and across different types of persons, settings, and times” (Cook & Campbell, 1979, p. 37).¹

Concerning extrapolation, then, we should distinguish the study and the target populations. They are, in fact, *different* populations and extrapolations from the former to the

¹ Although this distinction is deeply rooted in the methodological discussion, some critical voices are also present. See, for instance, Jiménez-Buedo & Miller (2010) and Reiss (2019).

latter are justified insofar as they resemble each other (or alternatively, insofar as they do not differ significantly). The fundamental question raised by causal extrapolation is, then, how we know that those populations are similar in the relevant respects (i.e. those factors which could have a non-negligible effect on the results obtained in the study population).

2.1. EXTRAPOLATION BASED ON STATISTICAL EVIDENCE

The standard statistical approach invokes randomized control trials (RCTs) for discerning causal relations in the study population. Both randomization and comparison between two subpopulations (the control group and the experimental group) are the strategies favoured by RCTs.

$$\text{Study population} \left\{ \begin{array}{l} \text{Experimental group } (P_1, P_2, P_3 \dots P_n) \\ \text{Control group } (\neg P_1, P_2, P_3 \dots P_n) \end{array} \right.$$

Experimental design according to RCTs looks for subpopulations which differ only in the alleged causal factor (P_1 , in this case). Random assignment justifies neutralization of those variables (P_i) apart from P_1 which could also be causally relevant for the dependent variable Q —i.e., the investigated effect. The necessary assumption here is that the statistical distribution of those variables which could affect Q apart from P_1 is the same (or nearly the same) in both subpopulations. Otherwise, significant differences discovered on Q -frequencies could be considered a consequence not just of switching $\neg P_1$ to P_1 in the control and the experimental group respectively, but a consequence of any of those other —be it known or unknown— underlying variables which could have an influence on Q . Then, after intervention, that is, after switching $\neg P_1$ to P_1 , differences in Q -frequencies obtained in both subpopulations should be considered as a consequence *exclusively* of P_1 . It should be added that different experiments —with different study populations— occasionally give similar results. After all, systematic reviews which encompass several RCTs, also labelled as “meta-analyses”, are the highest ranked sort of evidence in most hierarchies of evidence.²

Sometimes it is not possible to resort to RCTs due to practical or ethical reasons. Then, “quasi-experimental” designs (prospective cohort-studies, retrospective case-studies, etc.), in which randomization is absent and study populations are “real”, are used. Even for scientific fields which ubiquitously lean on statistical evidence that may be the usual predicament. Epidemiology is a good example. By and large we cannot intervene to induce dis-

² The GRADE (*Grading of Recommendations Assessment, Development and Evaluation*) approach in medicine and public health gives this recommendation: “Ideally, evidence profiles should be used to assess the certainty in the evidence and these should be based on systematic reviews” (GRADE working group, 2016). The European Parliament Directive 2001/83/EC on the Community code relating to medicinal products for human use is very clear in favouring RCTs: “In general, clinical trials shall be done as ‘controlled clinical trials’ and if possible, randomized; any other design shall be justified.” (Annex I, part 4). And there are more examples: for instance, those of the UK National Institute for Health and Care Excellence (2006), the CONSORT (*Consolidated Standards of Reporting Trials*) Group (2010), the Oxford Centre for Evidence-based Medicine (2011), and the Scottish Intercollegiate Guidelines Network (2019).

eases on humans so we must rely on quasi-experimental designs. RCTs on animals could overcome these ethical concerns, but then the issue is precisely extrapolating those results to humans. Nevertheless, since our target here is not detection (of causal relationships), but extrapolation, we will assume that a fairly good control on confounding variables may often be obtained and, consequently, that we can get conclusive reasons in favour of causal relationships by means of those quasi-experimental designs.³

Now the question is to what extent those results based on statistical frequencies could be reliably extrapolated to a different population (the target population). The short answer is that they can be extrapolated if the study and the target population are similar enough. From this point of view, the question is the same no matter if we depart from statistical evidence or evidence of mechanisms, that is, to justify the similarity assumption. The specific problem, however, is whether the statistical information about the study population overlooks what could be relevant differences between it and the target population so that the extrapolation is unreliable.

It is worth adding that we are not concerned here with deficiencies found in particular investigations due to practical constraints (funding, time, suitable training for human resources...). Rather, we should try to discern intrinsic limitations for statistical evidence about extrapolation which do not depend on these practical concerns. Those limitations usually invoked are related, firstly, to the assumption that there is a “universal response”, and secondly, to the differences between the respective contexts of both populations.

Extrapolating from RCTs (or quasi-experimental designs) assumes “universal response”, that is, that different individuals (or groups of individuals) will respond to an intervention, treatment, or drug in the same way. It is claimed that that assumption may be justified for “interventions with short causal pathways” at most, but it is highly debatable about “interventions involving long, complex causal pathways, or in large-scale evaluations where these pathways can be affected by numerous characteristics of the population, health system, or environment” (Victoria *et al.*, 2004, p. 402). An example is that of an antihypertensive treatment sensitive to the ethnicity of the patient. RCTs were developed in different ethnic groups and the design was motivated because of “plentiful evidence suggesting the operation of different pro-hypertensive mechanisms operating in different ethnic groups” (Clarke *et al.*, 2014, p. 347).

Furthermore, the study population is very frequently analysed in artificial or unrealistic conditions while the target population is “in the real world”. Then, to what extent can we expect that those results obtained in artificial conditions will be preserved in real populations? The fact that humans know that they are under experimental treatment could have distorting effects on the results. More importantly, extrapolation from *in vitro* trials to *in vivo* conditions about the effectiveness of a treatment demands mathematical models (IVIVC = *in vitro/in vivo* correlations) to relate *in vitro* effects to *in vivo* responses (Bell *et al.*, 2018). Concerning health policies, effective implementation may be troublesome. For instance, doses are precisely prescribed in experimental conditions (quantity and times), but the real situations to apply the treatment to patients are not so well-defined or

³ We will not distinguish here between “detecting a cause-effect relationship” and “assessing the effect size”. However, see Kincaid (2012) for some limitations of multiple regression analysis concerning the latter.

so controlled as those in the laboratory. The fact is that this sort of context-variance may put at risk extrapolations based on those correlations found in an experimental setting.⁴

Given this scenario, some scientists and philosophers of science consider that study-to-target causal inferences based on statistical evidence alone are not reliable and, consequently, that a different sort of evidence (i.e. evidence of mechanisms) should be taken into account to ground them.

2.2. THE “INDISPENSABILITY THESIS” ABOUT EVIDENCE OF MECHANISMS

The “indispensability thesis” about evidence of mechanisms claims that evidence of mechanisms is required in order to make a reliable causal inference (Russo & Williamson, 2007; Clarke *et al.*, 2013, 2014; Parkkinen *et al.*, 2018). Evidence of mechanisms is defined as evidence about the existence or the properties of mechanisms in the domain of inquiry (Illari, 2011). In that approach, a broad view of mechanisms is taken (Aronson *et al.*, 2018; Parkkinen *et al.*, 2018). Although a mechanism must be a real entity in the world, it can be a complex system (Illari & Williamson, 2012), a causal process (Salmon, 1998), or some combination of both. Moreover, there is no specific kind of evidence-gathering method through which evidence of mechanisms must be obtained. A wide variety of methods are considered legitimate, including case reports, autopsies, cohort studies, simulations, *in vitro* experiments, and RCTs.⁵

The underlying rationale of indispensability is that statistical evidence and evidence of mechanisms must be integrated in order to obtain reliable causal inferences. It is maintained that statistical evidence and evidence of mechanisms complement each other. Statistical evidence has some intrinsic limitations that require evidence of mechanisms to be overcome, and vice versa. Advocates of that approach claim that, in order to establish that *A* is a cause of *B* in a study population, it must be ascertained that they are correlated and also that there is an underlying mechanism that explains that correlation (Russo & Williamson, 2007; Clarke *et al.*, 2014). According to them, statistical evidence may successfully assess the net effect of the mechanism linking *A* and *B*, but it can hardly deal with shortcomings related to confounding factors and non-causal correlations. On the other hand, evidence of mechanisms is appropriate to cope with the latter difficulties, but it cannot assess the net effect of the involved mechanism. Furthermore, it is claimed that the demand of both statistical evidence and evidence of mechanisms is supported by medical methodology, uses of causality, and several instances of causal discovery (Clarke *et al.*, 2014).

Regarding causal extrapolation, advocates of the indispensability thesis maintain that, in order to extrapolate a causal claim from a study population to a different target population, one needs to establish similarity of relevant mechanisms in both populations (Clarke *et al.*, 2014; Parkkinen *et al.*, 2018; Williamson, 2019). Those authors claim that establishing similarity of relevant mechanisms in the study and target populations is analogous to

⁴ Regarding the social sciences (especially evidence-based policy), Cartwright (2012) has argued that, even when the study population is “in the real world”, differences in the context (e.g. cultural differences) between the study and the target population may undermine the extrapolation.

⁵ It should be noted that the distinction between statistical evidence and evidence of mechanisms is a distinction about objects of evidence, not about evidence-gathering methods. In fact, as Illari has argued, “there is no principled distinction between the *kinds* of empirical work by which we get evidence of mechanisms, and evidence of difference-making” (2011, p. 145).

establishing the existence of both a correlation and an underlying mechanism that explains it in the target population. They consider that “[b]y means of this similarity of mechanisms, one can use the claim that A is a cause of B established in the source population to further support the correlation claim in the target population” (Williamson, 2019, p. 53). In order to confirm the similarity of the relevant mechanisms, it is required to establish that in the target population there is a mechanism similar to the mechanism of action identified in the study population and there is no counteracting mechanism that masks its effect. That procedure involves both statistical evidence and evidence of mechanisms. Evidence of mechanisms can address the similarity between the mechanisms of action, but it can hardly rule out the existence of counteracting mechanisms. On the other hand, statistical evidence can rule out the existence of counteracting mechanisms, but it cannot address the similarity between the mechanisms of action.

It is time now to scrutinize those arguments presented in subsection 2.1 against the reliability of extrapolations based on statistical evidence, i.e., the “universal response” and the “context-variance” arguments. In the remainder of section 2, we will discuss whether evidence of mechanisms is required to overcome those difficulties —i.e., whether it effectively complements statistical evidence— and, consequently, whether the standard statistical approach should be somehow supplemented by considering evidence of mechanisms. Let us begin with the absence of universal responses.

It may be acknowledged that evidence of mechanisms is very useful to specify subpopulations within the target population with different responses to the intervention. Recall the example mentioned in the previous subsection. The effectiveness of a hypertensive prevention health policy may be increased on the basis of that kind of knowledge. Evidence of mechanisms pointed at ethnic differences as a crucial factor and RCTs should not overlook this fact. It is fair to say that, in order to minimize the potential heterogeneity of the treatment’s effects, different strategies to design trials have been developed: subgroup analysis, latent classes analysis (LCA)... (see below). Nevertheless, the role played by evidence of mechanisms in those strategies is similar to that played by background knowledge. It may be relevant to define the sample space, i.e., the content of those hypotheses which would be tested. Hence, in the previous example, a sensible RCT should include two groups —the experimental and the control group— which differ in ethnicity. But after then, frequencies would still be decisive since the goal is to check if there are statistically significant differences between both groups.

Then, it is worthwhile to distinguish between two different roles for evidence: the *testing role* —in an RCT, the frequencies obtained— and the *heuristic role* —i.e., that evidence not provided by the experiment itself but invoked to devise the sample space where testing takes place. The experiment’s outcome is the testing evidence for the causal hypothesis at issue, but the information previously taken into account for devising the experiment does not play the same role. Statistical evidence plays a testing role in RCTs, while the evidence of mechanisms incorporated when setting the sample space fulfils a heuristic role. Occasionally evidence of mechanisms may be very helpful, yes, but it should be remarked that here we are interested in the evidential support for our extrapolations, and not in how to devise an experiment.

We extrapolate a causal hypothesis provided that the *testing* evidence obtained in the experiment supports it. To insist that evidence of mechanisms *could* play a heuristic role in experimental design does not ground by itself pluralists’ claim for indispensability. It should be emphasized that sometimes background knowledge also plays that very same

role and it is not necessarily knowledge of mechanisms; it may be about previous known frequencies, expert judgment, common medical lore, etc. Common medical lore, for instance, may be based on very rough empirical generalizations. Properly speaking, this is not statistical evidence. But neither is it evidence of mechanisms. And the same could be said for some other sources of evidence (analogical reasoning from animals, biopsies...). Particularly, subgroup analysis is a standard statistical procedure where suspicious variables should be included as predictors (see Rothwell, 2005b). LCA is a further option where subgroups are established after some empirical data are gathered (Magidson & Vermunt, 2005). Both strategies rely, to a greater or lesser extent, on pertinent background knowledge but neither of them necessitates that it includes evidence of mechanisms.

Accordingly, since evidence of mechanisms is not the only option to avoid hasty generalizations —i.e., a “universal response” — the indispensability claim in favour of evidence of mechanisms does not follow from the fact that occasionally this sort of evidence is considered for setting the sample space.

Turning now to the “context-variance” argument, insofar as randomization and high control of potential distorting influences are defining features of RCTs, evidence provided by them is obtained in highly artificial conditions. Things could be different in a real scenario, after all, and even though several strategies have been developed in statistical practice for coping with this question, neither of them a priori guarantees that extrapolation will be successful.⁶ It should be reminded that evidence of mechanisms is presumably indispensable insofar as it compensates the *intrinsic* limitations of statistical evidence. And here we have an essential feature —the contrast between artificial and natural conditions— of the supposedly best statistical evidence we can get, i.e., that obtained in RCTs.

However, what are the prospects of avoiding the perils of context-variance by resorting to evidence of mechanisms? First of all, it is worth pointing out that evidence of mechanisms may be obtained from different sources. If it is recorded in a laboratory setting, we are in a parallel case to that of RCTs. It could also be obtained from case studies—autopsies, for instance. But, analogously, the degree of intervention when correlations are involved may vary. RCTs, observational and cohort studies, retrospective studies... offer a continuous range of options from more artificial to more “real world” conditions. Artificiality can hardly be considered, then, as a distinctive feature of statistical evidence when compared to evidence of mechanisms. Besides, even if we accept that getting statistical evidence involves, by and large, a more noticeable degree of intervention than evidence of mechanisms, context-variance could still be no less of a threat to evidence of mechanisms, as we will argue in the next subsection.

2.3. MECHANISMS, EXTRAPOLATION, AND CONTEXT-VARIANCE

The study and the target populations are, in fact, *different* populations. The question is how do we know that those populations are similar —or alternatively, that they do not differ— in the relevant respects, that is, in those factors which could have a non-negligible effect on the results obtained with the study population.

⁶ Two-stage designs, which involve over-sampling on the study population, is one example of such a strategy. See Rothman *et al.* (2008), chaps. 8 and 15, for further details.

From this point of view, the problem for both advocates of the standard statistical approach and those who underline the relevance of evidence of mechanisms is to justify the similarity claim. Trialists' efforts in RCTs are aimed at neutralizing those well-known factors which may bias the results (doctor-patient relationship, patient preference, and placebo effects). Blinded treatment allocation, exclusion of patients or clinicians with strong preferences, and placebo control are routinely applied. These issues are primarily related to internal validity but, as might be expected, there may be unsurmountable obstacles for full control on these factors outside trials, that is, in real-world situations.⁷ As the difficulties for neutralizing potentially biasing factors increase, the distance between the experimental population and the target population grows and the extrapolation claim gets riskier. All these complications demand caution in extrapolation, certainly. However, study-to-target extrapolations based on statistical evidence assume that both populations are similar, or at least, that they are not so dissimilar as to cast doubt on the inferences obtained in the study population.

On the other hand, mechanism-based extrapolations demand that mechanisms similar (in the relevant aspects) to the relevant mechanisms operating in the study population are present in the target population.⁸ Let S = study population, T = target population, and three indexed variables: I_i for interventions in those populations, M_i for mechanisms operating in them, and O_i for the respective outcomes of interventions. Suppose that we have good evidence that a particular intervention I_i brings about O_i through M_i in S . Then, a causal relationship between I_i and O_i has been ascertained in S . Nonetheless, for a reasonable mechanism-based extrapolation, it would also be required that a mechanism similar to M_i in the relevant aspects is present in T .

Steel (2008) has suggested a particular procedure ("comparative process tracing") to compare mechanisms operating in both populations and to discern their similarity. Given our limited knowledge about the mechanisms at work, it has been argued that comparative process tracing is usually unfeasible (Reiss, 2010; Howick *et al.*, 2013b; van Eersel *et al.*, 2019). We will not discuss Steel's proposal here (see section 3). Nevertheless, even if it is verified that a mechanism similar to M_i in the relevant aspects is present in T , contextual variation may undermine the causal extrapolation.

In the first place, consider the so-called problem of masking (Clarke *et al.*, 2013, 2014; van Eersel *et al.*, 2019). In the target population, there may be unknown disturbing mechanisms that influence the outcome. Even if a mechanism similar enough to M_i is present in T , there may be unknown mechanisms that affect the outcome and, eventually, mask its influence. Those unknown disturbing mechanisms could even interfere with the identified mechanism and undermine its own contribution to the outcome.

The challenge posed by masking could be illustrated by means of the relationship between exercise and weight loss.⁹ In a study population whose eating habits are regulated

⁷ Rothwell (2005a) offers detailed recommendations to minimize the risk when extrapolating RCTs results to particular clinical settings in daily practice.

⁸ A mechanism (or an aspect of a mechanism) is considered *relevant* for a causal relationship in a particular population if a change on it could make a difference to whether the causal relationship holds in that population (see Steel, 2008, p. 89).

⁹ Here we add minor variations on an example discussed in several places. See, for instance, Steel (2008) p. 68 and Clarke *et al.* (2014) p. 351.

(e.g. a military unit), it is the case that additional physical exercise causes more calories to be burned and results in weight-loss. However, in a target population whose eating habits are not regulated, a disturbing mechanism may hinder the loss of weight. Consider, for instance, people who suffer from knee pain. Prescription of adequate physical exercise to lose weight is recommended in order to alleviate that physical discomfort. However, provided that physical exercise stimulates appetite and that there is no regulation on eating habits, increase in caloric intake could minimize the expected weight loss—it could even bring about the opposite effect, that is, increasing weight gain—in the target population.

Secondly, mechanisms' absence of regularity should not be overlooked (Howick *et al.*, 2010, 2013b; van Eersel *et al.*, 2019). Similar mechanisms may be present in the study and the target populations, but not behave in a similar way. Even if a mechanism similar to M_j in several relevant aspects is present in T , it may have unanticipated and paradoxical effects in that population. Absence of regularity in mechanisms is of particular concern when the study population is artificially built in the laboratory. The fact is that the behaviour of a mechanism in the laboratory may differ significantly from its behaviour in a real setting (Howick *et al.*, 2013a, 2013b). By way of illustration, consider the following example about non-selective β agonists:

Given that β_1 and β_2 adrenoceptors are present in both the heart and the vasculature, and the existence of various neurally mediated systemic baroreflex functions, it is not surprising that the integrated response at the level of the individual to non-selective β agonists such as isoprenaline (isoproterenol) may be hypertension or hypotension and tachycardia or bradycardia. (Smith *et al.*, 2012, p. 178)

Given that, contextual variance seems as dangerous at least for evidence of mechanisms as it is for statistical evidence. Admittedly, there may be great differences about the potential effect of contextual variation in causal extrapolations, whether extrapolations are based on statistical evidence or evidence of mechanisms. A case-by-case scrutiny seems mandatory here. Nevertheless, we think that those considerations introduced in subsections 2.2 and 2.3 do suffice to discard the alleged complementarity afforded by evidence of mechanisms with respect to statistical evidence. Provided that complementarity is the main rationale to consider that evidence of mechanisms is necessary in study-to-target causal extrapolation, we conclude that the indispensability thesis should be put aside. Our conclusion, nonetheless, does not entail that evidence of mechanisms has no role to play when causal extrapolation is at issue.

3. *The (non-indispensable) role of evidence of mechanisms*

Evidence of mechanisms is not indispensable for the extrapolation of causal claims. As it has been argued, it is not necessary for dealing with the heterogeneity of responses and is hardly able to address the variance resulting from differences in the context. However, this does not entail that evidence of mechanisms is irrelevant or useless for causal extrapolation. In fact, most authors, including those who do not endorse the indispensability thesis, agree that evidence of mechanisms could be a valuable resource (Howick *et al.*, 2013a, 2013b; Parkkinen *et al.*, 2018; Marchionni & Reijula, 2019; Anjum *et al.*, 2020). The aim of this section is to discuss the proper role of evidence of mechanisms in causal extrapolation. For

that purpose, a distinction between a positive and a negative role of evidence of mechanisms in causal extrapolation will be introduced.

In the literature about the contribution of evidence of mechanisms for establishing causal claims in study populations, a distinction is often made between a positive and a negative role (Steel, 2004, 2008; Russo & Williamson, 2007; Illari, 2011; Clarke *et al.*, 2014). This distinction is nicely expressed by Steel as follows:

On the positive side, we can infer that X is a cause of Y if we know that there is a mechanism through which X influences Y . The negative flip side is that if no plausible mechanism running from X to Y can be conceived of, then it is safe to conclude that X does not cause Y , even if the two variables are probabilistically dependent. (Steel, 2004, p. 56)

This distinction has a prominent place in the analysis of the importance of evidence of mechanisms for establishing causal claims. It helps to accurately identify the strengths and weaknesses of evidence of mechanisms. In fact, some authors have argued that one aspect of evidence of mechanisms' contribution is actually helpful while the other is not. For example, regarding establishing causal claims in the social sciences, Steel claims that the positive role is significant, but the negative role "is undermined by the ease of imagining plausible mechanisms that could link nearly any two macrolevel social variables" (2008, p. 196).

A distinction between a positive and a negative role of evidence of mechanisms in causal extrapolation could also be made. It could be formulated as follows. On the positive side, if the relevant mechanisms at work in the study and the target population are highly similar in the relevant aspects, the extrapolation of the causal claim *is* justified. On the negative side, if the relevant mechanisms at work in the study and the target population differ in relevant aspects, the extrapolation of the causal claim *is not* justified. This claim does not entail that the causal relation does not hold in the target population or that the causal claim could not be legitimately extrapolated to the target population from another study population. It just means that, given the relevant differences between the mechanisms at work in the two populations, it is ungrounded to infer that the causal relation holds in the target population from the fact that it holds in the current study population.

It should be noted that the positive side is nearly gradual, while the negative is not. The more similarities are known between the relevant aspects of the relevant mechanisms in the two populations, the more justified is the extrapolation (Williamson, 2019). However, one relevant difference between them is enough for considering that the extrapolation is not justified. Even if the relevant difference (and its influence on the causal relationship) were known in detail, the extrapolation of the causal claim would hardly be legitimate. Certainly, in that scenario, it may be possible to infer the causal relationship held in the target population. But, given the relevant dissimilarity, the particular causal claim established in the study population could not be extrapolated to the target population.

The idea that evidence of mechanisms can support or undermine the extrapolation of a causal claim is not entirely new. Some authors have recently referred to it (e.g. Howick *et al.*, 2013a, p. 85; Aronson *et al.*, 2018, p. 1172; Parkkinen *et al.*, 2018, p. 12). Nevertheless, the two roles of evidence of mechanisms have never been analysed in depth nor taken into account for evaluating the relevance of evidence of mechanisms in causal extrapolation.

tion.¹⁰ We think that, as it was the case with the establishment of causal claims, clearly distinguishing between the positive and the negative scenarios and considering them separately could shed light on the appropriate role of evidence of mechanisms in causal extrapolation.

3.1. THE NEGATIVE ROLE

In the first place, the negative role of evidence of mechanisms in causal extrapolation will be considered. The basic idea is that if the relevant mechanisms at work in the study and the target population differ in relevant aspects, the extrapolation *is not* justified. In order to evaluate the actual relevance of the negative role, three aspects will be considered: (i) the availability of a procedure for obtaining the necessary information about the relevant mechanisms, (ii) how it is affected by the shortcomings encountered by the mechanisms approach, and (iii) the number and diversity of real cases that exemplify it.

Regarding the first aspect, it seems that a procedure for obtaining the information required for the negative mechanism-based inference is available. Consider comparative process tracing, which is supported by several advocates of the mechanistic framework (e.g. Steel, 2008; Clarke *et al.*, 2014). According to this procedure, firstly, the relevant mechanisms in the study population must be carefully scrutinised by means of process tracing or other methods. And secondly, the relevant mechanisms in the study and the target population must be compared in those stages in which the two are more likely to significantly differ. This procedure could provide the required information. Comparing the relevant mechanisms in both populations in those stages in which they are more likely to differ may result in the identification of a relevant difference between them. Moreover, the necessary knowledge to use this procedure is usually available. Comparative process tracing does not presuppose a detailed knowledge about the relevant mechanisms in the study and the target population. In order to identify a relevant difference by means of this procedure, it is only required to have knowledge about likely dissimilarities between mechanisms in both populations. Certainly, very often not all the stages in which relevant mechanisms are likely to differ are known. Nonetheless, that knowledge is not necessary for comparative process tracing to provide the information required for the negative mechanism-based inference. Given that not all relevant differences between the relevant mechanisms must be identified (it is only required to identify at least one relevant difference), knowing some stages in which they are likely to differ may be enough.

¹⁰ There are authors who have both considered the diversity of roles of evidence of mechanisms and discussed the relevance of evidence of mechanisms in causal extrapolation. Nevertheless, none of them have analysed the diverse roles of evidence of mechanisms in causal extrapolation. Their considerations about the diversity of roles of evidence of mechanisms refer to other issues (e.g. to establish a causal claim in a study population) and are not introduced in the discussion about causal extrapolation. For example, Steel (2004) distinguishes between a positive and a negative role of evidence of mechanisms to establish a causal claim in a study population. As it has been noted, that distinction acts as guide in his discussion of the importance of evidence of mechanisms to establish a causal claim. However, when Steel (2008) addresses the relevance of evidence of mechanisms in causal extrapolation, he does not distinguish between different roles for evidence of mechanisms. Actually, in the development of comparative process tracing, the diversity of roles of evidence of mechanisms is not taken into account.

The negative role of evidence of mechanisms is not severely undermined by those objections commonly raised against the mechanisms approach. Although they affect it, the negative role is still operative. One of the main shortcomings is the fragmentary knowledge about the mechanisms at work in *both* populations (Reiss, 2010; Howick *et al.*, 2013a, 2013b; van Eersel *et al.*, 2019). Even when a mechanism is identified in a population, the knowledge about it is frequently very limited. In particular, it is extremely difficult to fix mechanisms' boundaries, i.e., identifying all their component entities and activities. Nevertheless, it is possible to identify a relevant difference between the mechanisms at work in the two populations in this scenario. Even if not all the mechanisms' component entities and activities are known, a relevant difference between them may be identified. For instance, although there were significant gaps in our understanding of how a particular compound is metabolised by humans and mice, it would be possible to identify a relevant difference between their metabolic mechanisms. As it has been noted, the identification of all the relevant differences is not required in the negative scenario.

Another difficulty faced by the mechanistic stance, which has already been introduced in subsection 2.3, is masking (Clarke *et al.*, 2013, 2014; van Eersel *et al.*, 2019). Even if some mechanisms are identified and scrutinised in a target population, there may be unknown mechanisms that affect the outcome and, eventually, mask their influence. Nonetheless, masking is not especially threatening for the negative role of evidence of mechanisms. Admittedly, some disturbing mechanisms may go unnoticed. However, it is unlikely the case that, because of their influence, the negative mechanism-based inference leads to a wrong conclusion. Imagine that a relevant difference is identified between the mechanisms at work in the study and the target population and, consequently, it is concluded that the extrapolation is not justified. The presence of unnoticed disturbing mechanisms in the target population would only undermine that conclusion if those mechanisms operated so that they exactly compensated the identified difference. And that exact counterbalance is highly unlikely given the complexity —e.g., a high density of interactions among their components— of most biological and social mechanisms (Wagenaar, 2007; Howick *et al.*, 2010; Howick, 2011; Andersen, 2012).

A third important problem for the mechanisms approach is that mechanisms might not behave regularly (Howick *et al.*, 2010, 2013b; van Eersel *et al.*, 2019). As it has been argued in subsection 2.3, similar mechanisms may be present in the study and the target population but not behave in a similar way. As in the case of masking, mechanisms' possible absence of regularity is not very jeopardizing for the negative role. Surely, mechanisms may not behave in the target population as in the study population. However, this absence of regularity would hardly undermine the conclusion of the negative mechanism-based inference. Imagine that a relevant difference is identified between the mechanisms at work in the study and the target population and, therefore, it is concluded that the extrapolation is not justified. The irregular behaviour of a/some mechanism/s in the target population would only undermine that conclusion if it exactly compensated the identified relevant difference. Again, given the complexity of the studied mechanisms, it is very unlikely.

Finally, there have been identified many, and diverse, real cases where relevant differences between the mechanisms at work in the study and the target population undermined (or would have undermined) causal extrapolation. Some prominent examples are, in medicine, the non-recommendation of streptomycin therapy by the Medical Research Council in the 1940s (Clarke *et al.*, 2014); in evidence-based policy, the failure of the Bangladesh

Integrated Nutrition Program (Cartwright, 2012); and, in drug approval, the disallowance of abaloparatide by the European Medicines Agency (Aronson *et al.*, 2018).

To conclude, evidence of mechanisms can provide a basis for concluding that the extrapolation of a causal claim *is not* justified. As it has been discussed, a procedure for obtaining the necessary information is available, it is not severely affected by the difficulties faced by the mechanistic framework, and many and diverse real cases exemplify it. The negative role of evidence of mechanisms in causal extrapolation is so vindicated.

3.2. THE POSITIVE ROLE

Let us turn to the positive role of evidence of mechanisms in causal extrapolation. According to it, if the relevant mechanisms at work in the study and the target population are highly similar in the relevant aspects, the extrapolation of the causal claim *is* justified. In order to evaluate the actual relevance of the positive role, the three abovementioned aspects (i.e. availability of a procedure, difficulties faced by mechanism-based approaches, and real cases that exemplify it) will be considered.

With regard to the first aspect, it is not clear whether a procedure for obtaining the information required for the inference of the positive scenario is available. Specifying the degree of similarity between the relevant mechanisms at work in the study and the target population is a challenging question for comparative process tracing, the procedure suggested by advocates of the mechanism-based approach. In the first place, in order to specify the degree of similarity, that procedure requires knowing all (or, at least, most) stages in which significant differences between relevant mechanisms are likely to occur. However, previous knowledge about similarities and differences between populations is typically insufficient for identifying all (or most of) those stages (Reiss, 2010; Howick *et al.*, 2013a, 2013b; van Eersel *et al.*, 2019). In particular, knowledge about the target population is often remarkably incomplete. It is worth mentioning that, in some cases, that population does not even exist yet. And, provided that we manage to identify all the stages in which populations are likely to differ, there would be too many to compare and comparison of all (or most) of them would require a vast amount of resources and time. On account of this, Steel (2008, p. 90) has proposed a shortcut for making that comparison feasible. He claims that, when several stages of likely difference are upstream of another with respect to the causal chain, it may be enough to compare the downstream stage.¹¹ However, that circumvention can rarely be used. Steel warns that, in order to take advantage of the shortcut, two requirements must be satisfied: (i) there must not be a path that bypasses the downstream stage and (ii) the upstream stages must leave a distinctive mark upon the downstream stage. Otherwise, some relevant differences between the mechanisms at work in the study and the target population may go unnoticed. Nevertheless, given the fragmentary knowledge about relevant mechanisms, it is often impossible to confirm whether the requirements specified by Steel are fulfilled.

¹¹ Steel (2008) argues that comparing two causal chains in their latest stage of likely difference can reveal whether there is a significant dissimilarity between them in any stage of likely difference. Dissimilarities in upstream stages result in divergences regarding downstream stages. Consequently, any significant dissimilarity in a stage of likely difference, no matter where it is located, should manifest in that downstream stage.

Concerning the difficulties usually encountered by the mechanisms approach, unlike the negative role, the positive role seems to be seriously undermined by them. Firstly, consider the limited knowledge about the relevant mechanisms at work in the study and the target populations (Howick *et al.*, 2013a, 2013b; van Eersel *et al.*, 2019). Unfortunately, in many cases it is extremely difficult to fix the relevant mechanisms' boundaries and identify all their components. This situation poses a major challenge to the positive role. If only some components of the relevant mechanisms are known, it is very complicated to properly compare and establish if there is a high degree of similarity between them. That comparison is hardly feasible unless all (or most) components of relevant mechanisms are known. Secondly, think about the masking problem (Clarke *et al.*, 2014; van Eersel *et al.*, 2019). It also represents an obstacle for the positive role. Even if several relevant mechanisms were identified and scrutinised in both populations and it was confirmed that they are highly similar, there could be unknown mechanisms in the target population that influence the outcome. Those unknown disturbing mechanisms would modify the causal relationship held and, consequently, the extrapolation of the causal claim established in the study population would be erroneous. And thirdly, consider the absence of regularity (Howick *et al.*, 2010, 2013b; van Eersel *et al.*, 2019). It undermines the positive role of evidence of mechanisms too. Even if the relevant mechanisms at work in the study and the target population are highly similar, they may behave differently in the target population. That change in their behaviour would modify the causal relationship held in the target population. As a consequence, the extrapolation of the causal claim established in the study population would be mistaken.

Most advocates of the mechanistic framework consider that evidence of mechanisms should be complemented with statistical evidence (Russo & Williamson, 2007; Clarke *et al.*, 2013, 2014; Parkkinen *et al.*, 2018).¹² Statistical evidence is considered an adequate resource to address the main weaknesses of evidence of mechanisms. Consequently, it could be argued that the aforementioned contentions would not be a threat as long as evidence of mechanisms is properly used (i.e. in conjunction with statistical evidence). Complementing mechanism-based extrapolation with statistical evidence would enable us to overcome them. Nonetheless, that approach faces some relevant difficulties.

In the first place, statistical evidence could hardly mitigate certain problems of mechanism-based causal extrapolation. For example, statistical evidence could not help with the lack of knowledge about the relevant mechanisms. Statistical evidence is evidence about whether and to what extent the effect varies with the postulated cause, not about the existence or the properties of mechanisms in the relevant domain (Illari, 2011). As it has been noted (see footnote 5), the distinction between statistical evidence and evidence of mechanisms is precisely a distinction about objects of evidence. Consequently, by definition, statistical evidence could hardly increase our knowledge about the relevant mechanisms. Statistical evidence would also be of little help to address the absence of a procedure for obtaining the information required for the inference in the positive scenario.

¹² It should be noted that not all advocates of the mechanistic standpoint consider that evidence of mechanisms by itself cannot justify the extrapolation of a causal claim and must always be complemented. For example, Steel (2008) argues that evidence of mechanisms alone can adequately guide the extrapolation of a causal claim. Nevertheless, he admits that the mechanistic framework cannot always be applied.

Secondly, although statistical evidence could be helpful in order to overcome some problems faced by mechanism-based causal extrapolation, researchers can rarely benefit from its assistance. Certainly, statistical evidence could address the problem of masking (Clarke *et al.*, 2013, 2014). It measures the overall causal effect (i.e. the net outcome) in the target population and reveals if there are disturbing mechanisms that modify the causal relation. Likewise, statistical evidence could mitigate mechanisms' absence of regularity. By means of studying what happens overall, it identifies if the behaviour of the mechanisms as a whole is the same (i.e. they produce the same net outcome) in the study and the target population. Nonetheless, quite often the aid of statistical evidence is not available in causal extrapolation. Extrapolating causal claims is particularly relevant when we have limited access to the population of interest. We attempt to extrapolate a causal claim to the target population of interest from a different study population precisely because we cannot directly study the target population. Needless to say that if we had full access to the population of interest and could directly study it, we would focus on identifying the causal relationship held in that population (not on extrapolating a certain causal claim from a different population).¹³ The reasons of our limited access to the target population can be ethical, economic, political, technological, etc. For example, we attempt to extrapolate the effect of a nutrition policy in European children living now and in the future from its past effect in a particular French town because we cannot directly study all the current and future European children. This means that, when causal extrapolation is concerned, statistical evidence about the causal relation in the target population is usually very scarce and can hardly mitigate the aforementioned problems.

Lastly, there have been identified several real cases where a high degree of similarity between the relevant mechanisms at work in the study and the target population supported (or would have supported) causal extrapolation. Some notorious examples are related to animal models: the extrapolation of the carcinogenic effects of aflatoxin B1 from rats to humans (Steel, 2008), of the causal relation between cholesterol-rich diet and atherosclerosis from rabbits to humans (Gillies, 2018), and of the carcinogenic effects of benzo[a]pyrene from animals to humans (Wilde & Parkkinen, 2019). Nonetheless, there are fields of science where real cases that exemplify the positive scenario are hard to find. For example, few real examples of positive mechanism-based inferences are known in the social sciences (Reiss, 2010; van Eersel *et al.*, 2019). Steel (2008), after analysing several real cases of extrapolation in the social sciences, acknowledges that he could not find any one that exemplifies it. Furthermore, the main case study in support of the mechanisms approach in the social sciences—i.e., the study of the Bangladesh Integrated Nutrition Program—supports the negative role of evidence of mechanisms, but not the positive one (Howick *et al.*, 2013b).

Summing up, the actual relevance of the positive role of evidence of mechanisms in causal extrapolation is uncertain. Although in principle that sort of support is possible, a procedure for obtaining the required information is rarely available and positive mechanism-based inferences are significantly affected by the difficulties faced by the mechanisms

¹³ By “full access to the population of interest”, we allude to a scenario in which several experimental or quasi-experimental studies could be developed, so that compelling evidence about the existence of the causal relationship at issue in that population could be ascertained.

approach. Moreover, although there are several real cases that exemplify it, they are extremely hard to find in some fields of science.

4. Conclusion

Extrapolating a causal claim from a study population to another population of interest is a controversial issue. The standard statistical approach, which relies on RCTs or, in their absence, on quasi-experimental designs, is not devoid of difficulties. The response to an intervention may vary from one individual (or group of individuals) to another and differences in the respective contexts of both populations are not always neutral regarding the expected outcome. Accordingly, some scientists and philosophers of science have argued that evidence of mechanisms is indispensable for causal extrapolation. However, we have tried to show that evidence of mechanisms is not necessary for dealing with the heterogeneity of responses and that it is not less sensitive to contextual variance than statistical evidence. Nonetheless, this does not necessarily mean that evidence of mechanisms is irrelevant or useless for causal extrapolation. In order to identify the proper role of evidence of mechanisms in causal extrapolation, we have introduced a distinction between a positive and a negative role. In our view, evidence of mechanisms can support or undermine the extrapolation of a causal claim from the study population to the target population. In particular, while the positive role is highly questionable, the negative role is reliable and valuable for causal extrapolation.

Acknowledgments

Previous versions of this paper were presented at: *Fourth Meeting in Philosophy, Probability and Scientific Methodology* (University of Valencia), *TINT-Centre for Philosophy of Social Science* (University of Helsinki), and *Workshop on Reasoning about Evidence* (Ghent University). We are grateful to the audiences. We should also thank two anonymous reviewers for their helpful comments.

We gratefully acknowledge funding from the Spanish Ministry of Universities under grant FPU16/03274 (Saúl Pérez-González), and the Spanish Ministry of Science and Innovation under research projects FFI2016-76799-P (Valeriano Iranzo) and FFI2017-89639-P (Saúl Pérez-González).

REFERENCES

- Andersen, H. (2012). Mechanisms: what are they evidence for in evidence-based medicine? *Journal of Evaluation in Clinical Practice*, 18, 992-999.
- Anjum, R. L., Copeland, S., & Rocca, E. (2020). Medical scientists and philosophers worldwide appeal to EBM to expand the notion of 'evidence'. *BMJ evidence-based medicine*, 25, 6-8.
- Aronson, J. K., La Caze, A., Kelly, M. P., Parkkinen, V., & Williamson, J. (2018). The use of mechanistic evidence in drug approval. *Journal of Evaluation in Clinical Practice*, 24, 1166-1176.

- Bechtel, W., & Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36, 421-441.
- Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton University Press.
- Bell, S. M., Chang, X., Wambaugh, J. F., Allen, D. G., Bartels, M., Brouwer, K. L., ... & Jarabek, A. M. (2018). In vitro to in vivo extrapolation for high throughput prioritization and decision making. *Toxicology in Vitro*, 47, 213-227.
- Campbell, D. T., & Cook, T. D. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago, IL: Rand McNally College Publishing Company.
- Cartwright, N. (2012). Presidential address: Will this policy work for you? Predicting effectiveness better: How Philosophy Helps. *Philosophy of Science*, 79, 973-989.
- Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2013). The evidence that evidence-based medicine omits. *Preventive Medicine*, 57, 745-747.
- Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi*, 33, 339-360.
- Craver, C. F., & Alexandrova, A. (2008). No revolution necessary: Neural mechanisms for economics. *Economics & Philosophy*, 24, 381-406.
- Darden, L. (2018). Strategies for discovering mechanisms. In S. Glennan & P. M. Illari (Eds.). *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 255-266). Abingdon, UK: Routledge.
- Gillies, D. (2018). *Causality, probability, and medicine*. London, UK: Routledge.
- Glennan, S. (2017). *The new mechanical philosophy*. Oxford, UK: Oxford University Press.
- Glennan, S., & Illari, P. M. (2018). Introduction: Mechanisms and mechanical philosophies. In S. Glennan & P. M. Illari (Eds.). *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 1-9). Abingdon, UK: Routledge.
- GRADE working group. (2016). Criteria for applying or using GRADE. https://www.gradeworkinggroup.org/docs/Criteria_for_using_GRADE_2016-04-05.pdf
- Hedström, P. (2005). *Dissecting the social. On the principles of analytical sociology*. New York, NY: Cambridge University Press.
- Hedström, P., & Ylikoski, P. (2010). Causal mechanisms in social sciences. *Annual review of sociology*, 36, 49-67.
- Howick, J. (2011). Exposing the vanities—and a qualified defense—of mechanistic reasoning in health care decision making. *Philosophy of Science*, 78, 926-940.
- Howick, J., Glasziou, P., & Aronson, J. K. (2010). Evidence-based mechanistic reasoning. *Journal of the Royal Society of Medicine*, 103, 433-441.
- Howick, J., Glasziou, P., & Aronson, J. K. (2013a). Can understanding mechanisms solve the problem of extrapolating from study to target populations (the problem of 'external validity')? *Journal of the Royal Society of Medicine*, 106, 81-86.
- Howick, J., Glasziou, P., & Aronson, J. K. (2013b). Problems with using mechanisms to solve the problem of extrapolation. *Theoretical Medicine and Bioethics*, 34, 275-291.
- Illari, P. M. (2011). Mechanistic evidence: Disambiguating the Russo-Williamson thesis. *International Studies in the Philosophy of Science*, 25, 139-157.
- Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal of Philosophy of Science*, 2, 119-135.
- Jiménez-Buedo, M., & Miller, L. M. (2010). Why a trade-off? The relationship between the external and internal validity of experiments. *Theoria*, 69, 301-321.
- Kincaid, H. (2012). Mechanisms, causal modeling, and the limitations of traditional multiple regression. In H. Kincaid (Ed.). *The Oxford handbook of philosophy of social science* (pp. 46-64). Oxford, UK: Oxford University Press.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1-25.

- Magidson, J., & Vermunt, J. K. (2005). A nontechnical introduction to latent class models. *DMA Research Council Journal*, 2005, 1-15.
- Marchionni, C., & Reijula, S. (2019). What is mechanistic evidence, and why do we need it for evidence-based policy? *Studies in History and Philosophy of Science*, 73, 54-63.
- Parkkinen, V., Wallmann, C., Wilde, M., Clarke, B., Illari, P., Kelly, M. P., Norell, C., Russo, F., Shaw, B., & Williamson, J. (2018). *Evaluating evidence of mechanisms in medicine: Principles and procedures*. Cham, Switzerland: Springer.
- Reiss, J. (2010). Review of Daniel Steel's *Across the boundaries: Extrapolation in biology and social science*. *Economics & Philosophy*, 26, 382-390.
- Reiss, J. (2019). Against external validity. *Synthese*, 196, 3103-3121.
- Rothman, K. J., Greenland, S., & Lash, T. L. (Eds.). (2008). *Modern epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins.
- Rothwell, P. M. (2005a). External validity of randomised controlled trials: "to whom do the results of this trial apply?". *The Lancet*, 365, 82-93.
- Rothwell, P. M. (2005b). Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *The Lancet*, 365, 176-86.
- Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, 21, 157-170.
- Salmon, W. C. (1998). *Causality and explanation*. Oxford, UK: Oxford University Press.
- Smith, S. W., Hauben, M., & Aronson, J. K. (2012). Paradoxical and bidirectional drug effects. *Drug safety*, 35, 173-189.
- Steel, D. (2004). Social mechanisms and causal inference. *Philosophy of the Social Sciences*, 34, 55-78.
- Steel, D. (2008). *Across the boundaries: Extrapolation in biology and social science*. Oxford, UK: Oxford University Press.
- van Eersel, G. G., Koppenol-Gonzalez, G. V., & Reiss, J. (2019). Extrapolation of experimental results through analogical reasoning from latent classes. *Philosophy of Science*, 86, 219-235.
- Victoria, C. G., Habicht, J. P., & Bryce, J. (2004). Evidence-based public health: moving beyond randomized trials. *American journal of public health*, 94, 400-405.
- Wagenaar, H. (2007). Governance, complexity, and democratic participation: How citizens and public officials harness the complexities of neighborhood decline. *The American Review of Public Administration*, 37, 17-50.
- Wilde, M., & Parkkinen, V. (2019). Extrapolation and the Russo-Williamson thesis. *Synthese*, 196, 3251-3262.
- Williamson, J. (2019). Establishing causal claims in medicine. *International Studies in the Philosophy of Science*, 32, 33-61.

SAÚL PÉREZ-GONZÁLEZ holds a Ph.D. in Philosophy from the University of Valencia (Spain). His main areas of interest are philosophy of science, philosophy of the social sciences, and philosophy of the life sciences.

ADDRESS: Department of Philosophy, University of Valencia. Avd. Blasco Ibáñez, 30. 46010 Valencia. Spain. Email: saul.perez@uv.es. ORCID: 0000-0001-9594-9558

VALERIANO IRANZO is Professor at the University of Valencia (Spain). Visiting Fellow at Washington University (USA), University of Sheffield (UK), Rijksuniversiteit Groningen (The Netherlands) and University College London. His research interests are focused on philosophy of science, philosophy of probability, and epistemology.

ADDRESS: Department of Philosophy, University of Valencia. Avd. Blasco Ibáñez, 30. 46010 Valencia. Spain. Email: iranzov@uv.es. ORCID: 0000-0001-6432-9284