Comparative process tracing: yet another virtue of mechanisms?

Federica Russo

Philosophy, Louvain & Kent

f.russo@kent.ac.uk

Draft of 24 March 2009

To appear in *Journal of Economic Methodology*

Across the Boundaries suggests crossing frontiers. If you know already what the book is about, the title is adequate and suggestive; if you don't, the title is appealing and captivating—an effective encouragement to leaf through it. *Extrapolation*, which appears in the subtitle, is a term familiar to social scientists grown up in the quasi-experimental tradition, at least in its synonym form of 'external validity'. But the term is not ipso facto telling to philosophers of science. A search on the *Philosopher's Index* for the term 'extrapolation' returns 94 hits, of which only 5 or 6 are in fact relevant; 'external validity' doesn't fare much better, with only 27 hits and just a bunch of relevant records. Yet, this neglect shouldn't create a false impression—it is more a sign of an emerging awareness of the problem rather than of ignorance.

The first chapter brings the reader into the topic: extrapolation in the social and biological sciences is the problem of extending/exporting results, notably results about causal relations from one population—the observed population, or the 'model'—to a different population—the target population, or simply the 'target'. The peculiarity of the practice of extrapolation is that we cannot assume homogeneity between the model and target. Consider the following cases. In cancer research, attention was paid to the effects of aflatoxin. Research on this substance showed that the animal model best approaching human was the rat. Yet, a close examination of the mechanism in rats' and humans' metabolism, also showed significant differences that have to be taken into account in the process of extending results from the animal model to human beings. Around this example Steel builds and discusses much of his positive arguments. Another example, also examined by Guala in his *The methodology of Experimental Economics*, is taken from experimental economics. The Federal Communication Commission (FCC) conducted in 1993-94 a series of auctions of broadcasting licenses. Experiments were run in order to design and implement an auction mechanism that would, among other things, prevent monopolies and promote small business. A

number of auction mechanisms were proposed and tested in laboratories; eventually, the one selected and implemented in the real auction in 1994 delivered results that conformed to the expectations. Both these cases are well chosen examples of extrapolation, one from biological contexts to human contexts, the other from a restricted human population to a larger human population. What they have in common is that the results obtained in an experiment on a sample are meant to be 'exported' elsewhere. But what grants this 'export' process? This, in a nutshell, is what the book is about.

Steel's book fills a gap in the existing literature—only a few accounts address the issue of extrapolation. Steel stresses that those available are wanting because they do not provide adequate answers to two crucial questions, namely the *extrapolator's circle* and *the problem of difference*. The first is the challenge of successfully exporting information from the observed population knowing that such information is limited and partial. The animal model, for instance, has been oft criticised exactly because, it seems, its suitability rests on prior knowledge of the causal relationship in the target population. The second is the challenge of providing an account of extrapolation that is successful when differences between the observed and the target populations are present. Steel's solution lies in a *mechanism-based* approach to extrapolation. Earlier works, for instance Guala (2005), do recognise the centrality of mechanisms for extrapolation, but they also maintain that whether a model can be successfully extrapolated remains an empirical hypothesis. If so, Steel concludes, the extrapolator's circle is not answered, nor extrapolation is justified when relevant difference between the model and the target are present.

In describing very accurately the scientific practices of extrapolation, Steel puts forward a number of theses of primary theoretical importance. His account of mechanism-based extrapolation relies on the idea of comparing mechanisms in the model and in the target, where the mechanisms are most likely to differ, and is equipped with specific concepts, notably 'causal relevance' and 'disruption'.

Steel's develops his theses topically and gradually. Chapter 2 provides the reader with a 'toolkit' to deal with the rest of the book. The notions of 'intervention', 'causal effect', and 'causal relevance' are presented and explained at length. Yet the chapter is not merely descriptive, as much of what Steel says about 'causal relevance' relies on the 'standard literature' but also adds important clarifications. Notably, he distinguishes between a *comparative* and *monotonic* interpretation of causal relevance. According to the former, a claim of positive causal relevance indicates that the value of the putative effect-variable is greater when the value of the putative cause-variable is "raised from some basal, comparison value to any value within the interval" (Steel 2008, 22). According to the latter, positive

causal relevance indicates that the value of the putative effect-variable increases monotonically with the putative cause-variable throughout the interval. The intuition behind this distinction is that a cause may change either the variance of the putative effect-variable or its mean value. Much discussion in the literature focuses on average causal effects, and very little is about interventions on the variance of effect-variables.

Likewise, Chapters 3 and 4 discuss other key concepts: 'causal structure', 'mechanism', and 'disruption'. Mechanisms are identified with causal structures (in the relevant sense, this applies to both biological and social contexts), which generate probability distributions and indicate how these distributions change given a certain range of interventions. Social mechanisms, in particular, "are complexes of interacting agents—usually classified into specific social categories—that produce regularities among macrolevel variables" (p.48). Worth-noting is the 'disruption principle' which basically serves as a check for causal relevance. The idea is this (Steel 2008, 58-60). Suppose we know that a given factor X is a positive causal factor for Y in a population P. We then ask whether there exists a subset P' of P in which the effect of X on Y is nullified. This means that X is not causally relevant to Y in the subpopulation P', and we call factors acting this way 'disrupting factors'. Now, if X is a disrupting factor for Y in P', this means that every mechanism from X to Y is blocked. Disruption is a key concept in Steel's account because the extrapolation theorem relies on it.

The heart of the book is presented in Chapter 5, where comparative process tracing is presented and applied to extrapolation from animal models. Comparative process tracing is meant to assess the suitability of the model as a basis for extrapolation to the target. This involves a systematic comparison of the mechanisms in the model and in the target, especially at the most 'critical' points, that is where the mechanism in the model and in the target are most likely to differ. This methodology requires to learn about the mechanism in the observed population first, and *then* to compare it with the mechanism in the target population. The extrapolation theorem then goes along the following lines. Let the human population be our concern, in particular, we are interested in the question of whether a factor X (e.g., aflatoxin) is a positive causal factor (e.g., a carcinogenic substance) for Y, given that we know that X is a positive causal factor in, say, rats. We also know that the mechanism from X to Y in rats is 'positively consonant', i.e. "different combinations of mechanisms do not exert conflicting positive and negative influences" (Steel 2008, 108). We can then infer that X is a positive causal factor for Y in humans exactly if the relative frequency in those who possess an undisrupted mechanism is greater than zero. This, according to Steel, allows the

inference from mechanisms *plus* causal relevance in the model in rats to positive causal relevance in humans.

Chapters 6 and 7 digress on 'ceteris paribus' and 'reduction'. Steel presents original theses about those concepts. He distinguishes, following Schurz (2001, 2002), several senses of 'ceteris paribus' and argues that the ceteris paribus clause appearing in extrapolation claims is *not* a requirement of an absence of causally relevant disanalogies. Instead, ceteris paribus clauses ought to be understood as "indicating an *inference schema* that specifies sufficient conditions for extrapolating claims about positive or negative *causal relevance*" (Steel 2008a, p.124). Reductionism is approached from a singular perspective: Steel argues that his account of mechanism-based extrapolation does need a form of reductionism, notably that one level might turn out to be more fundamental than another, and yet this is compatible with pluralistic stances according to which there are autonomous levels of explanation.

Finally, the last two chapters tackle the problem of extrapolation in social science. Steel takes issue with Guala's account (2005) because although he endorses a form of mechanism-based approach, he is unable to offer an answer to the extrapolator's circle and to the problem of difference; the reason seems to be that Guala's account does not specify enough where and how comparisons between mechanisms have to occur. This problem was central for extrapolation in biology but so it is for economics and social science in general. In fact, in social contexts too we cannot have enough 'separate' knowledge of mechanisms in the model and in the target to grant extrapolation. This is the case in the experimental contexts already discussed by Guala, as well as in other domains such as psychology, sociology, or even non-experimental situations such as ethnography. Steel then takes side with Little—who thinks that social scientists need to make hypotheses about underlying mechanisms in order to establish causal claims—and against Kinkaid—who instead thinks it is false that, to establish whether X causes Y, at least a mechanism linking X to Y has to be identified. This emphasis on mechanisms is meant to address a well-known complaint of inaccuracy or approximation in the social sciences. This is, as economists know, the bulk of the so-called 'Lucas critique', according to which although econometric models may well succeed in making correct short-term economic forecasts, they can easily fail to predict the consequences of policy interventions because decision rules of agents would change as well. Steel seems to suggest that this does not speak against the reliability of economic or social science methodology; rather, this reveals that the literature hasn't emphasised enough that policy interventions are a *special* kind of interventions, which Steel calls 'structure-altering'. Structure-altering interventions make changes *in*

the causal structure, that is, they altogether change the causal mechanism. But the full-blown advantage of a mechanism-based approach is also meant to answer Kinkaid's objection. Contrary to 'direct causal inference', where the goal is simply to draw conclusions about relationships between variables from a large data set, Steel's approach allows to infer causal *mechanisms* relating macrolevel variables by reconstructing the interactions between parts of the system. This is certainly a step forward in understanding what causal analysis in economics and social science ought to do in order to provide successful explanations or forecasts.

The ideas defended in *Across the Boundaries* show a high internal consistency. It is certainly a merit that Steel is always attentive in giving the reader enough background to understand his points and the novelty of his ideas. The use of sources also deserves a note. Steel uses new sources—new to philosophers, at least—as well as mainstream literature. Steel manages to integrate skilfully the literature from biology and social science with the literature from philosophy. This gives a neat picture of the problem of extrapolation. The overall result is therefore remarkable. A book written with expertise and enthusiasm. The structure is articulated and the narrative fluid. Steel's vocabulary is technical at times, but when it is, it is always accompanied by useful explanations of formal language. The index and table of contents are excellent aids to understanding and utilisation.

There is much to be found there for philosophers, biologists, and social scientists. Steel puts forward a number of theses that deserve attention. Some, as I mentioned, concern refinements of familiar concepts, e.g. causal relevance or ceteris paribus, whilst others are novel theses, e.g. the *comparative* process tracing account of extrapolation. The book is certainly a major contribution to methodology of both biology and social science, and to mainstream and emergent topics in philosophy of science. The main merit of Steel's work is perhaps that he goes beyond a tradition initiated with Cook and Campbell (1979), which focuses too much on 'statistical threats' to external, e.g., statistical interaction between selection and treatment and between history and treatment. In *Across the Boundaries* we find a deeper examination: what allows the extrapolation is the presence and comparison of mechanisms in the observed and target population—this discussion is totally absent in Cook and Campbell.

Overall the approach is very promising and it will undoubtedly raise debate on an important issue that has been largely neglected in the philosophical literature. To be sure, discussion around Steel's ideas has already arisen (Guala 2008, Steel 2008a). This ping-pong debate between Steel and Guala is about *details* of their approaches. That comparative process tracing is, after all, a form of analogical reasoning is a concession, on Steel's side, that doesn't alter the basis of his project. This means, it

seems to me, that Steel and Guala are on the same track, hopefully the right one, and therefore they can work on issues of fine tuning. By and large, in tackling the problem of extrapolation and external validity, I tag along with Steel and Guala too. The points I am about to raise are thus expressions of minimal divergences that, if addressed, could make the account more complete and solid.

One specific issue concerns the extrapolator theorem. The theorem allows us to infer positive causal relevance in the target from mechanism *plus* positive causal relevance in the model. However, this is not yet the end of the extrapolation, as eventually extrapolation is about inferring the *mechanism* in the target. So, it seems to me, a step is missing. A mechanism is, according to Steel, a causal structure. In turn, a causal structure is that which produces probability distributions and indicates how these distributions will change given intervention. But even then, there might be a number of different mechanisms that give rise to the same probability distributions—a situation of underdetermination familiar to social scientists. So even if we successfully infer positive causal relevance in the target from knowledge of the mechanism *plus* positive causal relevance in the model, we haven't inferred the *mechanism* yet.

A more general issue concerns the exact meaning of extrapolation and external validity. Steel uses extrapolation in the following sense: "transferring causal generalizations from one context to another when homogeneity cannot be presumed" (p.3). This definition hides the same ambiguity of Cook and Campbell's definition of external validity, which equivocates between two senses: (i) generalising from the sample to a larger population, (ii) generalising to populations/settings other than those studied. But perhaps those two senses also require different extrapolation tools? Some social scientists, for instance Lucas (2003), seem to suggest so.

Related to this last point is the following problem. At times, Steel equivocates between two distinct questions. One is to extrapolate that a causal factor operating in the model also operates (or does not operate) in the target, *assuming* that the same mechanism operates in the model and in the target. Another is to extrapolate that a causal factor operating in the model also operates (or does not operate) in the target, *comparing* the mechanisms in the model and in the target. The aflatoxin and the auction examples fall under the latter question—comparative process tracing seems a good tool to use here because we track, along the mechanisms, how the particular factors behave, notably we check whether it is disruptive or not. Yet, extrapolation from randomised clinical trials (RCT) appears to be different. RCTs establish whether a given drug is effective or not. However, successful extrapolation in RCTs does not involve comparisons between two distinct mechanisms. Causal generalisations established by RCTs usually do not involve a description of the 'functioning' of the

mechanism. *First*, a comparison is made between the control and the test group, but it does not involve mechanistic considerations at all. *Second*, results of RCTs are applied to a larger population, but again, this does not involve mechanistic considerations. Steel might counterargue that this practice is wrong and that mechanistic considerations must enter extrapolation from RCTs. If so, this is certainly an area to look at more closely.

Finally, it is worth investigating whether extrapolation has the same characteristics and requires the same tools in other areas. An interesting domain is qualitative research in social science. Ethnography might be an interesting test-case as it is by and large observational and knowledge of the mechanism in the target population is not usually presupposed—the goal of extrapolation is just to 'translate' the mechanism from one population to another, rather than making comparisons (LeCompte and Goetz, 1983). Also, some authors suggested that external validity might also go the other way round, namely it is not used to make inference about larger/different population from the results of experiments, but it is sometimes used to test the consequences of a theory, as it happens in experimental psychology (Lucas 2003, Mook 1983).

At the beginning of the book Steel set two goals: to be of practical relevance to scientific methodology and to be of significant interest general philosophy of science. Steel manages to reach them both. Comparative process tracing describes and prescribes how extrapolations ought to be done. The account is also well equipped with a number of concepts that are certainly of much interest to philosophers of science. Steel has opened up a new path of research. Extrapolation and external validity deserve more philosophical discussion than they had so far. The topic is relevant to methodology also because it presents again, in a slightly new form, a perennial question about the differences and similarities between the social and the natural sciences, and between experimental and observational sciences. It is therefore an easy and safe prediction that comparative process tracing will keep philosophers and scientists busy for a while.

Acknowledgements I wish to thank the Fellows (Claus Beisbart, Erik Curiel, Laura Felline, Ulrich Krohs, Flavia Padovani, Chris Pincock, and John Norton) at the Center for Philosophy of Science (University of Pittsburgh) for very stimulating discussions on extrapolation and external validity. Comments by Caterina Marchionni have also been valuable and helpful. Financial supports from the FSR-FNRS (Belgium) is also gratefully acknowledged.

References

Cook, T. and Campbell, D. (1979). *Quasi-Experimentation. Design and Analysis Issues for Field Settings*. Rand MacNally, Chicago.

Guala, F. (2005). *The methodology of experimental economics*. Cambridge University Press.

Guala, F. (2008). Extrapolation without process tracing. Paper presented at the Philosophy of Science Biannual Meeting, Pittsburgh 2008.

LeCompte, M. D. and Goetz, J. P. (1983). Problems of reliability and validity in ethnographic research. *Review of Educational Research*, 52(1):31–60.

Lucas, J. W. (2003). Theory-testing, generalization, and the problem of external validity. *Sociological Theory*, 21(3):236–253.

Mook, D. G. (1983). In defence of external validity. *American Psychologist*, 38:379–387.

Schurz, G. (2001). Pietroski and Rey on ceteris paribus laws. *British Journal for the Philosophy of Science*, 52:359-370.

Schurz, G. (2002). Ceteris paribus laws: classification and deconstruction. *Erkenntnis*, 57:351-372

Steel, D. (2008). *Across the boundaries. Extrapolation in biology and social science*. Oxford University Press.

Steel, D. (2008a). A new approach to argument by analogy: extrapolation and chain graphs. Paper presented at the Philosophy of Science Biannual Meeting, Pittsburgh 2008.