F53's being a straightforward instrumentalist methodology was that 'if it walks like a duck, and quacks like a duck... then it's a duck', nothing more.

For a conference volume, this book contains some worthwhile essays about the history of F53. Unfortunately there is also recycled material and some papers spiced up with unnecessary critical comments that seem unfair to this reviewer.

Lawrence Boland
Simon Fraser University, British Columbia

REFERENCES

Boland, L. 1979. A critique of Friedman's critics. *Journal of Economic Literature* 17: 503–22.
Boland, L. 1985. Reflections on Blaug's *Methodology of Economics*: suggestions for a revised edition. *Eastern Economic Journal* 11:450–454.
Boland, L. 1997. *Critical Economic Methodology: A Personal Odyssey*. London: Routledge.
Boland, L. 2003. Methodological Criticism vs. ideology and hypocrisy. *Journal of Economic Methodology* 10: 521–526.
Mäki, U. 2003. 'The methodology of positive economics' (1953) does not give us *the* methodology of positive economics. *Journal of Economic Methodology* 10: 495–505.
Mayer, T. 2003. Symposium: Fifty years of Milton Friedman's 'The methodology of positive economics'. *Journal of Economic Methodology* 10: 493–530.
Wong, S. 1973. The 'F-twist' and the methodology of Paul Samuelson. *American Economic Review* 63: 312–325.

doi:10.1017/S0266267110000325

Across the Boundaries: Extrapolation in Biology and Social Science, Daniel P. Steel. Oxford University Press, 2007. xi + 241 pages.

The problem of extrapolation is of as much philosophical interest as it is of practical significance. We often cannot experiment directly on populations whose wellbeing we ultimately care about because of ethical, technological or financial constraints. When we cannot experiment directly on a population of interest and nevertheless seek experimental evidence about that population, we must experiment on a different but related population and infer what we want to know about the former from what we do know about the latter. This is Steel's core question: what are reliable rules of inference from an experimental population to a target population of interest? Steel's is the first monograph-length study of this question, and it fills an important void.

Suppose you would like to establish the efficacy of a new drug or welfare programme for an actual population such as that composed of

individuals over age 59 currently living in the UK or a hypothetical population such as that of unemployed Canadians living now and in the future. It is clear that, if for no other than practical reasons, these populations cannot be experimented on directly. If experimental evidence is sought nevertheless, it must be regarding a surrogate or *model* population such as a sample human population or a population of animal models. But how do we know whether the model is a good model for our population of interest or *target* population in the sense that inferences from what we know about the model are reliable indicators of what is true of the target? We know that the model is good when we know that the result established with its help is true of the target as well. In order to know that we need to know what is true of the target. But if we could learn what is true of the target directly we would not need a model to begin with. This is what Steel calls the *extrapolator's circle* (p. 4): to know whether a model is good, one needs to know what is true of a target; but in order to know what is true of a target, one needs to know whether a model is good.

This way put, the extrapolator's circle is very similar to what I have called the 'fundamental problem of measurement' (Reiss 2008a: 64). To establish the correct value of a property we require an accurate measurement instrument if that property is not directly observable (such as temperature or the inflation rate). To know whether the instrument is accurate, we need to know the value of the property. But in order to know the value of the property, we need to know whether the instrument is accurate. A fact that exacerbates this problem is that model and target are often known to differ in (at least potentially) causally relevant ways. Animal models are useful (when they are) precisely because they differ from humans in important ways, for instance. Steel considers two solutions to this problem that have been proposed before: simple induction and inference from knowledge about a property's causal powers or capacities (pp. 80–85). Simple induction is the rule (p. 80):

> Assume that the causal generalization true of the base [i.e., model] population also holds approximately in related [i.e., target] populations, unless there is some specific reason to think otherwise.

The formulation in terms of populations indicates that Steel has biological applications in mind when discussing the alternative inference rules. Indeed, 'relatedness' is understood as 'phylogenetic relatedness': the more recent a shared ancestor, the more related two populations are. The main problem with simple induction is that using this rule would lead to many wrong inferences because the fact that two populations are related in this way doesn't guarantee that one is a good model for the other. For instance, Fischer rats are phylogenetically closer to mice than to humans but they turn out to be a better model for the latter than for the former when predicting carcinogenicity of aflatoxin

B1. Nevertheless, Steel concludes not that simple induction is wrong or useless for extrapolation but rather that it is limited and in need of supplementation with 'some more sophisticated inferential strategy' (p. 82).

A possible strategy is to appeal to 'capacities'. Causal powers or capacities are stable causal tendencies that continue to operate, to 'try to produce their effects', even when disturbing factors are present. The paradigm example for a causal power is a physical force such as gravity. The Earth, say, continues to exert its downward pull on a body even when that body in fact rises for instance because a magnet drags it upwards. What is important is that causal powers leave traces even in situations where they do not fully produce their characteristic effects. In the magnet case gravity is noticeable because the body's upward movement is slowed down by the Earth's pull.

Stable causal powers or capacities are thus closely related to extrapolation: where they exist, they continue to contribute to an outcome even when their operation is disturbed by other factors. Therefore, learning about a causal power or capacity in an ideal experimental situation where no disturbing factors are present is informative also about situations outside the experiment because it tells us what a factor contributes to an outcome.

But causal powers don't wear their virtues on their sleeves. In particular, when a factor has been shown experimentally to have caused an effect it is not transparent whether this outcome is due to a causal power or rather due to a local causal law that is true only of the specific set up (or population) examined in the experiment. Not all factors have causal powers they can exercise independently of what other factors are present. In biology and social science, the sciences of Steel's primary interest, it is more common to find factors that are *interactive*, i.e., factors whose ability to produce an effect depends on what other factors are present.

Suppose for instance that a person swallows a poisonous substance and an antidote at the same time. It would then be wrong to say that the substance continues to be poisonous and that this causal power is manifest in the outcome. Rather, the antidote destroys the ability of the compound to poison – the two substances interact. Finally, whether causes operate additively, as required by the causal powers/capacities approach, or interactively is an empirical matter that can only be decided on a case-by-case basis. Steel therefore concludes (p. 85):[1]

[1] Steel might underestimate the resources of the capacities approach to address the problem of external validity. See for instance Cartwright (forthcoming). For an application to extrapolation in the social sciences, see Reiss 2008b. These hadn't been published at the time of Steel's writing.

> [T]he proposal that one estimate context- or population-sensitive causal relationships, and then assume that these hold approximately in related populations unless there is some evidence to the contrary, is simple induction. And ... simple induction is often not a sufficient basis for extrapolation from animal models.

The additional information that Steel thinks is required to improve the reliability of inferences from model to target concerns *mechanisms*. In biology and the social sciences, causal relations are seldom fundamental, i.e. 'brute' or 'not further analysable'. Instead, causal relations obtain on account of an underlying structure, and causes produce their effects through a series of intermediate steps. The carcinogenicity of aflatoxin B1, for instance, depends on details of the metabolism of the organism that ingests the compound. Money affects real variables such as aggregate output through the so-called transmission mechanism. Steel adopts the widely discussed Machamer–Darden–Craver definition of mechanism, according to which, 'Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions (Machamer *et al*. 2000: 3).

Comparative process tracing now proceeds by first examining the mechanism by which the cause produces the effect in the model and then comparing the stages at which the mechanism is likely to differ between model and target. In the example Steel discusses at length, the effect of aflatoxin B1 on liver cancer, significant differences are likely to be found in the metabolism because carcinogenicity often depends on metabolism. Indeed, it turned out that the metabolism of mice is much more effective in detoxifying aflatoxin than that of rats or humans. This is why rats are a better model for predicting carcinogenicity in humans than mice are (or why rats are a better model for humans than for mice).

Does comparative process tracing avoid the extrapolator's circle? At first sight, the answer is no. In order to compare the mechanisms of model and target, it seems as though one needs to know the mechanism in both model and target. Since knowledge of the mechanism by which one variable causes another entails knowledge *that* that variable causes the other it seems as though one needs to know the causal relation in the target to apply comparative process tracing.

Steel blocks this unwanted conclusion by pointing out that the method does not require comparison of the entire mechanism from start to finish but rather only of those stages that are downstream from a point where differences are likely. Suppose (this is Steel's example on p. 90) that a cause C causes an effect E through the mechanism $C \rightarrow X \rightarrow Y \rightarrow A \rightarrow Z \rightarrow B \rightarrow E$. X, Y and Z signify points at which the mechanisms are likely to differ whereas A and B are likely to be similar in model and target. Then, if upstream differences must result in downstream differences, it is

only necessary to compare the mechanisms at *Z*. This reduces the amount of information about the mechanism in the target necessary to apply comparative process tracing successfully, and thereby, argues Steel, the extrapolator's circle can be avoided.

Moreover, the aflatoxin case study demonstrates that extrapolation can be successful even when causally relevant differences between model and target are present. The main metabolic difference between rats, mice and humans concerns the level of DNA adducts in the liver. Cross-species studies have shown that the more DNA adducts in the (blood around the) liver, the more susceptible to carcinogenesis a species is. Fischer rates have the highest level of DNA adducts among the animal species studied. However, humans have much higher levels than even Fischer rats. Therefore, even though there is a causally relevant difference between this model and the target – the level of DNA adducts makes a difference to carcinogenicity – the claim that aflatoxin is carcinogenic can be extrapolated because it is likely that humans are, if anything, *more* susceptible to carcinogenesis than Fischer rats.

There is a straightforward objection to using this kind of mechanistic evidence for causal inference: even when it has been established that *C* does indeed cause *E* through some mechanism, this does not mean that increasing the level (or probability) of *C* by an intervention will result in an increase in the level (or probability) of *E*. This is because *C* and *E* may be connected through a variety of mechanisms, some affecting *E* positively, others negatively, and the overall result depends on which of these mechanisms are triggered in a given case (see for instance Elster 2006: ch. 1). In order to deal with cases such as this, Steel develops a concept of *consonance*. Roughly speaking, a mechanism set (such as the set of all mechanisms between *C* and *E*) is consonant just in case different combinations of mechanisms do not exert conflicting positive and negative influence (p. 112). To drive his case home, Steel argues that the carcinogenic effects of aflatoxin are never negative and that there appears to be no plausible mechanism by which aflatoxin might prevent liver cancer. Hence, there is evidence that the mechanisms by which aflatoxin affects liver cancer are positively consonant and that the result can be extrapolated from animal model to human target.

Whatever the usefulness of comparative process tracing for extrapolation in the biomedical sciences, readers of this journal will be more curious about applications in economics and perhaps other social sciences. Unfortunately (from the point of view of readers of this journal), the treatment of this topic in chapter 8 appears to be more of an afterthought than a core concern of the book. Steel examines two case studies: welfare reform and preference reversals. With respect to the former (pp. 161–168), Steel argues quite at length that the circumstances are likely to be unfavourable for comparative process tracing and

consonance. Information about mechanisms is less useful than in the aflatoxin example because interventions are often structure altering in the sense that the intervention changes the causal relation between the variable intervened on and the outcome variable of interest. Further, in the social sciences mechanistic knowledge is harder to come by in general, which makes it more difficult to judge similarities and differences between the mechanisms in model and target. Finally, it is likely that a new welfare programme has both positive as well as negative effects on the outcome variable (such as income).

The second case study (pp. 169–173) concerns the widely observed 'preference reversals' in which subjects have to choose between and evaluate different types of bet. A typical behaviour when the options are a high probability/low payoff bet (a P-bet) and a low probability/high pay off bet (a $-bet) with similar expected value is that subjects choose the P-bet but value the $-bet higher. These results are robust enough as to suggest they do not depend on specific laboratory settings (e.g. Guala 2005: 225). Steel now argues that the question whether the results can be extrapolated from the laboratory into the 'real world' depends on the mechanism by which they are produced. According to the scale-compatibility hypothesis, the weight of the aspect of the bet in which the response is made is enhanced. Thus, when choosing, subjects regard probability as more important; but when providing a monetary evaluation, the money value of the bet. This 'mechanism' would, if responsible for the laboratory preference reversals, have relevance outside the lab because preferences are expressed in various ways, including but not exhausted by probability and prices, and hence preference reversals should be a prevalent feature of everyday life.

By contrast, the market-price-reversal hypothesis asserts that subjects interpret questions about the prices of bets as questions about their market value. They may prefer the P-bet but nevertheless evaluate the $-bet higher because they, not unreasonably, expect some market participants to be risk-seeking and therefore willing to pay a higher price for the $-bet. However, choosing and specifying prices are strategically equivalent under laboratory but not under market conditions; if the market-price-reversal hypothesis is true, the results should have little or no relevance outside the laboratory. Steel concludes that in this case too it is mechanistic considerations that determine whether a result can be exported or not.

Having summarized the contents of Steel's book, let me now turn to critical discussion. This is the first monograph-length study of a problem that is extremely important. When, as is very frequent, there are ethical, technological or financial limitations to experimenting with the subjects of our ultimate interest – humans in general, or past and future populations, or large-scale societies, researchers must resort to surrogate systems or

populations to draw conclusions about the ultimate subjects of interest. Steel's book provides an up-to-date and in-depth discussion of the dangers and pitfalls of inferring from model to target, as well as a novel, original and thoughtful proposal of how to overcome them.

Steel does not oversell his own proposal but rather offers it as one way to approach the extrapolation problem among others, a way that has its own conditions of applicability and limitations. For instance, in order to apply comparative process tracing, we need to know a great deal about the mechanism by which the cause produces the effect in model and something about the mechanism of the target, and the stages at which they are likely to be similar and different. If comparative process tracing is to avoid the extrapolator's circle what we know about the mechanism in the target must be less than what we know about the mechanism in the model. Steel argues that only downstream differences matter. But this (as he mentions himself, see p. 90) depends on whether there are upstream differences that affect the outcome on a path that doesn't go through the downstream part of the mechanism. Whether there exists such a path is something that cannot be ruled out a priori of course, so additional empirical evidence is required. Similarly, the source or input variable (such as aflatoxin exposure) may affect the outcome through various mechanisms, and the effectiveness in a certain direction of any one of them can be undermined by the existence of another that has the opposite effect. In his extrapolation theorem (p. 113) Steel rules out such cases rather trivially by assuming positive consonance. But positive consonance once more is present or absent by way of empirical fact and belief in and action on it must therefore be substantiated by evidence.

In my view the book has two weaknesses. The first is that Steel's positive account of extrapolation is entirely based on a single case, that of aflatoxin metabolism. There are no further supporting case studies, nor is there any indication whether and to what extent this case is typical, even in the biomedical sciences (in fact, even in the context of carcinogenicity or toxicity studies). (It is again a sign of intellectual honesty that Steel mentions this in his 'Looking ahead' chapter 10 but the point seems too important to brush it aside in this way.)

The aflatoxin case is special in various ways. For instance, the compound had already been established to be carcinogenic in animals and the question was whether it is also carcinogenic in humans. In drug testing we may face a different situation, namely one in which a drug has *not* been shown to be toxic in a limited set of animal models and the question is whether it is toxic in humans. But if it is not toxic in the studied animals, there is no mechanism to extrapolate to begin with. Further, the cross-species studies in the aflatoxin case are reasonably consistent – aflatoxin causes liver cancer in all species, albeit it less so in some than in others. There is much more variability in other cases.

Aflatoxin is also special in that a variable was found – DNA adducts – that strongly correlates with carcinogenicity: high levels of DNA adducts predict high numbers of liver cancers. Biomedical researchers aren't always so lucky. Commenting on testing for toxicity of thalidomide – a sedative and hypnotic that causes terrible birth defects when taken during pregnancy – the following has been said (Manson and Wise 1993: 228):

> An unexpected finding was that the mouse and rat were resistant, the rabbit and hamster variably responsive, and certain strains of primates were sensitive to thalidomide developmental toxicity. Different strains of the same species of animals were also found to have highly variable sensitivity to thalidomide. Factors such as differences in absorption, distribution, biotransformation, and placental transfer have been ruled out as causes of the variability in species and strain sensitivity.

Moreover, it is not clear to me whether the aflatoxin case is one of *extrapolation* at all. Before trying to extrapolate any result from animal models researchers had established that there is a high correlation between exposure to aflatoxin and liver cancer. There was therefore prima facie evidence that aflatoxin causes liver cancer. Since liver cancer cannot cause exposure to aflatoxin, before concluding that aflatoxin indeed causes cancer, possible confounders – common causes of exposure and cancer – have to be ruled out. This is often a difficult endeavour in epidemiology. An alternative strategy to statistically controlling confounders (one Steel discusses himself in his 2004 publication, which is republished in revised and expanded form as chapter 9) is to investigate the possible mechanisms by which the putative cause might cause the putative effect. The mechanism of hepatocellular carcinogenesis was largely unknown for a long time (Harris 1990). In the early 1990s, however, (human) liver cancer patients have been found to show evidence of guanine-to-thymine (G $\rightarrow$ T) mutations of the tumour-suppressor gene p53 (Bressac *et al.* 1991; Hsu *et al.* 1991). What is more, these mutations are very unlikely to be caused by anything but aflatoxin: 'The mutagen aflatoxin B1 . . . which is the main food-contaminating aflatoxin species in Africa and China, binds preferentially to G residues in G+C-rich regions and induces G $\rightarrow$ T substitutions almost exclusively' (Bressac *et al.* 1991: 430; footnotes removed). The DNA adducts Steel mentions also played an important role in disentangling the effects of aflatoxin from other causes of liver cancer because new techniques that enabled the detection of aflatoxin-albumin adducts in human serum for the first time allowed to perform case-control studies at the individual level (Wild *et al.* 1993).

None of this is to deny that animal models may have played a role in all this, for instance, in suggesting possible pathways of aflatoxin carcinogenesis. But I feel that Steel still has to make the case that extrapolations from animal models played a *justificatory* – as opposed

to merely *suggestive* – role in establishing aflatoxin carcinogenicity in humans. (Steel explicitly rejects the view that animal models have mere heuristic power, see section 5.4.)

The second weakness concerns Steel's treatment of social science examples. Here Steel doesn't even try to promote comparative process tracing as a reliable method of extrapolation, for the reasons mentioned. But from a book whose subtitle is *Extrapolation in Biology and Social Science* one would have expected a somewhat more thorough discussion of the prospects for and methods of extrapolation in social science. Perhaps the arguments given against applying comparative process tracing double up as arguments against model-based reasoning in social science more generally. Or perhaps there are alternative approaches that are not subject to these arguments. Especially since reasoning in parts of the social sciences is so heavily model-driven, these issues could and should have played a more prominent role.

These slight misgivings notwithstanding, *Across the Boundaries* is a great achievement and fills a gap in the literature. One can only hope that the book spurs some interest in this important, intricate and hitherto neglected problem and that more of its kind will follow.

Julian Reiss
Erasmus University, Rotterdam

REFERENCES

Bressac, B., M. Kew, J. Wands and M. Ozturk. 1991. Selective G to T mutations of p53 gene in hepatocellular carcinoma from Southern Africa. *Nature* 350: 429–431.

Elster, J. 2006. *Explaining Social Behavior: More Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

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

Harris, C. 1990. Hepatocellular carcinogenesis: recent advances and speculations. *Cancer Cells* 2: 146–148.

Hsu, I., R. Metcalf, T. Sun, J. Welsh, N. Wang and C. Harris. 1991. Mutational hotspot in the p53 gene in human hepatocellular carcinomas. *Nature* 350: 427–428.

Machamer, P., L. Darden and C. Craver 2000. Thinking about mechanisms. *Philosophy of Science* 67: 1–26.

Manson, J. and D. Wise 1993. Teratogens. In *Casarett and Doull's Toxicology*, 4th edition, ed. Mary Amdur. Oxford: Pergamon.

Reiss, J. 2008a. *Error in Economics: Towards a More Evidence-Based Methodology*. London: Routledge

Reiss, J. 2008b. Social capacities. In *Nancy Cartwright's Philosophy of Science*, ed. S. Hartmann and L. Bovens, 265–288. London: Routledge.

Wild, C., L. Jansen, L. Cova and R. Montesano 1993. Molecular dosimetry of aflatoxin exposure: contribution to understanding the multifactorial etiopathogenesis of primary hepatocellular carcinoma with particular reference to hepatitis B virus. *Environmental Health Perspectives* 99: 115–122.