Economics and Philosophy, **20** (2004) 217–246

REVIEWS

DOI: 10.1017/S0266267104211397

Fact and Fiction in Economics: Models, Realism and Social Construction, edited by USKALI MÄKI. Cambridge University Press, 2002, vii + 384 pages.

In 1997 the conference 'Fact or Fiction?' celebrated Uskali Mäki's inauguration as Chair of Philosophy at Erasmus University, Rotterdam. The papers presented at the conference have now been published in the volume under review. Some of them appear in print for the first time, others have already been published in journals, but the volume presents the added value of having them organised in coherent groups – recreating at least in some cases the dialectical atmosphere of the conference from which they originate.

The most successful example of this can be found in Part III of the volume, devoted to economic modelling. Modelling has been one of the hottest topics in the methodology of economics during the last decade, and here we have five top-rate contributions by some of the leading scholars in the field. Bob Sugden's 'Credible worlds: The status of theoretical models in economics' is one of my favourite papers, as well as a pedagogic masterpiece. Sugden illustrates and discusses two well-known microeconomic models, George Akerlof's 'market for lemons' and Thomas Schelling's checkerboard model of segregation. Using them as reference points, he reviews critically some of the major views on the connection between theoretical models and the real world, including the views that models are tools for conceptual exploration, that they are instruments for prediction, that they are metaphors, and that they are caricatures; he finishes with a critical discussion of the method of isolation and of Mill's inexact deductive method. Sugden argues that none of these provides the right characterization of the link between models and reality, a gap that is bridged by a process of inductive inference. A model is a special entity somehow representative of a wider class of entities, a class that is supposed to include real economies among its elements. When engaged in theoretical modelling, economists ask their audience to believe that some of the properties of the model are shared with some of the real economies. According to Sugden, the crucial feature of a good model (one that supports this sort of generalization) is that it must portray a *credible counterfactual world*, a world that could have been true even though we all know that it does not, in fact, exist. Like a good novel a good model must be inhabited by credible characters, characters that are like real people, who perform credible actions in a credible environment.

One of the striking aspects of Sugden's account is that it ignores entirely issues of empirical testing. But here Sugden is concerned primarily with semantic rather than epistemic issues. The credibility of a model is, then, the epistemic consequence of its being related in the 'right' counterfactual way with the real world. But credibility, I take it, is not supposed to provide a surrogate proof of the empirical validity of the model. The next two papers in this section tackle the issue of empirical testing more directly. Nancy Cartwright briefly outlines a view she has defended extensively in her 1999 book, *The Dappled World*. She starts from an ontological analysis of the origins of the regularities measured by empirical scientists. Regularities are not fundamental, but result from the repeated, unimpeded work of causal structures or, as she calls them, 'nomological machines'. Models then typically do not describe regularities: they are 'blueprints' of the nomological machines. All this is fairly plausible, given the difficulty of providing a purely empiricist interpretation of most theoretical models in physics and economics (an interpretation, that is, of the relations in the models as representative of regularities between actual events). The controversial aspect of her proposal comes next. Cartwright argues that much of the empirical-statistical work of estimation performed in science makes sense only under the assumption that a data-set has been generated by a unique and stable nomological machine. Only under such an assumption, in fact, are we entitled to apply standard statistical inferential techniques to estimate the underlying probabilistic relations between the variables in our models. Given that the uniqueness and stability assumptions are in several cases very implausible in non-experimental contexts, the argument constitutes a radical challenge to empirical statistics in the natural and social sciences. Cartwright backs up the abstract argument by discussing three models, one from the physical theory of superconductivity and two from economics.

The debate between Nancy Cartwright and Kevin Hoover was one of the highlights of the Rotterdam conference, and I was glad to see that the debate has lost little vigour and depth in print. Hoover does something remarkable in his paper: he criticizes a philosophical argument by means of a piece of social science. Using statistical data on job vacancies and unemployment, he shows how one of the models used by Cartwright (Chris Pissarides' model of skill loss during unemployment) can be put to work for estimation purposes. The exercise aims to show how econometrics practice can proceed on much weaker assumptions than those deemed necessary by Cartwright. The existence of a stable nomological machine such as that represented in the theoretical model, in particular, is not a prerequisite for statistical investigation because much econometrics is about discovering robust regularities rather than measuring causal structures. Models then, according to Hoover, are not blueprints of nomological machines, but rather describe *possible* mechanisms that *may* give rise to the rough regularities that econometricians are concerned with. In fact Pissarides does not invite us to estimate his 'nomological machine', but rather a much more generic model that is only vaguely compatible with it.

Hoover's point once again raises the issue of the relation between models and the real world. All the contributors with an economics background tend to stress the loose, informal nature of the relation between theoretical models, empirical models, and reality. Mary Morgan argues that informal stories play a key role in linking these different levels. She builds on previous accounts by Gibbard and Varian and by McCloskey, trying to identify more precisely where and when stories help in putting models to work. The key idea is that models are used to answer questions, and informal stories are essential in answering them. They are used first to trigger and manipulate the model and then to recover the complexity of the real world that is partly lost in the (necessarily simplified) structure of the model. The mathematical structure then works as a resource (by highlighting the 'essential' elements needed to answer the question) and as a constraint (because it limits the sort of things that you can do with a model, the results you can obtain from it).

Roger Backhouse sums up by commenting on the above papers. He unpacks the notion of 'story' into four separate elements that he thinks are conflated in Morgan's account and backs up Hoover's argument by suggesting that theoretical models are used primarily to prove the coherence and hence the possible existence of a mechanism that might conceivably underlie some observable empirical phenomena. He argues that it is dangerous to impose too strict requirements on the methodology of economics (*à la* Cartwright), given that most economic practice relies on looser informal methods. He recognizes, however, that we are still a long way from a satisfactory account of these informal procedures, an account able to recover their rationality and probably also their hidden systematicity. This is an exciting research agenda for economic methodology in the twenty-first century.

I have spent more time on these papers because they constitute the most coherent part of the book. Indeed, I predict that they will all be included in the syllabus of most graduate courses in economic methodology. Two other papers – by Mark Blaug and Partha Dasgupta – are also matched so as to constitute a 'virtual' dialogue. In recent years

Mark Blaug has become increasingly critical of the direction taken by contemporary economics. In his chapter he identifies three 'ugly currents' within it: general equilibrium theory, game theory, and the New Classical macroeconomics (with its most recent offspring, real business cycles theory). All three currents share the same defects, according to Blaug: the cult of formal modelling pursued for its own sake, the lack of interest for any real world (policy) issue that cannot be tackled mathematically, and a general contempt for grounding theories inductively on empirical facts. Blaug also identifies some counterparts of the formalist disease at the methodological level, in particular the varieties of postmodernism that downplay the importance of empirical evidence, and the approaches (like Critical Realism) that deny the possibility of prediction in economics. All of them, he suggests, encourage economists to persist in their complacent attitude and to avoid the confrontation with real world issues and empirical data.

Like Blaug, Partha Dasgupta is a respected economist who can look back at the recent history of the discipline from the vantage point of his own experience. Unlike Blaug, however, he has not been actively involved in methodological theorizing and, more importantly, has a completely opposite outlook on the recent developments of economic science. Dasgupta defends economics as currently practised from some of the most common critiques that are levelled against it, with vigour and no lack of concrete examples. I said that his dialogue with Blaug is 'virtual' in character because his essay is effectively a long review of Robert Heilbroner's recent work and covers much more ground than the worries raised in Blaug's chapter. Formalism and the alleged detachment from real-world problems are discussed by Dasgupta, who also tackles other key issues like the relation between ethics and economics, the programme of microfoundations, the state–market dichotomy and a lot more. The 'marriage' between the Blaug and the Dasgupta papers may well have been 'arranged' by the editor, but works quite well nevertheless.

The other papers are broadly connected by the overarching theme of realism in economics. The only paper *directly* concerned with realism is Uskali Mäki's chapter, a systematic critique of (counter-) arguments purporting to show that one *cannot* be a realist about economics. Mäki does a good job of showing that realism is a viable position, i.e. that it is entirely reasonable to believe that an entity X *might* exist or that a theory T *might* be true. The argument, however, seems quite lame if compared with most realist arguments in the philosophical literature. Such arguments typically try to demonstrate that one *ought* (rationally) to be a realist about certain entities or theories (think about no-miracles arguments, for example). Perhaps in the context of economic methodology it is still worth defending the very *possibility* (rather than rationality) of a generic realist philosophy of economics against the skeptics' challenges, but at the end of this chapter-appetizer one is left hungry, waiting for the main meal. Hopefully Mäki will provide us soon with his own defence of a well-defined, substantial realist position.

Deirdre McCloskey also defends a very broad notion of realism – so broad as to make it vacuous and intentionally so. The idea is that in order to participate in any discourse (scientific, political, philosophical, or whatever), one needs to make a commitment to the existence of certain entities or the truth of certain claims. However, according to McCloskey, the commitment is not ultimately based on epistemic but on moral foundations. It is based on the decision to participate in a given discourse, to accept (tentatively, perhaps) the rules that govern speech and conversation within a given community. Overall, I found this chapter rather disappointing, not for the thesis it defends (I happen to agree, for example, that philosophy and large parts of economics are basically normative enterprises) but for its style. The arguments are sketchy and rely heavily on the bashing of an unlikely straw man – the 'analytical philosopher', who for some reason also happens to be a realist (in McCloskey's mind). I might not belong to her intended audience, but as far as I'm concerned this time McCloskey has used the wrong rhetoric, I'm afraid.

There is just enough space left to summarize the contents of the remaining chapters. Wade Hands reviews a well-known debate in science studies (the so-called 'chicken controversy') and uses this example to show how social constructivism has injected new life into old debates in the social sciences such as the 'individualism *vs*. holism' controversy. Raimo Tuomela and Wolfgang Balzer propose an analysis of social ontology along the lines of those defended in more or less recent times by Barry Barnes, Margaret Gilbert, and John Searle. Ilkka Niiniluoto shows how idealized models do not constitute an insurmountable problem for the realist, once they are understood as counterfactual conditionals. Philip Pettit's paper features one of several recent outlines of his own interpretation of rational choice explanations as describing 'virtual' dispositions. Such dispositions explain the resilience of a phenomenon by identifying the factors that *would* restore the phenomenon should some force try to disrupt it. Virtual factors like egoism may be inactive in normal circumstances but are ready to be triggered when, for example, persisting in routine or norm-driven behaviour would go blatantly against an agent's self-interest. In his paper Pettit shows how, under this interpretation, rational choice and functional explanations in evolutionary biology and the social sciences turn out to make use of very similar explanatory strategies. Shaun Hargreaves-Heap puts forward a plea for introducing more explicitly hermeneutic elements in economic theory and argues that this should not be at the expense of a realist attitude regarding economic theory itself. Indeed, not only the hermeneutic elements (especially the 'common cultures' that underpin individuals' interpretation and evaluation of their own actions) should be read realistically, but the most successful applications of standard rational choice models also turn out to be vindicated by the hermeneutic approach. Bruce Caldwell and Neil de Marchi contribute the two most distinctively historical chapters, devoted respectively to Friedrich Hayek and John Stuart Mill. Caldwell focuses on Hayek's theory of cultural evolution, in particular on its relation to his simultaneous endorsement of methodological individualism. De Marchi reconstructs the intellectual genesis of Mill's famous 1836 methodological essay and tries to explain its strong emphasis on the *a priori* deductive method. In order to provide strong foundations and convincing power to his own programme of social reform, Mill was looking for a method providing the same certainty as that achieved in the realm of natural science. *A priorism* was the chosen solution, after a long struggle with the problem of induction and harsh disillusionment with the historical method.

Finally, let me say a few words on Jesús Zamora-Bonilla's exercise in the economics of science, a field that seems to be undergoing a second renaissance these days. The paper starts from the assumption that scientists (and hence economists) aim neither at epistemic nor at wealth-maximizing goals but at personal recognition. Since there does not seem to exist a 'market' where scientists directly exchange recognition, Zamora-Bonilla conjectures that scientists reach an agreement on 'constitutional' (read: methodological) norms regulating the collective recognition of research. This opens fairly standard methodological questions: are the rules 'good'? Can they serve general epistemic as well as selfish individualistic goals? Unfortunately, there is no guarantee that the constitution will produce epistemic efficiency, as problems of path-dependency, multiple equilibria and inefficiency are pervasive. As Zamora-Bonilla recognizes, these epistemic failures seem to be the direct consequences of recognition-oriented motivation, a result that does not strike the layman as particularly surprising. (The opposite result, that non-epistemic motives generate epistemically optimal outcomes, would surely be more surprising, *à la* Adam Smith.) The interesting point is how to make sure that the negative effects of selfish motivation are neutralized or, to put it another way, how to make sure that scientists 'reveal' their true epistemic opinions. Zamora-Bonilla suggests that coalition-formation at the level of theory-choice may allow scientists to 'jump' to a better equilibrium when they are trapped in inefficient equilibria by following the constitutional rules. Translated out of jargon, this is the decision by some scientists to accept a theory despite its *prima facie* inferiority in light of commonly accepted methodological standards. There are several cases in the history of science of theories endowed with such a power to defeat even the most entrenched methodological standards (quantum mechanics and the abandonment of determinism is such a case; Newtonianism and the abandonment of action-at-a-contact is another). The economics of science offers a fresh perspective on a phenomenon that seems to play an important role in the interaction between scientific and meta-scientific theory choice.

At the end of this *tour de force*, let me sum up the main virtues of *Fact and Fiction*. Every ten years or so, an edited volume comes out that offers an overview of the best work done in the discipline in recent times. Despite the fact that this may have been neither its original nor its primary goal, the Mäki volume does this rather brilliantly. If we compare it with, say, Roger Backhouse's (1994) *New Directions* volume, we can appreciate that economic methodology has come a long way since the early nineties. Some debates, like the Lakatosian controversy, have almost completely died of exhaustion. The partnership between methodology and history of economics has also suffered during the last decade, yet the level of empirical detail and the richness of case studies in methodological writings has increased rather than diminished. The areas of modelling, causation and econometric inference have been flourishing, and new projects like the ontology of economics and the economics of science seem to be coming back with great impetus. For someone heavily involved in research in economic methodology (as I am), *Fact and Fiction in Economics* carries good news: the discipline seems to be alive and well, with plenty of stimulating things going on.

Francesco Guala

University of Exeter

REFERENCES

Backhouse, R. (ed.). 1994. *New Directions in Economic Methodology*. Routledge

DOI: 10.1017/S0266267104221393

Causality in Macroeconomics, by Kevin D. Hoover. Cambridge University Press, 2002, xiii + 311 pages.

Jung claimed that the three basic needs of mankind are food, sex and causality. Philosophers have anguished over causality for centuries. Natural scientists tend to use the word sparingly, although it can be argued that some notion of causality underlies most of their efforts to understand the physical and biological world. Epidemiologists have given careful attention to the distinction between risk factors and causes, requiring the isolation of the latter to justify public health recommendations. Bradford Hill's (1965) guidelines on when a risk factor is relatively likely to be causal are often appealed to. There are broadly parallel discussions in empirical sociology. Statisticians, perhaps characteristically, have been cautious about causality, at least until relatively recently. They note, following Karl Pearson and Yule, that correlation is not causation and, following R. A. Fisher, that perhaps randomized trials may be a secure base for causal inference but that single sets of observational data and single studies or even replicated studies of similar structure are rarely so unless very large effects are involved. R. A. Fisher's dictum that to make observational studies more nearly causal one should *make one's theories elaborate* is relevant. More recently, however, ideas from the theory of randomized experiments have been adapted, especially by Rubin, to clarify the circumstances and assumptions under which, in some sense, causal conclusions may follow from observational material; fruitful applications have been in social research and public policy issues. Computer scientists have developed causal networks in connection with Bayesian knowledge or belief systems; the recent book by Pearl (2000) is particularly noteworthy, as also are developments from the Department of Philosophy at Carnegie Mellon by Glymour, Spirtes and Scheines.

The present book sets out in detail a view of causality directed at macroeconomics. The book introduces the discussion via the philosophical and economic ideas of David Hume, set in a modern context. The next three chapters develop the notion of causal structure to which the author attaches key importance. Chapter 5 defends the view that macroeconomics is to be regarded as standing on its own, i.e. does not have microeconomics as its fundamental basis. Chapter 6 argues in particular that because of the special character of macroeconomics, temporal ordering cannot be taken as a crucial element in the discussion. Chapter 7 develops the special character of causality in macroeconomics and chapter 8 concerns how to determine the direction of causality. There are then two case studies, one about the direction of the relation between taxes and government spending and the second on the causal direction between money and prices. There is a brief summary of the argument in a final chapter.

The book is a careful and deeply scholarly synthesis of the economic and philosophical views of this topic. I profited from and enjoyed reading the book but must confess that I was somewhat bemused by the end. The review editor of this journal probably did the author no kindness by choosing a statistical reviewer, likely therefore to have a rather pragmatic approach to these issues. Did the book give me helpful new guidance on what sort of data to collect and what sort of analyses to do in order to achieve conclusions that are closer to being causal? Perhaps I should have seen such implications, but I did not. It might have been easier for the reader if the author had separated out more strongly his own view of how to proceed from critiques of other approaches, interesting though those discussions are.

One somewhat surprising aspect is the virtual omission of any discussion of confounders, especially unobserved confounders. In epidemiology, absence of these is regarded, surely correctly, as crucial to any justification of a passage from risk factors, i.e. empirical associations, to causal interpretation, unless the effect involved is a very large one. Yet the discussion here, both of Suppes's probabilistic causality and of the associated time series notion of Granger causality, makes little mention, so far as I can see, of possible alternative explanations by observed or unobserved features. True, in the initial definition of Granger causality a background causal field is mentioned, but in applications it is suggested that only two time series may be involved and that it is enough to show that the potential cause helps predict the outcome series given the history of the latter. To consider this to be a weak basis for a claim of causality is in no way to deny the usefulness of Granger's concept, only to be uneasy at possible over-interpretation. In the author's scheme of causal structure, what assurance is there that an omitted variable might not change the whole interpretation?

It is presumably assumed that all important variables are entered into the representation of the causal process being exploited, but does this not to some extent beg the question?

Some of the philosophical discussion is concerned with the precise meaning to be attached to the word *cause* and, to the extent that this clarifies the different usages by different workers, the discussion is helpful. On the other hand, some of the points raised for discussion seem somewhat contrived, to put it gently. There is considerable discussion of examples along the lines of one example due to I. J. Good: Watson, attempting to save Sherlock Holmes from certain death at the hands of Dr Moriarty, pushes a boulder at the latter but misses and knocks Mr Holmes to his death. Did Watson cause the death of Sherlock Holmes? Does this example show more than that in a relatively complex network of outcomes and potential outcomes, it may be essential to specify a whole process to achieve reasonable understanding? Describing bits in isolation may mislead.

In some ways a central issue is the establishment of the direction of causality and the role of time. A conventional view in many fields is that causality acts forward in time and not instantaneously. The author rejects this for macroeconomics, essentially on the grounds that macroeconomic data are so aggregated in time that possible effective simultaneity of cause and effect must be accepted. An alternative view would be, following Fisher (1970), that while causality works forward in time via stochastic difference equations (or better, stochastic differential equations) all that can be observed is either a long run equilibrium or time aggregates; of course, this does not resolve the central issue, trying to determine the direction of causality. The author's solution is to look, following massive changes in the system (what Bradford Hill termed *natural experiments*), for stability in components of the two possible factorizations of a joint probability distribution of two variables into marginal and conditional distributions.

The essence of the author's view is that causality is about structure. This is, I think, not far from one view of causality implicit in the natural sciences. That is, causality is about the inevitably provisional explanation of a phenomenon in terms of some underlying process. In other fields this underlying process may be explicitly justified by empirical data of a different kind, for example the explanation of a physiological process at cellular level. It may be in terms of well-established theory, such as the laws of classical physics used to 'explain' the strange behavior of flames, or it may be purely hypothetical. All three have their value, although clearly the third is the least satisfactory. Note that the development of macroeconomics as an outgrowth of microeconomics would correspond to the first of these routes and that the author argues strongly against this.

The remarks above deal with only some of the issues raised in the book. It contains a great deal of thoughtful and wide-ranging discussion of important topics.

D. R. Cox

Nuffield College, Oxford

REFERENCES

Bradford Hill, A. 1965. The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58: 295–300

Fisher, F. M. 1970. A correspondence principle for simultaneous equation models. *Econometrica*, 38: 73–92

Pearl, J. 2000. *Causality*. Cambridge University Press

DOI: 10.1017/S026626710423139X

The Methodology of Empirical Macroeconomics by KEVIN D. HOOVER. Cambridge University Press 2001, xii + 186 pages.

In this volume, Kevin Hoover displays some fruits of his enduring engagement with macroeconomics and its econometrics, methodology and philosophy of science. He develops a distinctive approach to methodological questions by trying to occupy a middle ground between the (nearly) pure prescriptivism of a large part of philosophy of science and the (nearly) pure descriptivism of much recent historicism and postmodernism. The intermediate position starts from individual case studies that are regarded as typical and significant. These cases are discussed and analysed on the basis of considerations of economic relevance alone, and they are used to introduce and illustrate methodological problems. Solutions to these problems, in turn, are sought from many different angles, including history and philosophy of science, economic methodology, economic history and contemporary practice. Thus Hoover shares with the prescriptivists a concern for general methodological issues but rejects their top-down approach; he shares with the descriptivists a detailed knowledge of the subject matter at hand and its troubles but rejects their particularism, that is, their aversion to make inferences beyond a single case. All in all, he assumes an explicitly pragmatist stance: good method is whatever is useful for the purpose at hand and what is useful is to be determined by the practitioners themselves.

Hoover discusses a great number of methodological issues including the relationship between models and stylised facts, the relationship between theoretical and empirical variables, idealisation in economics, the notion of empirical success, the possibility of micro foundations, how micro relates to macro and causality in macroeconomics. The book is too rich to even discuss a fair selection of these. In particular, I leave out Hoover's discussion of causality, because his other new book (2001), which examines the topic in much more detail, is also reviewed in *Economics and Philosophy* (this issue). I will thus focus on two aspects: the relationship between economic models and laws and the relationship between micro and macro entities.

THE RELATIONSHIP BETWEEN ECONOMIC MODELS AND LAWS

The concept of scientific law has been at the heart of the philosophical debate about science in the larger part of the twentieth century. The reasons are straightforward: laws codify our scientific knowledge; laws help us to predict, explain and control phenomena; laws mark what is general and of scientific interest from what is particular and accidental; and, last but not least, laws present an irresistible philosophical puzzle.

Viewed in this light, it seems surprising that the notion of law plays a relatively insignificant role in economics. True, there are claims that bear the name of a law: the law of supply and demand, Gresham's law, the iron law of wages, Say's law and so on. But in the overwhelming part of both theoretical as well as applied work one does not find much talk about laws. While in other disciplines knowledge seems to be codified in terms of laws, in economics knowledge is codified in terms of models and what economists call 'stylised facts', that is, 'rough-and-ready empirical generalisations'.

Thus, two questions arise. First, how do models and stylised facts relate to each other? Second, how are the functions that laws fulfill in other disciplines accomplished in economics? Hoover's response to the first question criticises Nancy Cartwright's notion of a model as a 'blueprint for a nomological machine' (Cartwright 1997a). Cartwright is famous for rejecting the Humean and logical positivist image of laws. She replaces it with an image in which properties with causal capacities are basic and laws arise only in very special circumstances. Such circumstances, stable arrangements of factors with stable capacities in a shielded environment, Cartwright calls *nomological machines* (Cartwright 1997b). Models represent nomological machines.

Hoover disputes that typical theoretical models in economics represent nomological machines. In order to show where he disagrees, Hoover takes up an exemplary model discussed by Cartwright. The model is constructed by Christopher Pissarides (1992) to explain persistence in the level of unemployment or 'hysteresis'. He hypothesises that this phenomenon is due to a thin market externality created by the fact that workers lose skills when unemployed. The intuitive idea is this. Firms supply jobs and are randomly matched with work-seeking employees. An external negative employment shock lowers the average level of skill of the pool. Since the expected profit from less-skilled workers is lower, firms offer fewer jobs. But in so doing, they prolong periods of unemployment which further reduces employees' skills. Hysteresis arises.

One aim of Pissarides' paper is to formalise this intuitive idea. He presents us an overlapping generations model with equal cohorts of identical agents whose life span is two periods. Workers and firms are matched in each period at random. They must agree to either cooperate to produce in that period or to do without production (unemployed workers produce no output), and they enter into a Nash bargain, splitting the marginal return to the worker's production equally between them.

So why does Pissarides' model not represent a nomological machine? While possibly correct in some cases of physics models, Cartwright's claim does not apply to economics. The difference between physics and economics models lies in their relationship with reality. Many models of physics, according to Hoover, represent real systems with real properties. He discusses the case of the harmonic oscillator. One can, for example, take the model of the harmonic oscillator and use it as a blueprint to build a real one. But, by and large, one cannot use an economics model to build a real economic system. Economic models, taken part by part, do not – directly – represent reality. This is mainly due to the extensive use of certain kinds of idealising assumptions. What, in the economy, is the referent of two equal cohorts of identical agents living two periods? And who determines salaries by a Nash bargain? What is the basis for treating matches as random?

To understand the nature of these idealisations, Hoover invokes ideas developed by Leszek Nowak (1980). Nowak thinks that economic outcomes are the result of an interplay between so-called primary factors (or factors of interest) and secondary factors (or disturbing factors). The aim of idealisation is to isolate the operation of the primary factors by setting the secondary factors to extreme values. Unfortunately, as Hoover notices, there is a problem with this suggestion. The assumptions of the hysteresis model are not quite of this kind. Assuming, for instance, that workers' lives last two periods is not like assuming away air resistance: 'The length of a worker's life is a material variable in Pissarides' optimization problem' (p. 38). That is, in making this assumption, Pissarides does not set a secondary factor to an extreme value in order to isolate a primary factor.

What, then, is really going on? Hoover points out that only *parts* of an economic model mimic *parts* of real systems purely qualitatively, and even these parts relate only vaguely. For example, Pissarides claims that an implication of his model is that a plot of the unemployment and the vacancy rate should make counter-clockwise loops over the business cycle, while the overall relationship between them should be an inverse one. But neither unemployment nor vacancies are even mentioned in his model. That is, concepts in the theoretical model have to be *interpreted* in order to apply to data.

The first step in interpreting a theoretical model is to build a second, empirical model. Pissarides develops two empirical equations, one that relates vacancies to variables that influence demand for labour, number of job seekers and the duration structure of unemployment and one that relates the matching rate to vacancies, number of job seekers, intensity of search and mismatch (discussed on pp. 10–11). These concepts are regarded as measurable.

The second step is to interpret results of the theoretical model in the light of empirical results. Hoover writes: 'The relationship between the model and the empirical equations, even after they are estimated, is analogical. It requires an imaginative leap to draw conclusions about the model from the empirical questions or about the equations from the model' (p. 26). It is important to point out that there is no principled way for the imaginative leap. In fact, the analogy may be so sloppy that the causal order implied by the two kinds of model differs (which is the case here: cf. pp. 112–13).

This brings us to the second question regarding the functions traditionally fulfilled by laws. I take it that the robust stylised facts at hand (e.g. the counter-clockwise loops and the negative relationship between unemployment and vacancies) were known prior to Pissarides' model. Thus, if these are the model's only implications of interest, its purpose cannot be the *prediction* of novel facts. It may still *explain* known facts. But it is unclear how it does that job. The covering-law model of explanation will not work in this case because no statements about empirical facts are deduced. Hoover himself rejects the covering-law model, if for different reasons (pp. 19ff.). Causal models will not work either because of the nature of the idealisations involved. We have also seen that the causal order differs between theoretical and empirical models. Nor will unification models work because they, too, require a deductive relationship between explanandum and explanans. Further, I cannot see a way in which it would allow us to *control* economic quantities. Whatever the model does, hence, it does not play the role traditionally played by laws.

Hoover therefore correctly rejects 'law talk' (p. 54). But the problem is that in traditional philosophy of science where laws mattered, there were viable accounts of how scientific practice (e.g. the search for and use of scientific laws) connected with the aims of science (e.g. prediction, explanation and control). Hoover is right to try to let economists speak themselves for what they regard as the aims of their discipline. I do not see that they have given up on the more traditional aims. Hoover literally writes that Pissarides seeks to *understand* why shocks to unemployment persist for long periods (p. 7). That economists also seek *prediction* is clear from his discussion of empirical success (e.g. p. 39) and that *control* matters is clear from his discussion of causality (ch. 4). I do not understand what a model such as Pissarides' does to further any of these aims. Thus, while it may be right to say that it does not represent a nomological machine, it remains unclear what else it does – in relation to the traditional aims.

THE RELATIONSHIP BETWEEN MICRO AND MACRO

The Lucas critique can be interpreted as making two claims. First, the causal parameters estimated using the econometric methods of the structural equations approach do not represent 'deep parameters', i.e. parameters that are stable under intervention. A corollary is that policy recommendations that are based on the models of this approach are likely to generate at best unforeseen but more probably adverse results. Second, only models based on individual agents' optimisation problems are likely to represent deep parameters.

Clearly, these two claims are independent. It is conceptually possible that purely macro models contain deep or stable parameters while models based on individual optimisation fail to do so. Many economists, however, have understood the two theses as the two sides of the same coin and sought micro foundations as a consequence. Hoover does understand them as independent; he accepts the first but rejects the second thesis. About causality I will not talk here; let us focus instead on Hoover's rejection of micro foundations.

For Hoover, the burden of proof should be on the side of the advocates of micro foundations. This is because of what he calls the 'Cournot problem'. Essentially, the problem is that there are too many causal factors interacting at the level of individuals to derive the aggregate result by summing up individual influences. So he demands that the proponent of micro foundations show that he or she can solve the problem of aggregation. By demonstrating that no aggregation strategy works, Hoover has a good argument against the demand for micro foundations.

Hoover finds two aggregation strategies in economics, aggregation by summation and aggregation by analogy. The former constructs a macro entity by summing over micro entities; the latter treats the macro entity *as if* it were a micro entity.

What are the conditions for treating the mereological sum of things as one thing? Let us consider a few examples. If all coins in my pocket are valid Euro coins, I can treat them as if they were a single coin when I pay my restaurant bill in Paris. It does not matter whether the vendor on Portobello Road market uses two one-pound weights on the pan to balance apples or one two-pound weight. For the purpose of building a solid wall, many small bricks or few big bricks do the same job (in some respects). A necessary condition for treating a sum as a single object for some uses seems to be that they share a property that is essential to the purpose at hand – be it 'price', 'weight' or 'length and solidity'. For economics, the good news is that all commodities share the property 'price'. The bad news is that its unit is currency/good, so one can aggregate only similar goods, and most goods are not very similar.

Now, this is a problem for the construction of price indices. More difficulties appear when one is trying to construct composite *goods* such as the GDP. Here, whether one can regard the aggregate as a single good depends, among other things, on whether relative prices are stable, whether tastes are identical and what the respective utility functions are, and on the income distribution. The point is that what Hoover calls 'perfect aggregation' is possible only under most stringent conditions that are practically never fulfilled for a real economy.

The second strategy is aggregation by analogy. Representative-agent models treat the macro economy as if it were an individual making optimal choices. But there is no real agent 'who maximises a utility function that represents the whole economy subject to a budget constraint that takes GDP as its limiting quantity' (p. 83). Thus, in fact, these models do not take micro foundations seriously; they use micro mathematics but macro concepts. One may regard these models as 'as if' models in Milton Friedman's sense. But then, of course, it is only empirical success that matters. Nothing could or should be sacrosanct in model construction.

Prospects for the feasibility of economic policy look pretty dismal, then, if one takes the Lucas critique to be both valid and involving the two theses simultaneously: policy feasibility depends on 'deep parameters', but to get deep parameters one needs micro foundations; however, there is no way to get from micro to macro. Hoover's way out of this dilemma is to accept the reality of macro economic entities in their own right. Macro aggregates such as the GDP, the price level and the federal funds rate may be as real as any quantity and there may be causal relations among them. To be sure, causal relations can differ in their degree of stability. Some may be altogether fragile, others relatively stable but not autonomous and still others will be autonomous. But this is a matter of causal inference and not of micro modelling. Hoover calls macro entities real, but he presents no argument to that effect in the present volume (for that, see his 2001 book). He does, however, tell us something about the relation between micro and macro, which he believes to be one of supervenience. Hoover explains that 'if two parallel worlds possessed exactly the same configuration of microeconomic or individual economic elements, they would also possess exactly the same configuration of macroeconomic elements. It is not the case, however, that the same configuration of macroeconomic elements implies the same configuration of microeconomic elements' (2001: 120).

Now, this is odd. It is odd because the concept of supervenience has usually been used by proponents of reductionism in the face of difficulties with more fully blown reductivist strategies such as type-type identity (which entails that types of macro states are *identical* to types of micro states). But Hoover opposes reductionism. So why does he believe that macro entities supervene on micro entities? Part of the story is that there is a kind of ontological dependence: one could not have macro without micro. But this is a trivial claim. No one (except maybe some Platonists) believes that one could have ethical, mental, aesthetic or whatever properties people have thought of as supervening without having the corresponding lower-level properties as well. I do not see how the macro could supervene on the micro in any stronger sense than this.

One point is that, clearly, macro entities causally influence micro entities (when, for example, agents react to inflation or recessions or changes in the federal funds rate). This contradicts both the spirit and the letter of supervenience theories. The second point is that it is not clear whether one would really 'fix' the macro entities by 'fixing' the micro entities. Imagine we could duplicate all micro entities of some economy. Would that duplicate the macro economy, too? Not necessarily. This is due to the relative liberty with which macro aggregates are constructed. There is no one way in which, say, the price level can be measured. Importantly, different ways of measuring have different, and in some cases very different, results. These results, in turn, may have effects that spread throughout the economy.

Consider the controversy about measuring the cost of living (see, for example, the various papers in the *Journal of Economic Perspectives*, Winter (1998), 12:1 and Gordon 2000). In 1996, the Boskin commission reported that the CPI overstates the 'true' cost of living by about 1.1 per cent due to various biases. The Boskin report received much criticism in the aftermath. Much of that criticism was, however, directed at the social and political implications of the report because many payments such as taxes and social security benefits are CPI indexed. Hence, had the cost of living been measured 'correctly' in the past, many payments would have been very different, which in turn would have influenced all sorts of individual economic decisions.

Therefore, even if one assumes strict determinism at the micro level, copying a micro economy would not ensure that the history of the duplicate would be identical to the history of the original. In order to ensure that, one would also have to fix the methods of measurement, but this is exactly the 'additional bit' macro entities have and whose existence supervenience theorists deny.

The irony is that the rejection of supervenience helps Hoover's overall project. He want macro aggregates to be real entities with real causal relations among them. Finding that there are two conceptually distinct and causally interacting levels surely does not make that claim more difficult to argue for.

Hoover's book is an important contribution to economic methodology. He raises a great number of methodological issues, only a fraction of which could be discussed here. His proposed solutions to problems are highly innovative without exception. Every serious student of economic methodology should read this book.

Julian Reiss

London School of Economics

REFERENCES

Cartwright, N. 1997a. Models: The blueprints for laws. *Philosophy of Science*, (Supplement) 64:S292–S303

Cartwright, N. 1997b. Where do laws of nature come from? *Dialectica*, 51:65–78

Gordon, R. 2000. The Boskin Commission Report and its aftermath. *NBER Working Paper* No. w7759

Hoover, K. D. 2001. *Causality in Macroeconomics*. Cambridge University Press

Nowak, L. 1980. *The Structure of Idealization: Towards a Systematic Interpretation of the Marxian Idea of Science.* Reidel

Pissarides, C. 1993. Loss of skill during unemployment and the persistence of unemployment shocks. *Quarterly Journal of Economics*, 107:1371–91

DOI: 10.1017/S0266267104241396

Reflection without Rules: Economic Methodology and Contemporary Science Theory, by WADE HANDS. Cambridge University Press 2001, xi + 480 pages.

In the book that did more than any other to promote the emergence of economic methodology as a field of study, Mark Blaug (1980) provided a section entitled 'What you always wanted to know about the philosophy of science but were afraid to ask.' His target was economists interested in methodology but lacking the relevant philosophical background. Nearly a quarter of a century later, Hands is aiming at the same audience – people who, if they do not remember the economics, can refresh their memory of it very quickly, but who have no training in philosophy. In his own words, 'The book *concerns* economics; it *explains* science theory' (p. 9). Like Blaug, he offers a survey of developments in economic methodology, though because of the proliferation of such work in the past two decades, where Blaug's survey stretched back into the nineteenth century, Hands is able to concentrate on recent work. He goes on to provide a survey of what economic methodologists need to know, and it is here that the contrast with Blaug (or, as I will argue later, an apparent contrast) emerges.

Twenty-five years ago, Blaug could take it for granted that the literature that was relevant to economic methodology was philosophy of science, and two comparatively short chapters were sufficient. In contrast, Hands argues that economic methodology needs to draw on a much broader range of work, for which he uses the label 'science theory'. This includes philosophy of science but also sociology and interdisciplinary work that fails to fit within conventional disciplinary boundaries. He organizes the material into four chapters. His starting point is the breakdown of the received view within philosophy of science (ch. 3). This covers much of the material covered by Blaug but, writing so much later, his perspective is very different. Lakatos is presented as a 'first-round', 'quasi-historical' move and is followed by a discussion of realism covering, amongst others, Boyd, Bhaskar and van Fraassen. Though still aiming at economists, Hands offers coverage of the philosophy that is both more up-to-date and more technical. Where Blaug's book could be read as an introduction to the subject, Hands' book is aimed at a higher level – academics or Ph.D. students who know some of the issues and want to study the subject more rigorously.

However, the following three chapters take the material well away from what Blaug covered. Chapter 4, the heart of the book, is on 'The naturalistic turn'. Hands introduces this as 'a turn away from *a* *priori* philosophy and toward a philosophical vision that is informed by contemporary scientific practice' (p. 129). He starts with a discussion of Quine's naturalized epistemology, in which epistemology is based on psychology. From there Hands moves on to consider naturalized epistemology of different sciences. Cognitive science takes him to the work of Alvin Goldman and Herbert Simon. From there he proceeds into biology and evolutionary epistemology and to David Hull. This takes him back to Popper, though this time the evolutionary epistemology is interpreted by Peter Munz. Here he points out that economic elements in the argument for evolution involve competition, a theme that is more explicit in the work of W. Bartley and Gerard Radnitzky, and concludes that he has come full circle, back to Quine's suggestion that there might be 'encouragement in Darwin' (p. 164). The chapter concludes with a section on eliminative materialism and the philosophy of mind.

Chapter 5, 'The sociological turn', covers material that has been much more widely discussed in economic methodology. Arguing that the sociological turn is a variety of naturalism, not an alternative to it, he surveys various sociological approaches to science. He begins with the work of J. D. Bernal and Robert Merton and then proceeds to the sociology of scientific knowledge – approaches that provide sociological explanations of the content of science, organized under the headings 'the Strong program', 'social constructivism' and 'contemporary developments'. The last of these includes the reflexivity school and actor-network theory. After a discussion of social realism (focused on arguments about the 'epistemological chicken' by Harry Collins and Steven Yearley), he points out, as in the previous chapter, the extent to which economic arguments or analogies are being used. In an italicized passage he writes that 'many studies in SSK [sociology of scientific knowledge] look much like what an economist would write about science or the behavior of scientist agents' (p. 208). We find the literature using metaphors of investment, capital accumulation, exchange and markets.

Chapter 6 considers further turns – the pragmatic turn; neo-pragmatism and the discursive turn; and the turn to a more situated methodology, associated with feminism. Discussion of Peirce, Dewey and Rorty leads into discourse analysis and the enormous literature on the rhetoric of economics. The section on feminism is welcome as it is a strand in the literature that is not often brought into broader discussions of methodology.

After these four chapters on 'science theory' there follows a long (79-page) chapter on 'Recent developments in economic methodology', under four headings: the Popperian tradition; the Millian tradition; realist themes; and cognitive and semantic themes. This chapter makes it clear how far the field of economic methodology has progressed since Blaug's *Methodology of Economics*. Not only is the cited literature enormous but, since this is the chapter in which Hands summarizes and passes on from earlier work, it reveals (especially when taken together with the discussions of discourse and feminism in the previous chapter) the range of approaches now found within the field.

Many of the chapters in *Reflection without Rules* end by demonstrating how economic arguments enter into accounts of science. These remarks are brought together in the penultimate chapter, 'The economic turn'. This starts with what is described as the economic turn in contemporary science theory. Growth theory and microeconomics have increasingly been used to explain science. This leads Hands to argue that economics could be the basis on which philosophy of science is naturalized and that economics and epistemology are deeply intertwined. Paralleling the discussion of sociology in chapter 5, he moves on to the economics of science and then to the economics of scientific knowledge, covering the work of Philip Kitcher, Partha Dasgupta and Paul David, and Jim Wible. Kitcher's work is arguably seen as the most significant, for two reasons. He provides an example of a philosopher using game-theoretic arguments of a type that one would expect an economist to use. In addition, he tries to show how one can still have progress even though scientific knowledge is the outcome of social processes in which agents pursue a variety of non-epistemic goals.

This paves the way for the final chapter, which argues the case for a 'New Economic Methodology'. In this, 'the Received View' and 'the Legend' are gone, the search for a few narrow methodological rules to demarcate good science from bad has gone, and use is made of a much broader range of philosophical rules than 'a few decades ago' (pp. 396–400).

Reflection without Rules is an impressive, valuable survey of a broad body of literature that deserves to be widely read. Before turning to the points where I disagree with Hands, it is important to state what should be obvious: there is much in this book with which everyone should agree. The Received View has been undermined. We need to turn to a wider range of philosophical resources than economic methodologists were using around 1980 (I would attach particular importance to Pragmatism). Sociological, evolutionary and even economic perspectives on science are here to stay; and so on. For anyone who came to economic methodology through, for example, Blaug's *Methodology of Economics*, this book shows how much more there is for economic methodologists to draw upon than Popper, Kuhn and Lakatos.

However, despite all this, I suggest that economic methodology should be turning in a direction that is very different from the one Hands proposes. My skepticism about his arguments centres on two issues: whether the book really is about the methodology of economics, rather than methodology in general; and whether his approach marks as much of a break with the old methodology as Hands claims. In the remainder of this review, therefore, I ask what the book has to say about *economic* methodology. I also argue that, though Hands adopts the now-common rhetorical device of arguing for a 'new' economic methodology, his proposed turn keeps uncomfortably close to the 'old' economic methodology that he wishes to move forward from. These discussions lead to my conclusion that economic methodology should be (and has been) turning in rather a different direction.

A good starting point in trying to justify the above claims is with the question of the book's audience. As I mentioned earlier, Hands presents his book as aimed at people with a background in economics who are interested in economic methodology. However, where Blaug was trying to create such an audience, addressing economists who knew relatively little philosophy, Hands is able to speak to a community that has been discussing these issues for two decades. Blaug produced a book that could be used by both students and professional economists. In contrast, this book is at a higher level, aimed at those who have already gotten into the field. The tone of the argument suggests that he believes that this community has not properly recognized that the field has now moved on. At the same time, it is hard to resist the conclusion that Hands also seeks to interest 'science theorists', including philosophers, in the application of economics to the study of science. His conclusion that there is much to be said for naturalizing philosophy, not on generic science, on psychology or on biology, but on economics, is surely one that he wants scholars trained in philosophy to hear.

Given this, what should we make of Hands' claim, cited above, that *Reflection without Rules* concerns economics but explains science theory? Clearly, he chooses to explain science theory because he believes that economic methodologists coming to the subject from economics are not as fully informed about it as they should be. This is undoubtedly correct. It is a book that I, for one, needed to read and that I will find it useful to refer to. But what of the other claim about the book: that it *concerns* economics. This is where I am not convinced. I understand 'concerns economics' to mean something like 'seeks to explain or understand what is going on within economics'. Insofar as Hands is trying to educate those who are engaged in this task, providing them with tools that they need to investigate economics, it might seem obvious that the arguments he discusses concern economics. However, to show that they really concern economics (and are not just presented as if they concern economics), he needs to show that these are the right tools for investigating economics. As an aside, I should add that using economics to understand science (naturalizing on economics) does not mean that the book concerns economics as I understand it.

Most of the science theory that is surveyed in this book could be applied to any science. He draws on literature by a wide range of authors, most of whom are not concerned with economics. One way of defending the thesis that this concerns economics would be to argue that economics is like all other sciences – a unity of science hypothesis – from which, provided we accepted that economics was a science, the claim would follow trivially. On the other hand, suppose we accept that all sciences are different (and I suspect Hands would agree with this). In that case, any discussion of economic methodology would need to pay attention to the peculiarities of economics (in addition, of course, to places where economics has things in common with other sciences). This is the big thing missing from this book if it is seen as a contribution to economic methodology.

Here I come back to the remarks cited in my first paragraph, concerning the book's audience. Hands contrasts his book with books by philosophers, who frequently offer 'an elaborate discussion of some particular aspect of economic theory' (p. 9). The implication here is that economists, who know or can readily brush up on the relevant theory, do not need this, but philosophers need to be taught (or to teach each other) economics. However, surely the reason that philosophers discuss the details of economic theory (my examples here would be Daniel Hausman, Harold Kincaid or Nancy Cartwright) is not that their audience does not understand economics, but that these details are necessary for their claims about how economic arguments work. By turning away from such arguments, Hands is, I suggest, avoiding some of the most important issues concerning economic methodology.

This point is so important that it is worth elaborating on it with an example, Hausman's *The Inexact and Separate Science of Economics* (1992a). Hands provides an accurate account of Hausman's Millian methodology – of tendency laws and the inexactness of economic claims. He illustrates Hausman's attitude to prescription in a one-sentence reference to economists' attitudes to recent psychological claims of utility theory. Though Hands' assessment is correct, he misses the point that, in this case study, Hausman is trying to understand what is peculiar about economics, something he can do only through working through an example in detail. It is the basis for his claim that economics is what he terms a 'separate' science. Another way of looking at Hausman's book is that he is trying to get away from economic methodology that stems from 'theory' – whether off-the-shelf philosophy of science or more general science theory – to create what he terms an 'empirical' philosophy of science (cf. Hausman 1992b, ch. 16). Analysis of detailed case studies in economics is essential; it is not just to aid the reader who does not know the economics. Another way to put it is to say that Hands is telling us that there are many ways in which we could be doing economic methodology, but Hausman, in this case study, is actually doing it.

Hands' book, for all new ideas it introduces into the discussion of economic methodology (and there are many) is much less radical than he implies. He is rightly critical of some early work on economic methodology for taking off-the-shelf philosophy of science and applying it to economics. But it is arguable that he is doing much the same, albeit at a much higher level of sophistication. This is implied when he talks about 'naturalizing' on something: naturalizing on generic empirical science or naturalizing on some specific science, such as cognitive psychology, evolutionary biology or sociology. Though Hands does not see it this way, this seems suspiciously like replacing 'off-the-shelf' philosophy of science with 'off-the-shelf' science of various types. This is consistent with his reference to 'science *theory*' as the body of ideas on which economic methodologists need to draw. One might contrast this approach with *empirical* philosophy of science. This is not because we can somehow find evidence that is untainted by theory (clearly we cannot) but because emphasizing science *theory* biases us away from looking, via detailed analysis of economics, at the peculiarities of economics. Perhaps there is a case for approaching economic methodology not from science theory (though this will undoubtedly remain in the background) but from the history of science or even the history of economics.

Reflection without Rules is backward-looking in that, in addition to suggesting that economic methodology should move in new directions, it continues to fight battles that have been engaged with in economic methodology for the past twenty years. He is still debating the Received View. If 'the Legend' (his other rhetorically loaded label for it) is dead, why does he spend so much time saying so? In the past decade or more, students of economic methodology have begun to move beyond this, in what might be termed (using the terminology that Hands uses when he wishes to express approval) 'the Empirical Turn'. I have already cited Hausman's work on economists' reaction to experimental work. To this could be added work on how economic models are used; on the relationship between macroeconomics and microeconomics; on causality in macroeconomics; on the use of econometric techniques; on the way economics use empirical evidence; and more.

This comes back again to the question of audience. Unlike Blaug, I suggest, Hands is not aiming at regular economists. The issues he discusses are several levels removed from the issues that economists are interested in. His targets are philosophically informed economists and, though he does not claim this, science theorists. If practising economists are to be engaged in methodological discussions, an empirical approach, making much more limited claims, would seem much more likely to succeed. It would not simply be a discussion of practical methods (that would be tedious) but would explore in detail the variety of practices that are found in economics – bottom-up rather than top-down philosophy. It would be unfair to describe *Reflection without Rules* as simply an updated and expanded version (with a radically different conclusion) of Blaug's two chapters on 'What you always wanted to know about the philosophy of science but were afraid to ask'. It goes far more deeply into the subject and points in many new directions. However, it is still telling us ways in which we might do economic methodology, whereas economic methodologists who have taken the Empirical Turn have already begun to do it. *Reflection without Rules* might be seen as a step backwards in the sense that, for Blaug, philosophy of science was clearly the *hors d'oeuvre* before a series of case studies analyzing economics, whereas for Hands, science theory is more like the main course.

Roger E. Backhouse

University of Birmingham

REFERENCES

Blaug, M. 1980. *The Methodology of Economics: How Economists Explain.* Cambridge University Press

Hausman, D. 1992a. *The Inexact and Separate Science of Economics.* Cambridge University Press

Hausman, D. 1992b. *Essays on Philosophy and Economic Methodology.* Cambridge University Press

DOI: 10.1017/S0266267104251392

Trust and Trustworthiness, by RUSSELL HARDIN. Russell Sage Foundation, 2002, xxi + 234 pages.

In *Trust and Trustworthiness*, Russell Hardin pulls together a vision of trust as 'encapsulated interest' that he has been developing in a series of essays over the past decade. Along the way, he provides a very useful overview of the large and rapidly growing recent literature on trust. He argues sensibly that there is no single thing called 'trust', and indeed that the English word has many different meanings, so there may be many useful accounts of trust that differ from Hardin's. There may also be useless and confused accounts of trust, and part of Hardin's effort is critical. In particular, he faults other accounts of trust for (a) failing to distinguish between trust and trustworthiness; (b) taking trust to be a two-place relation, 'A trusts B', or even a one-place relation, 'A is trusting', rather than a three-place relation, 'A trusts B to X'; (c) supposing that a single notion of trust will capture phenomena ranging from trusting a friend, trusting a business partner, trusting a government official or trusting an institution or government to trusting a stranger; and (d) romanticizing and 'de-rationalizing' trust and forgetting that distrust is often sensible and even wise.

Hardin's real concern is *cooperation*. Trust is of interest as one of the crucial factors that explains cooperation. He is thus concerned to put his account of trust to work and show how people manage and use trust, how they learn when to trust and to distrust and how trust enters into accounts of cooperation. One delightful and rewarding feature of the book are the extended discussions of examples from literature and opera that make Hardin's observations about trust come alive.

On the first page of the preface, Hardin writes, 'I trust you because your interests encapsulate mine to some extent – in particular, because you want our relationship to continue' (p. xix). On page 1, he writes, 'On this [encapsulated-interest] account, I trust you because I think it is in your interest to take my interests in the relevant matter seriously in the following sense: You value the continuation of our relationship, and you therefore have your own interests in taking my interests into account. That is, you encapsulate my interests in your own interests.' It thus appears that Hardin's view is that A trusts B to do X if and only if A believes four things:

1. that B will do X (unless there is some interfering factor),
2. that it is in B's *interest* to do X,
3. that (2) is true because B believes it is in A's interest that B do X, and
4. B's belief that it is in A's interest that B do X motivates B to do X, because B has an interest in a continuing relationship with A.

Notice that on Hardin's view trust is cognitive not behavioral. It is not something that one does or decides to do. Although Hardin argues against behavioral views of trust (58–9), it would be bizarre to deny that people ever decide whether to trust somebody. Rather than making such a denial, it is better to take his insistence that trust is a matter of belief rather than action as a proposal for regimenting terminology. Hardin is in effect insisting that one restate claims such as, 'A decided to trust B to do X' as 'A decided to *act on* his trust that B will do X'.

Implicit in this account of trust is a theory of trustworthiness, which Hardin makes explicit in chapter 2. Hardin's encapsulated-interest view of trustworthiness takes B to be trustworthy in interaction with A with respect to X if clauses 1–4 above are true of B. So another way to state Hardin's view of trust is that A trusts B to do X if and only if A believes that B is trustworthy with respect to X.

So, for example, in Dickens' novel, Mr Grimwig disagrees with Oliver Twist's benefactor, Mr Brownlow, about whether Oliver Twist can be trusted to deliver Mr Brownlow's books and a £5 note to the bookseller. Dickens poses this as a disagreement about Oliver's moral character, and Oliver himself is portrayed as more distressed that his kidnapping by Sikes

https://doi.org/10.1017/S0266267104251392 Published online by Cambridge University Press

will lead Mr Brownlow to think him ungrateful and wicked than that it costs him his freedom and comfort. But Mr Brownlow could judge Oliver to be trustworthy in this regard if he merely regarded Oliver as prudent. It is obviously in Oliver's interest to attend assiduously to Mr Brownlow's interests, because it is so clearly in his interest to maintain his relationship with Mr Brownlow, who is the only person standing between Oliver and a life of misery, degradation, or even starvation. Oliver would have to be a fool as well as a scoundrel to run off with the books and the money. Oliver's interests partly encapsulate Mr Brownlow's, and Mr Brownlow's trust is clearly rational.

Although Hardin's general view may seem clear enough, I found *Trust and Trustworthiness* a very difficult book to understand. It seems to me that Hardin may not yet have made up his mind on what his 'encapsulated-interest' theory is. One finds multiple expositions that are not systematic or complete and which do not appear to be fully consistent with one another or with the examples that Hardin gives.

In the passages above, it seems evident that Hardin's account includes condition 4 – that B's interests encapsulate A's because of B's interest in a continuing relationship with A. Indeed Hardin makes the point even more insistently: 'In the encapsulated-interest account, I must know that the agents or institutions act on my behalf because they wish to maintain their relationships with me' (156). But I doubt that this is in fact his view. Sometimes he leaves out the condition (11). At other points, he treats B's interest in a continuing relationship with A as merely a common or typical explanation for why B's interests may encapsulate A's (4, 14). More seriously, he explicitly asserts that there are other explanations (23) for how B's interests can come to encapsulate A's. For example, A may have reason to believe that B's interests incorporate A's because of B's interest in establishing a reputation for trustworthiness (139). A may have a reason to believe that B's interests incorporate A's because of B's interest in maintaining a relationship with an intermediary rather than with A (140). There is also love and altruism (23). Indeed it is hard to see how Hardin *could* insist that in order for A to trust B to do X (in the encapsulated-interest sense) B must aim to maintain the relationship. Surely the interests of a nurse who treats me in the hospital *can* encapsulate mine, even if we are both confident that I will recover and leave the hospital in a week. It may be in his interests to attend to my interests – to make my interests partly his own – because of his horror at suffering, his fear of causing harm, or his desire to make a favourable impression on his supervisor (or my unmarried sister). It would be arbitrary to maintain that trust as encapsulated interest is not possible unless the trustee has an interest in maintaining his or her relation with the trustor.

Hardin might reply that he focuses on relations of trust in which B has an interest in continuing relations with A not because he denies that there are also relations of trust in which B does not have such an interest but because he is only offering 'an account of a particular but important class of trust relations' (xix). Indeed, he repudiates the project of offering an essentialist account of what trust *is*, and at a number of points in the book, he treats some alternative theories of trust and trustworthiness as complementing his account rather than as competing with it. One can certainly grant Hardin that B's interest in maintaining a relationship with A is a common and significant ground for B's interests to encapsulate A's. Even so, it is hard to see why Hardin would want to separate off a species of trust and a species of trustworthiness that has just this explanation.

Why then separate off an encapsulated interest account? I was puzzled by Hardin's treatment of his encapsulated-interest theory as just 'an account of a particular but important class of trust relations' (xix), which is compatible with alternative accounts of other classes of trust relations. As far as I can tell, the only thing that distinguishes the class of trust relations to which the encapsulated-interest theory applies from the class to which it does not apply is that what explains B's trustworthiness and thus grounds A's trust is that B's interests encapsulate A's. The encapsulated-interest theory is not a theory of trust between strangers, trust in commercial relations, trust among friends, trust of professionals. Rather than a theory of some class of trust relations, it seems that Hardin is discussing one important *explanation* for trustworthiness and trust. Near the end of *Trust and Trustworthiness* Hardin writes,

> In the academic literature, there are four main theories or models of trust that are actually brought to bear in empirical claims and research. Three of these are based on the kinds of reasons for judging the trustworthiness of the potentially trusted, and we could as sensibly say that these are three different theories of trustworthiness as that they are theories of trust. These kinds of reasons are encapsulated interest, moral commitment, and commitment from character. Two of these – moral commitment and character – are dispositional reasons, and the other – encapsulated interest – is a reason from interests. The fundamentally important common feature of these three theories or models of trust is that they require cognitive assessments of the trustworthiness of the potentially trusted. The fourth theory is purely dispositional trust that is not grounded in the assessment of the trustworthiness of the trusted and therefore is not at all a theory of trustworthiness. (197–8)

Hardin specifically grants that people are sometimes trustworthy because of moral commitments and commitments from character: 'The standard views of trustworthiness seem likely each to be true of some people in some contexts' (199). Indeed it does not appear that the three views disagree concerning what trust and trustworthiness are. B's trustworthiness with respect to X is a disposition to do X because A trusts B to do X. A's trust that B will do X is A's belief that B is trustworthy with respect to X. Rather than defending an alternative theory or model of trustworthiness or trust or a theory or model of some independently specifiable class of trust relations, it seems that Hardin is arguing that there is an explanation of trustworthiness and hence of trust in addition to moral commitment and commitment from character – in particular an explanation in terms of interests. If this interpretation is correct, Hardin is not defending an encapsulated-interest theory of trust or trustworthiness. His view is instead that trust and trustworthiness have several explanations, and his objective is to clarify one of those explanations that is, he believes, of great importance and generality.

In reading the earlier essays upon which the chapters of this book are based, I rather took Hardin (perhaps mistakenly) as defending the view that people will be trustworthy if and only if it is in their self-interest to be trustworthy – that other explanations of trustworthiness are of negligible importance. The reason why some people are trustworthy and others are not lies in the self-interested benefits the trustworthy can obtain from continuing relationships, not from any difference in moral character. If there is sufficient profit tomorrow to B from her attending to A's interests concerning the matter of X today, then she will be trustworthy with respect to the performance of X, and it is rational for A to trust her. I took Hardin to be defending a theory of trust that competed with, rather than complemented, accounts that relied on moral commitment or character.

Whether or not this reading of Hardin's essays is justified by the texts, in *Trust and Trustworthiness*, Hardin has, as we have seen, explicitly repudiated such a reading. He argues, for example, that in some communities, generalized norms may secure trustworthiness without interest encapsulation (185). He maintains that 'trust can be grounded in the belief that another person is guided by norms or moral dispositions' (52). In conceding the existence of other grounds for trust and other bases for trustworthiness, Hardin obviously cannot claim to be offering 'the' theory of interpersonal trust. Presenting his discussion of the encapsulated-interest theory as an account of a 'class' of trust relations may represent an incomplete retreat from the attempt to offer a theory of trust and trustworthiness to the more modest – but still highly ambitious – objective of sketching a particularly common explanation for trustworthiness and ground for trust.

Whether a competing theory of trust or an alternative explanation for trust, the encapsulated-interest view needs to be distinguished from accounts in terms of moral commitment or character, and it is not clear to me that Hardin succeeds in doing so. One problem is that Hardin does not distinguish clearly between interests, moral commitments, and character. He says very little about what constitutes a person's interests, other than that those interests are a proxy for a whole story of 'well-being through the use of resources', where well-being is not necessarily selfish (23–4). In Hardin's view, it is in the interests of a doctor who treats me that I recover, if she cares about her patients? If a salesman believes that it is wrong to

https://doi.org/10.1017/S0266267104251392 Published online by Cambridge University Press

tell a lie and would feel diminished as a human being if he lied, is it then in his interest to tell the truth? Without some sharper delineation between interests, commitment and character, it is not clear how to distinguish Hardin's explanation for trust and trustworthiness from the alternatives.

Indeed, it seems that there is double-edged risk of trivialization. Suppose that George, a freshman from rural North Dakota at Columbia University, goes out walking and gets lost. He approaches Mary and asks for directions back to his dormitory. He knows that there are unsafe areas around Columbia, and he trusts Mary to provide him with a safe route home. He regards her as trustworthy not – or so one would think – because her interests encapsulate his, but because he believes that she, like almost everybody else (even in the big city), will be moved by norms of truth-telling and minimal civility. Suppose George is right to trust Mary and that, in addition, she is concerned about the young stranger who asks for directions. He's obviously not streetwise, so she takes special care to determine a safe route back for him, and she makes sure that he understands her directions.

This is clearly *not* the sort of trust relationship that Hardin set out to capture. Though he would not deny that George trusts Mary to give him directions home and that Mary is trustworthy in this regard, Hardin is concerned with trust that is founded on more knowledge of the trustee and that plays a role in longer-lasting relationships. But, especially once one grants that moral commitments can explain trustworthiness, there seems to be no way to draw a non-arbitrary line between the sort of trust that George has in Mary and the sort of trust that Hardin is concerned with. Indeed, it is no longer clear that this is not an instance of interest encapsulation after all. In the story above, doesn't Mary acquire an interest in George's safe return? It is not a very deep or important interest, but it is an interest. If she were to learn that he met with harm on his route home, her well-being would be diminished.

Hardin's account is also in danger of collapsing into the supposed alternatives because there is a sense in which all trustworthiness – regardless of its explanation – implies an encapsulation of interest. As Annette Baier (1986) emphasizes, to be trustworthy is not just to do X if one is trusted to do X. If B is trustworthy and learns that A is mistaken in seeking X and will be harmed by X, B's trustworthiness will incline her *not* to do X. In accepting A's trust, B is obligated to act on A's behalf. B should attempt to further A's interests. This involves judgment and discretion, not a blind commitment to perform X. Perhaps this fact about trust doesn't necessarily imply that B's *interests* encapsulate A's interests, because the explanation for B's actions may turn on B's benevolence or B's moral commitments. But if B is genuinely trustworthy, A's interests will *motivate* B somehow. In all plausible accounts of trustworthiness, the interests of the trustor motivate the trustee. Sometimes the interests of the trustor may not literally be *interests* of the trustee, but it is unclear whether the notion of an interest can bear this much weight and whether the competitors to Hardin's views could also be labelled as encapsulated-interest accounts of trust.

What distinguishes Hardin's view is not, I think, after all, his insistence that B's interests encapsulate A's interests. In describing his account as an encapsulated-interest theory of trust Hardin bundles together three theses that should be distinguished:

1. A trusts B to do X if and only if A believes that B will be trustworthy with respect to doing X.
2. B is trustworthy toward A with respect to doing X if and only if B is motivated in part by A's interests, concerns, or preferences.
3. Trustworthiness may be and often is in the interest of the trustee.

The first two theses, and particularly the second, are common to many accounts of trust, though Hardin's emphasis on these theses, especially the first, has been extremely constructive. Although the third thesis is not 'news' either, it has been forgotten often and even more often underestimated. As I see Hardin's contributions – and I think they are substantial – he has defended two central theses and explored some of their many rich implications. First, he has argued that trust is a three-place relation, and insofar as expectations of trustworthiness are rational – and they often are – trust itself is entirely rational. Second, he has argued that trustworthiness is very often in the self-interest of the trustee, particularly because of the value of maintaining a relationship with the trustor.

It thus seems to me that *Trust and Trustworthiness* should be read as an exploration of ways in which trustworthiness serves the interests of trustees and of the ways in which awareness of this influences when people trust one another and the consequences of their trust. Even if one refuses – as Hardin wisely does – to go on cynically to question whether people are ever motivated by more than their own self-interest, it is clearly of the utmost importance to investigate to what extent and in what circumstances trust and trustworthiness can rest on self-interest and how these facts influence the scope, possibilities and character of cooperation. Although a puzzling and a difficult book, *Trust and Trustworthiness* is ultimately a deeply rewarding one.

Daniel M. Hausman

University of Wisconsin, Madison

REFERENCE

Baier, A. 1986. Trust and antitrust. *Ethics*, 96:231–60