
REVIEWS

What Price Fame? TYLER COWEN, Harvard University Press, 2000, 248 pages.

One of the oldest issues in social and political thought derives from three observations. First, that short of outright tyranny people will each generally prefer to have order rather than disorder in their society; second, that for any relatively benign order to prevail people will each have to behave according to certain more or less demanding constraints; and third, that even when this is clear, people cannot be individually relied upon to act according to the saliently required constraints. The issue raised by these observations is so central to social and political thought that it might be described as *the* socio-political problem.

Down to the eighteenth century the problem was mainly posed in relation to political life and, as a result, it was understood as one of how to encourage civic or public virtue in those citizens – propertied, mainstream males as they generally were – who had the power in public life either of challenging those who ruled over them, whether in a representative or monarchical fashion, or of being rulers themselves. The standard view was that while the problem of achieving social order could be reduced by imposing a system of checks and balances on those in public life – this was the ideal of the mixed or blended constitution – its resolution still required citizens to have a developed, essentially moral sense of what the public good required of them and to be responsive to such a moral sense of how they ought to behave. But how could people be expected to respond to their moral sense of such civic requirements?

Social and political thinkers from Aristotle and Cicero to Machiavelli and Montesquieu argued that many people could be expected to respond to such requirements, at least on many occasions, so far as they were possessed of civic virtue. But all of these thinkers were obliged to admit that civic virtue was sometimes in scarce supply. So was there anything, then, to serve as a backup and reinforcer for such virtue? Was

there any motivational device that could play the role of increasing the chance that even when virtue fails, still people will behave in a virtuous manner?

The answer that was given by almost all thinkers from the time of the Greeks and Romans down to the end of the eighteenth century is that yes, there is such a device. According to the received wisdom, it consists in the universal human yearning to be well thought of and, perhaps even more pressingly, to avoid being thought of badly. Polybius gave nice expression to the idea about the beginning of the first century B.C. when he argued that the *philotimos*, the person who loves honour, may be expected on that account to be also a *philagathos*, someone who loves the good. But the theme was there before him in the work of Plato and Aristotle and it assumed great importance among the enthusiasts for the Roman republic that Polybius's histories encouraged, most notably Cicero.

The theme continued in later Western thought, figuring in the recognition by Aquinas that honour is the primary external good that human beings naturally seek – *honor is primum inter bona exteriora* – and achieving particular prominence in the wake of the Renaissance. It found perhaps its most sober and sensible expression in the argument of John Locke, deployed in the *Essay on Human Understanding*, that the law of opinion binds people to seeking the good opinion and fleeing the bad opinion of others and that in favourable circumstances – in circumstances where public standards are clear and public scrutiny available – it can be relied upon to keep them on the straight and narrow path of virtue. Seeking to be honoured, he thought that people could generally be relied upon to do what is honourable.

But while the belief in the law of opinion, as we may call it, was almost universal among Western thinkers in this period, it continually attracted a certain ambivalence. The lover of honour may come to be a lover of the good, but to love honour is still to love something distinct from the good; it is to fall from the highest plane of virtue. The desire for opinion may be a saving vice – a vice that can keep people virtuous – but it is a vice for all that. It comes dangerously close to an eagerness to please and when it leads someone to behave virtuously then it involves a sort of deception: the person behaves in a way that is designed to elicit the incorrect attribution of virtue, not the correct attribution of a desire to be thought virtuous. The love of honour may be a source of continence or self-control, in Aristotle's language, but it falls well short of virtue pure and simple.

This ambivalence shows up among classical authors in the fact that while they praise the desire for good opinion as a source of much social good, and as a source of assurance against the corruption of public life, they do not recommend it in their more personal, ethical reflections on

the nature of the good life. Thus Plato looks in *The Republic* for a reason why the just person should love justice, even when justice is combined with being thought to be unjust. And Cicero, the great enthusiast for the effects of men's love of glory, derides in his more stoic moments the shallowness of the glory-loving soul.

But the fact that such authors criticize the love of honour, treating it as yet another sort of vice, should not blind us to their belief in the social power and public benefit of that desire. They see in it a force for good that works like a providential mechanism for the benefit of human kind. They cast it as 'an intangible hand' – something that may even manifest the hand of God – that snatches public good out of private bad, public virtue out of private vice. (The 'intangible-hand' phrasing is used in Geoffrey Brennan and Philip Pettit, 'Hands Invisible and Intangible', *Synthese*, 1993, pp. 191–225 ; see too Brennan and Pettit 'The Hidden Economy of Esteem', *Economics and Philosophy*, 2000, pp. 77–98.)

If this is right, of course, then Bernard Mandeville was not very original in the themes he struck when he argued early in the eighteenth century that private vice is public virtue. What he was doing, at least at one level, was giving a cynical twist to a piece of received wisdom. Where the tradition had identified a benign form of hypocrisy that produced social good – the simulation of virtue, as it was called – he rehearsed the discovery as if what he had revealed was the more malign hypocrisy of dissimulation. To simulate is to act in a virtuous manner, representing oneself falsely as having been motivated by virtue itself; to dissimulate is to act in a non-virtuous manner, representing oneself falsely as having acted virtuously. The distinction had been available in English since at least the time of Francis Bacon but Mandeville collapsed it for his own satirical purposes.

Mandeville also did something else, however, that was of much greater moment in the development of social and political thought. Without distinguishing the devices clearly, he pointed out that not only does the simulation of virtue often do just as well as the real thing in producing a certain social order; there is another, equally welcome sort of order that is produced by the naked pursuit of self-interest and not by any recognizable sort of virtue, real or feigned. In other words he drew attention, not just to the intangible hand whereby the love of honour produces honourable deeds, but also to the invisible hand whereby non-honourable or even dishonourable deeds produce that benign form of social order.

Adam Smith was the first to give the name of 'the invisible hand' to this device and to document the extent to which the invisible hand could operate in social life, producing all the cost-reducing, coordinative wonders of the competitive market. He distanced himself carefully from Mandeville's cynicism, however, and rejoiced in the recognition that the

intangible hand – though he did not give it that name – also had an important part to play in social life. He even suggested that it was because of the intangible hand – because of the love of esteem and distinction – that the rich could be relied upon to seek luxury services and goods and keep the economy active enough to ensure the general welfare.

The invisible hand that was celebrated in Smith's writing, however, soon came to eclipse the intangible hand in which he also believed, and it was not long before the discipline that he launched had forgotten completely about that traditionally more hallowed mechanism. It was not long indeed before the broader discipline of social and political thought had forgotten about it also. With a very few exceptions, the mechanism that had been hailed and invoked for two millennia – always, it is true, with a certain ambivalence of attitude – fell into almost complete neglect in the nineteenth and twentieth centuries. The invisible hand was used to explain the benign, unintended results of the competitive economy and was often extended in the attempt to explain allegedly similar results in other spheres of social and political life. It became the hallmark of the sort of social science that took economics as its model and while that was not by any means the only sort of social science pursued in this period, the competing, more collectivistic paradigms ignored the role traditionally assigned to the intangible hand with the same determination as they ignored the role ascribed by their rivals to the invisible.

Happily, this period of indifference to the role played in human and social life by the love of esteem is beginning to pass. There have been a number of writers in recent decades who have begun, not just to notice the scarcely avoidable fact that esteem matters to people, but also to describe how the desire for esteem figures as an important determinant of macro-level social patterns. Tyler Cowen is one of these writers and his recent book, *What Price Fame?*, is a good introduction to the riches awaiting those who choose to look with the eyes of the social scientist at this long-ignored dimension in people's desires and at the dealings with one another that those desires help to shape.

The book starts from the premise that people care about fame, though it says little or nothing about the long tradition of support for that claim. So far as it has an historical side, it tends to quote from the downbeat passages in which writers are morally critical of the desire for esteem rather than in the upbeat tracts in which they wax lyrical about its social, virtue-supporting aspects. But this distortion of the tradition is easily forgiven, for the book proceeds in the elegant, eye-catching style of a Thomas Schelling to document a series of effects, many of them quite paradoxical, that characterized a system in which the few take steps to realize a demand for fame and the many make responses that serve to supply it.

The book begins with a chapter that documents the intensity of the contemporary fame culture, giving us a brisk overview of the travail of the famous, the toil of their fans and the outcomes, frustrating and satisfying, of the exchanges between these two groups. The second chapter then goes on to display the way in which fame and merit often separate – contrary, so it seems, to the traditional wisdom I described above – and the reasons why they come apart. One reason is that information is expensive and that people are quickly tipped into going along with what seems to be the general opinion of someone's worth; thus even modest recognition in the right quarters can snowball, justifiably or unjustifiably, into widespread fame. Another reason is that fame in a mass society is more easily coordinated around salient, often simplistic characteristics than around the nuanced features of true merit. Yet another is that fans like being fans – in particular, fans together – and opportunities for achieving the satisfaction of fandom, no matter how independent of real merit in the target individual or group, are likely to be grasped with enthusiasm. And another reason still why fame and merit may come apart is that there are a variety of ways in which those who can best achieve fame for others – the reviewers, the critics, the connoisseurs – can be given a stake, openly or covertly, in that result.

Cowen is quite ingenious and persuasive at describing these sorts of devices and the best chapters of his book are devoted to such descriptions. The chapters are organized around the different angles from which such devices can be seen at work, not around the different types of devices that he identifies, which makes the book sometimes repetitive. But, repetitive or not, there is no denying the cumulative effect that the book achieves by means of this strategy. As example piles up on example, description on description, the reader can hardly help but succumb to the sheer bulk of the evidence marshalled. There is nothing for it but to join the author in acknowledging the extraordinary presence and role played by the fame industry in our society.

But while Cowen thinks that the dynamics of that industry do drive a wedge between fame and merit – and do, as he emphasizes, support the downbeat theme in the tradition I described – he still maintains, with the Mandevillian relish that economists often display, that overall the system works for good, and not for ill. Although the theme is given less prominence as the book progresses, he opens with praise of the society in which fame bulks as large – and as merit-indifferently – as it does in ours and he contrasts it favourably with the fame-hostile, merit-loving society of which he suggests that earlier moralistic thinkers like Plato dreamed.

In sounding this theme, he sometimes reads like Popper on the open versus the closed society. He is happy to admit that from the point of view of a Plato, characterized as interested in truth with a capital 'T', a

fame-centred society like ours is bound to look unattractive; it is bound to present itself as an arena in which genuine merit is put out of view and truth is displaced in favour of falsehood and pretence. But he insists that in this rather staged choice, we should choose in favour of our own, non-Platonic world. Even if fame comes apart from merit, the fame industry still serves us well. It provides satisfaction for the fans and the famed alike, and it ensures that people are well informed about the alternatives represented by those producers, ranging from film stars and musical artists to academic researchers and authors, who seek to win popular fame.

But is not the collective pursuit of fame by rival contenders a zero-sum game in which each spends more and more resources, only to find themselves no better off than if no one had spent anything? No, he maintains. We may be able to envisage a state of affairs in which no one can win any more fame without someone else losing out, but we are certainly not at that stage yet: 'fame remains positive-sum at its current margin' (p. 114). Thus, so the argument goes, there is nothing inherently self-defeating in the aggregate pursuit of fame and there are real benefits achieved in the society that sponsors that pursuit.

Let me conclude with one general reservation about the argument in this intriguing and informative study. As the book does not display an awareness of the older tradition to which it in effect would reconnect us, so it is not sensitive to a distinction that anyone in that tradition would have recognized. On the one side of that distinction is the esteem you enjoy, or the disesteem you suffer, so far as someone else thinks well or badly of what you do. On the other side is the fame you enjoy, or the infamy you suffer, so far as it is a matter of common belief in a certain community that almost everyone thinks well or badly of you in that way: that is, when almost everyone thinks (rightly or wrongly) that almost everyone thinks well or badly of you, when almost everyone thinks that this is the case, and so on in the usual hierarchy.

Cowen is clearly concerned, as the title of the book suggests, with the economy of fame – and, to a much lesser extent, infamy – but it would have been very useful had he related the way that fame and infamy motivate people, and the effects they have in the aggregate, with the way in which the broader economy of esteem and disesteem operates. This might have influenced his treatment, for example, of morality. He recognizes that fame does not attach in our society to straightforward moral behaviour as such and argues, plausibly, that this is because moral behaviour is relatively common and does not make a person stand out among his or her fellows (p. 63). But he worries about the implications of this for the stability of moral behaviour in a fame-oriented society and invokes some considerations designed to allay that concern.

Had he been clear about the distinction between esteem and disesteem, then he would have seen that the worry is not really serious. Although moral behaviour is not a matter of fame, it may still attract esteem and it may still be supported by people's desire to be esteemed. And even if it is so common as not to attract positive esteem, that very fact should be sufficient to trigger a correspondingly higher degree of disesteem for any failure to behave morally, in which case the desire for esteem – specifically, the aversion to disesteem – will continue to operate in a supportive role.

But to make this point is only to draw attention to other work that can be pursued in this area, not really to take away from the achievement of Cowen's book. It is an original and fine contribution to what will surely be an area of continuing investigation.

Philip Pettit,

*Research School of Social Sciences,
Australian National University.*

Foundations of Causal Decision Theory, JAMES M. JOYCE. Cambridge University Press, 1999, xii + 268 pages.

Joyce's *Foundations of Causal Decision Theory* is one of the best books in decision theory to have appeared in recent years. It is one of those rare works that combines philosophical and technical sophistication, that manages to be both accessible and path-breaking, and that both refreshes our understanding of well-established areas in decision theory, while making significant advances of its own. It is essential reading for anyone interested in decision theory and highly recommended for those with an interest in causality, epistemology, philosophy of science or philosophy of mind

This book has much to offer those coming to decision theory for the first time. The first three chapters and a good deal of the fourth and fifth chapters are devoted to an exposition of the central elements of expected utility theory in its various instantiations. The presentation of Savage's work is one of the clearest and most insightful around, and although this is well-trodden territory, Joyce's book takes discussion of it to a new level of sophistication. Similarly, his use of a unified framework for formulating and comparing evidential and causal decision theory makes for an easier understanding of what is at stake between them. (The reformulation of Savage's theory is in fact no trivial matter, for it is far

from clear that there are propositional equivalents of Savage's actions. It is unfortunate that Joyce does not spell out how it can be done.) If there is any disadvantage to making this one's first book in decision theory, it lies in the fact that one does not get to appreciate the freshness of Joyce's presentation of, and reflections on, the material (for which some familiarity with 'standard' debates is required).

For the expert, on the other hand, the book offers a wealth of innovative results and new insights. Highlights include a novel solution to the problem of uniquely determining rational degrees of belief in Jeffrey–Bolker decision theory, a 'cutting-edge' exploration of the nature of conditional belief and supposition, and a representation theorem for a generalized conditional expected utility theory that supports a partition-independent version of causal decision theory. Any one of these would suffice to make his book worthy of our attention. And although I am quite critical of some of his proposals in the more detailed evaluation that follows, it is clear that this is a book that no one interested in the field can ignore.

1. PRAGMATISM AND THE PROBLEM OF UNIQUENESS

Decision theories are theories of rational agency that make claims both about what a rational state of mind consists in (e.g., that degrees of belief are probabilities) and about what decisions a rational agent should make, given their state of mind (e.g., that they should maximize expected utility). Decision theorists typically try and justify their theories by proving representation theorems for them: canonically a demonstration that a small number of axiomatic constraints on an agent's preferences suffices to ensure that they are (uniquely) representable as consequences of degrees of belief and desire that fit the proposed theory of rational states of mind.

One of the strengths of Joyce's book is his discussion of the status of such representation theorems. Debate concerning the axioms of preference underlying these theorems often fails to recognize the different roles played by axioms of rationality, which express claims about the nature of rational agency, and structural axioms, which make the representation problem mathematically tractable. (Let alone the fact the representation theorems themselves serve different roles depending on whether the theory is interpreted normatively or descriptively.) The issue of the completeness of preferences is a case in point. As Joyce points out, it is no requirement of rationality that agents have complete preferences. Indeed completeness is not even necessary for expected utility representations of preference to exist. Decision theorists assume completeness because they cannot prove strong representation theorems without the assumption, and in the hope that by showing what proper-

ties an agent's degrees of belief and desire must have when their preferences are complete, they shed light on what properties they should have when they are not. (The idea is essentially that the degrees of belief and desire of an agent with incomplete preferences must satisfy the canons of decision theory if their preferences are to be 'coherently extendible' to a complete set.) While this hope clearly needs substantiation, it is fatuous to declare decision theory false because people's preferences are not typically complete.

Joyce's care pays ample dividends in the evaluation of Savage's representation theorem. Discussion of Savage has been dominated by debate over the status of the sure-thing principle and its purported 'falsification' by the Allais paradox and the results of experimental psychology. But, as Joyce makes clear, the real problem with Savage's theory lies not in the rationality assumptions, but in the structural ones. In particular, Savage assumes that for *any* given consequence of acting, there exists an action (called a 'constant act') that has this consequence in every state of the world. But this is highly implausible, given that consequences are supposed to identify all that matters to an agent.

No such implausible assumptions are to be found in Bolker's representation theorem for Richard Jeffrey's decision theory. There is a cost, however: Bolker's axioms of preference do not suffice to determine unique probability representations of an agent's beliefs. Indeed, they do not ensure that all representations of the agent's partial beliefs will agree on whether or not one proposition is more likely to be true than another. Joyce has an interesting solution to the problem. He places axiomatic constraints on agents' judgements of comparative likelihood as well as on their preference judgements and shows that this suffices for the unique determination of their degrees of beliefs as well as the determination of their degrees of desire up to choice of scale.

Joyce is well aware that his solution conflicts with what he calls Pragmatism: the idea that the laws of rational belief cannot 'stand alone' and should be underwritten by the laws of rational desire or preference. Pragmatism is the default philosophy of decision theorists, for the most part because they view the 'reduction' of partial belief and desire to preference to be an essential part of the 'reduction' of non-observable mental attitudes to observable behaviour. Joyce argues that the non-uniqueness problem in Jeffrey–Bolker decision theory shows that Pragmatism is false. This claim is clearly a little hasty, however, as is demonstrated by the existence of an alternative solution to the non-uniqueness problem, based on an extension of the domain of the preference relation to non-Boolean conditionals, that is perfectly consistent with Pragmatism (if not with the kind of strict behaviourism that decision theorists sometimes espouse). (See Bradley, R. 'A representation theorem for a decision theory

with conditionals', *Synthese*, 116:187–229, 1998.) Though I think it is best to regard these solutions as complementary rather than conflicting, Joyce needs to amplify his argument against Pragmatism. Perhaps because he is concerned with decision theory as a normative rather than an empirical theory, he only argues against the necessity of conceptual reduction. I suspect, however, that most decision theorists adhere to Pragmatism on methodological grounds; in the belief that the nature of the admissible evidence for mental states (namely, observations of people's behaviour) forces us to get to belief via preference. They are unlikely to be moved by anything less than a radical re-examination of the question of the evidential basis for our attributions of mental states.

2. CAUSAL DECISION THEORY AND CONDITIONAL BELIEF

One notable strength of Jeffrey's decision theory is that it offers a partition-independent expression for the desirability of an action. And Joyce regards the fact that in causal decision theory the utilities of actions have to be calculated relative to the 'right' sort of partition as one of the main problems facing it. His proposed solution is both bold and surprising. He suggests that causal decision theory ought to grant that Jeffrey has the correct theory of desire – that '*all value is news-value*' (page 178 of his book; the italics are his). The difference between causal and evidential decision theory is located, he argues, not in the theory of value, but in the epistemic positions they adopt for the purpose of assessing the desirability of actions. While evidential decision theory judges the value of an action A by asking 'how good would things be, if I learn that I will perform A?', causal decision theory does it by asking 'how good would things be if I were to perform A?'.

Joyce pursues his hypothesis by a thorough examination of the different kinds of suppositions involved in asking these questions. The current literature on suppositions tends to be highly technical and in the end it is not clear how much it enhances our understanding of the distinction between indicative and counterfactual supposition. Joyce's presentation is as clear and helpful as always, however, and for my taste there is enough of interest here – including an axiomatic characterization of the class of suppositional probability functions and an interesting contribution to the problem of old evidence – to justify the effort needed to absorb the technicalities. But at the very least this is an area of 'work in progress' and the crucial task of characterizing 'correct' subjunctive supposition remains to be achieved.

Let us return to Joyce's central insight: that causal and evidential theory differ primarily with regard to the kinds of suppositions involved in the evaluations of actions. This is cashed out formally by giving a schema for the value of X on the supposition that Y, $V(X||Y)$, relative to

an assignment, u , of utilities to atomic possibilities (worlds, w) and a suppositional probability measure on propositions, $P(\|)$:

$$(CV) V(X\|Y) = \sum_w \frac{P(w \& X\|Y)}{P(X\|Y)} \cdot u(w)$$

Different kinds of supposition give rise to different instantiations of CV. In the case of indicative supposition, where $P(\bullet\|\bullet)$ is the conditional probability function, $P(\bullet|\bullet)$, we obtain an expression for the desirability of X conditional on the evidence of Y being true:

$$(CD) V(x|Y) = \sum_w P(w|XY) \cdot u(w) = V(XY)$$

And in the case of counterfactual or subjunctive supposition, where $P(\bullet\|\bullet)$ is the imaging function $P(\bullet|\bullet)$, we obtain an expression for the desirability that X would have were Y true:

$$(CU) V(X\backslash Y) = \sum_w \frac{P(w \& X\backslash Y)}{P(X\backslash Y)} \cdot u(w)$$

Joyce’s suggestion is that causal decision theory be construed as saying that the utility of an action, A , is its desirability were it to be performed, $U(A) = V(A\backslash A) = \sum_w P(w\backslash A) \cdot u(w)$ i.e., the news-value of A on the subjunctive supposition that one will perform it. Evidential decision theory, on the other hand, is construed as saying that it goes by its news-value or desirability given that it will be performed, $V(A|A) = \sum_w P(w|A) \cdot u(w) = V(A)$. Now note that it follows from CV that the value of $V(X\|Y)$ is partition invariant in the sense that for any partition of X , $\{X_1, X_2, \dots, X_n\}$:

$$V(X\|Y) = \sum_{i=1}^n \frac{P(X_i\backslash Y)}{P(X\backslash Y)} \cdot V(X_i\|Y)$$

It follows then that the quantity $V(A\backslash A)$ too must be partition invariant. So Joyce’s proposal frees causal decision theory from partition-dependence.

3. THE REPRESENTATION OF CONDITIONAL PREFERENCE

The *pièce de résistance* of Joyce’s book is a representation theorem for his generalized conditional expected utility theory. Technically the innovation here is to extend Bolker’s mathematics to the representation of conditional preferences and likelihood judgements – comparisons of one thing on the supposition that something is true to another on the supposition that something else is true. There is little doubt that Joyce’s

work represents a considerable advance over extensions of Savage's theory to conditional preferences. But I have several reservations about how far it goes towards providing a solid foundation for causal decision theory: in particular, I do not believe we are capable of making robust preference comparisons of the kind that Joyce's representation theorem presupposes, while those judgements that we can make are in conflict with Joyce's theory.

Let me elaborate on the first point. I see no special problems with judging whether one prefers X to Y on the supposition that A (special, that is, in the sense of being different or more acute, from those facing ordinary preference comparisons). But the situation is altogether different when being asked to compare X on the supposition that A to Y on the supposition that B . A useful analogy is with judgements of relative height made from different vantage points. If asked to compare the height of X and Y from the perspective of vantage point A , one can move to A and take a look. But if asked to compare the height of X from the perspective of A to that of Y from the perspective of B it is much less clear how one is supposed to make a judgement (or indeed what one is being asked to judge). I do not deny that there are ways of arriving at an answer. One could for instance apply the standard quantitative theory of height measurement and compare by subtracting the heights of A and B respectively from those of X and Y and then seeing which number is greater. So, too, one could use conditional expected utility theory to guide one's conditional preference comparisons. But this is putting the cart before the horse as far as the representation problem is concerned. It is our pre-theoretical preference judgements that are supposed to justify conditional expected utility theory and not the other way around.

My second concern is that, insofar as we are able to make preference comparisons of the requisite kind, they do not conform to Joyce's theory. Let me illustrate this in the first instance with reference to the evidential instantiation of Joyce's conditional expected utility theory. Note that it follows from CD that $V(X|Y) = V(Y|X)$ and take as X and Y respectively the propositions that I am going to lose my train ticket and that I have enough money to buy a new one. It seems clearly desirable that I have enough money to buy a new ticket given that I am going to lose the one I have, but not that I am going to lose the ticket, (even) given that I have enough money for a replacement. So, contrary to CD, $V(Y|X)$ should exceed $V(X|Y)$. (For the record, I think the right expression for the conditional desirability of X given Y is not $V(XY)$, but $V(XY) - V(Y) + V(T)$, where $V(T)$ is the desirability of the tautology. I defend this view in Bradley, R. 'conditional desirability', *Theory and Decision*, 47:23–55, 1999. The representation theorem that I prove there only supposes that we are capable of comparing one thing to another against the background of a fixed supposition.)

This example illustrates the difficulties of getting comparisons relative to different suppositions right. The notion of news-value does, however, give us one particularly robust toehold on the issue; namely the principle that the value of the news that something is true, when you already know it to be, is worthless. Since Joyce accepts that all value is news-value he should accept this principle. Now the news that A on the supposition that A is clearly no better (and no worse) than the news that B on the supposition that B, whatever the news-values of A and B. So it should be the case that $V(A|A) = V(B|B)$ (as it is in the expression proposed in the last parenthetical note). But according to CV, this will not be the case unless $V(A) = V(B)$. So Joyce's proposal is not really consistent with his claim to be providing a theory of conditional news-value.

A final note: although I think that Joyce has got conditional expected utility wrong, none of this in any way undermines his claim that what distinguishes evidential and causal decision theory is the form of supposition involved. Indeed, at the very least, this hypothesis deserves further serious consideration.

4. CONCLUDING REMARKS

In this final section, I would like consider what kind of foundation his representation theorem, or any one based on a characterization of different kinds of suppositions, can give to causal decision theory. What Joyce's representation theorem shows is that comparisons of prospects of the form 'X on the supposition that Y' suffice (under the right conditions) to determine a representation of the agent's degrees of conditional belief and desire that conforms to CV. As conformity of belief and desire with CV is consistent with acting in such a way as to maximize expected 'news-value', Joyce's representation theorem is actually neutral between causal and evidential decision theory. Ideally, however, what causal decision theory needs is a demonstration that conformity of an agent's judgements with principles of rational comparative preference and belief suffices to ensure that they act in the way that causal decision theory recommends.

Recall that, according to Joyce, what distinguishes the evidential and causal decision theory is the kind of supposition that they recommend for the evaluation of actions. He shows that these can be teased apart by further constraints on the agent's comparative beliefs, but he does not claim (and rightly so) that considerations of rationality require one to perform one kind of supposition or another, without further specification of the context in which the supposing occurs. Causal decision theory requires me to adopt a particular epistemic perspective when evaluating the effect of performing a particular action, but does not say that I should not evaluate the action from a different perspective under different

circumstances. So when we add further constraints to the theory of rational comparative belief and preference, as Joyce suggests we do, we succeed in *characterizing* the perspective of the evidentialist or causalist, but we do not succeed in *rationalizing* or justifying them. Ironically, in view of his rejection of Pragmatism, justification of causal decision theory can only be obtained by a further constraint on preferences for actions – one that forces a rational agent to prefer ones with higher causal expected utility rather than those with higher news-value. It remains an open question as to whether this can be done or not.

Richard Bradley

London School of Economics

Models as Mediators. Perspectives on Natural and Social Science, MARY S. MORGAN and MARGARET MORRISON (eds.). Cambridge University Press, 1999, xi + 401 pages.

Once upon a time, philosophers of science thought it was their business to provide a theory of scientific theories. What came to be known as 'the received view' identified theories with languages. In particular, it took it that the language of first-order logic provided the framework in which the syntactic structure of a theory (conceived as an axiomatic system) could be cast. Issues of interpretation (what is a theory a theory of?) were mostly relegated to finding the right correspondence rules which link the language of theory to the world (and especially to the relevant empirical phenomena). But soon, this orthodoxy was replaced by another, one reason for the replacement being that the 'received view' failed to explain adequately how theories hook onto the world. Semantics and set-theory (i.e., mathematics) took centre stage in the new characterization of theories. Theories were no longer identified with languages, but instead with a (class of) set-theoretic structures or, more informally, models. Where Rudolf Carnap, for instance, took models to have no more than an aesthetic or didactic or, at best, a heuristic value, the new orthodoxy – exemplified in the writings of Patrick Suppes, Fred Suppe, Bas van Fraassen and Ronald Giere – saw models as the fundamental unit of scientific theorizing, theories themselves being families of models. However, both the old and the new orthodoxy were in essential agreement on the legitimacy of the philosophical project of offering a theory of theories. Both engaged in a kind of rational reconstruction of actual scientific theories, though the new orthodoxy – also known as 'the

semantic view of theories' – insisted that theirs was in closer contact with actual scientific practice. But, by focusing on models, the new orthodoxy claimed to be in a better position than the old orthodoxy to explain how theories get applied to the world and how they represent aspects of it. Yet, the story cannot end with the fortunate clause 'and they lived happily ever after'. For they did not, and the book reviewed here shows one major reason why.

Models as Mediators, (henceforth *MaM*), has the same interest in models as the Semantic approach. In fact, it is even more enthusiastic about the ineliminable and central role of models in scientific theorizing. But it is also more pessimistic (and certainly more realistic) about the prospects of a general and ubiquitous characterization of models. Where the semantic view treated (more or less) all models as models of a theory (theoretical models), *MaM* emphasizes the diversity of models, their (partial) independence from theory and the plurality of the ways in which they can represent whatever they do. In fact, one can argue that although *MaM* inflates the role of models in science, at the same time it deflates the need to offer a general philosophical theory of models. The slogan that is encapsulated in the title of the book is, in a sense, the most informative *general* statement about models and their role that *MaM* offers. The book stems from what Nancy Cartwright (p. 241) calls 'the LSE/Amsterdam/Berlin modelling project' and consists in a number of detailed case studies in physics, chemistry and economics which aim to illustrate (if not prove) the slogan that 'models mediate between theory and the world' (p. 242). The choice of focusing on different sciences is not accidental, and creates an impression of diversity and unity at the same time. On the one hand, there seems to be no substantive feature shared by all the models discussed in this volume. On the other, *MaM* purports to show not only that models are central to both the social and the natural sciences, but also that models function in (more or less) the same autonomous way in both domains.

MaM is a collection of papers written by different authors. Four of them are about economics, five about physics and one about chemistry. Since most of them discuss in some detail actual models, they tend to be rather technical and the readers who are not familiar with either physics or economics will find the task of following the details quite difficult. Yet patience with the book is rewarded, since most chapters have important methodological and philosophical insights and arguments. The heterogeneity of the chapters is counterbalanced by the important leading chapter by Margaret Morrison and Mary Morgan. Morrison and Morgan's paper does two things. First, it outlines 'an account of models as *autonomous* agents' and sketches 'how they function as *instruments* of investigation' (p. 10). Second, it locates all other pieces in the volume *vis-à-vis* this account. So, the prospective reader is recommended to start

with the leading piece and then, according to her interests, to move through the rest. Reading this chapter creates the firm impression that all the essays in the volume share some background views in common and engage in the same project. The reader, however, should not be misled into thinking that there is a 'Mediators' theory of theories. The authors do not seem to share any substantial thesis about the nature of scientific models, and the contributions turn out to be less homogeneous than one would expect them to be after reading the leading piece.

One central issue that crops up right at the beginning is the relation between models and theories. *MaM* presupposes throughout that there are theories and models and the world. Leaving the world aside, *MaM* presupposes that there is a distinction between theories and models. We are told that 'we should be mindful of the ways that models and theories do interact' (p. 8). We are also told that models are (partially) independent from theories, that models may represent 'some aspect of our theories about the world' (p. 11), that models are situated 'outside the theory-world axis' (p. 18) and suchlike. But we are not told what theories are and how they are different from models. This omission is revealed in various chapters. Both Ursula Klein and Mary Morgan, for example, rely on an intuitive distinction between different levels of concreteness, labelling what lies at the most abstract end of the scale 'theory', and what lies towards the other end 'models'. However, nowhere in the book is such a distinction fully articulated and made explicit. The scientists' 'rough and ready distinction' (p. 18, fn 3), which is briefly mentioned in the piece by Morrison and Morgan, seems to be at odds with what most of the individual authors have to say on models. Most of them reject (implicitly or explicitly) the 'rough and ready' view that, relative to theories, models are 'less certain or incomplete in important respects' (*ibid.*). On the contrary, the models discussed in this book seem to be more complete than theories (because they are endowed with more concrete details), and more certain than theories (because it is primarily the models, as opposed to the high-level principles, that are confirmed by the empirical evidence). We think that the problem we raise here is not a mere quibble. Its investigation would help demarcate clearly the view of *MaM* from the Semantic conception. As things stand, it is not entirely clear whether *MaM* offers an alternative – fundamentally different – view or whether it suggests ways in which the Semantic view could be complemented. To be sure, Morrison and Morgan state, in passing, that 'theories consist of general principles that govern the behaviour of large groups of phenomena' (p. 12). But this statement is consistent with both the 'received view' and the Semantic view. In particular, it can be construed in such a way that the general principles single out precisely the theoretical models of the Semantic view.

Things get more complicated when it comes to the characterization

of models. Here Morrison and Morgan are very explicit: 'we do not see ourselves as providing a "theory" of models' (p. 12). This might be disappointing for all of those who expected *MaM* to address this philosophical issue. But the individual chapters make clear why *MaM* cannot possibly offer such a theory: the models dealt with in the book are so diverse and disparate that they cannot really be covered by a general description. Morrison, for instance, talks of 'theoretical models' which can be derived from theory (p. 46), of models which are not strictly theoretical, but not phenomenological either (e.g., the nuclear models), and of phenomenological models which are 'motivated solely by the phenomenology of the physics' (p. 54). According to Morrison all these models depend on theory – so she rightly dismisses the view that phenomenological models are theory-free (p. 44). Mauricio Suarez, on the other hand, focuses his attention on 'mediating models', which "'fill . . . in" the abstract descriptions afforded by the theory', and contrasts them with other kinds of model (pp. 168–9). In support of this, he offers the case-study of superconductivity and concludes that the relevant model has an 'independent and non-reducible phenomenological' dimension (p. 187). The general picture that emerges is one of diversity.

This impression is accentuated by the papers which deal with economic case studies. Most of these papers simply describe models in economics, with little discussion of what models are in general. This probably reflects the relative lack of a shared paradigm in the philosophy of economics, where the Semantic view has won little consensus and where no one seems to agree on which philosophical problems (if any) are worth tackling in the first place. Geert Reuten's chapter, for instance, is mostly devoted to illustrating and trying to make sense of Marx's 'Schema of Reproduction'. Although some general methodological remarks are attached at the end of a lengthy case study, the overall impression is that this paper is mostly driven by exegetical preoccupations. So, unless one is interested in the exegesis of Marxian economics, it is not clear what to make of this detailed historical reconstruction. Similarly, Adrienne van den Bogaard's chapter is a nice piece of history of ideas, with an eye on the institutions that influenced (but were also conditioned by) the usage of different models and statistical techniques. But little philosophical elaboration can be found in this paper either.

Perhaps, a general descriptive statement about what models are that emerges from the book is that they are kinds of 'representative structure' (p. 33). But even here, we get little by way of an account of representation. R. I. G. Hughes offers a suggestive summary of his DDI account, which renders representation a function of three things: (a) the Denotation of elements of the subject of the model by elements of the model; (b) the Demonstration within the model of several conclusions; and (c) the Interpretation of these conclusion in terms of the subject of the model

(p. 125). Morrison talks of 'structural dependencies' (p. 63) but she also allows a (rare) kind of 'mirroring', where we get 'an increasingly realistic picture of the actual object or physical system' (p. 60). Yet, her general point is that 'there is no one way to characterise the nature of . . . representation' (p. 64). Suarez takes issue with the view that representation proceeds via 'deidealizations of theory' (p. 182). Cartwright dismisses the notion of representation as 'picturing' and suggests that models 'resemble the situation they represent' (p. 262). Finally, Stephan Hartmann connects representation to a loose 'story' that accompanies the interpreted formalism of the model (p. 344). Morrison and Morgan sum it all up by saying that 'a representation is seen as a kind of rendering' (p. 27).

If all these sound like weaknesses in the book, it also has its strengths which compensate for them. *MaM* might not offer a theory of what models are, but it does offer a kind of theory of *what* models *do* and *how* they *do* it: models mediate between theory and the world and they do that by being autonomous agents, that are irreducible to either theory or the data. Now, insofar as the Semantic View denies the autonomy of the models (an issue on which, we think, the jury is still out), *MaM* offers a substantially different approach. This approach is broadly *functionalist*. Models are not individuated by their content, nor by any account of how they represent. Rather, they are individuated – and distinguished from theories – by the *functions* they perform, of which there are four. The first relates to how they are constructed: seldom does the theory provide the entire stock of building blocks for the model; in most typical cases, elements from one or more theories, other models, and the data cooperate to build a model. Models are 'autonomous' from each one of these sources, in the sense that they are not derived uniquely from any one in particular. The second function relates to their being used as instruments for the exploration and development of theory as well as, more directly, for more accurate measurements. The third relates to their *ability* to represent. And the fourth function relates to their ability to enhance learning: this is not exhausted in the construction of the model; it is supplemented by the use of the model.

The foregoing functionalist conception is best seen as emerging from a set of broad generalizations about the way models function, their role, their relation to theory and empirical evidence, the way they are used, and the sort of knowledge they embody and can generate. Such generalizations are supposed to be derived inductively from the case studies in this book. An especially suggestive way in which models function is proposed by Cartwright. She notes that most concepts of high-level theories are abstract and that models – what she calls 'interpretative models' (p. 257) – are indispensable in giving concrete content to them. In fact, her thesis seems stronger than that, since she notes that "Force"

... being abstract, it can only *exist* in particular mechanical models' (p. 257, emphasis added). Cartwright's paper also provides the most direct attack on the Semantic view, which is criticized as a specific instantiation of what she calls the 'vending machine' view of modelling. According to this view, the model is already 'in' the theory, and the scientist's job is reduced to the (non-trivial) task of choosing the best machine or theory that is able to generate an appropriate model. Cartwright's own example from superconductivity, as well as other case studies in this volume, show convincingly that modelling is much more complicated than that. Her chapter puts to work the idea, also highlighted in Marcel Boumans's chapter, that modelling is a creative enterprise. Boumans introduces and illustrates two theses that are echoed at several other stages later in the volume. The first claim is the heterogeneity of elements that make up a model. The second thesis is that the ways in which the various ingredients are put together vary from case to case, and follow no general rule: there is no general recipe for model building. However, a proponent of the Semantic view could always reply to Boumans and Cartwright that irrespective of how models are actually built, each model will always be 'in the theory' because the theory simply *is* the set of (highly theoretical, interpretative and representative) models that make it up. We shall leave it to the reader to decide whether the Semantic view should be praised for its generality, or – as the authors in *MaM* seem to suggest – criticized precisely on the grounds that this alleged generality fails to distinguish between different kinds of models and the ways in which they are created.

In any case, the proposed functionalist account of models is original enough to be an important new contribution to the subject. It would have been better if the individual chapters had instantiated this account in a more coherent and systematic way. But even as they stand, each case study highlights some aspect(s) of this account. A remaining worry, however, relates to the philosophical implications of this account. It is one thing to describe how models function and to explain this function by means of their autonomy; and it is quite another thing to engage in the philosophical issue of how models are vehicles for substantive knowledge of the world. To be sure, some of the papers in the volume (e.g., the papers by Morrison, Hughes, Suarez and Cartwright) do deal with this philosophical issue. But there does not seem to be an informative, overall approach. In fact, there are conflicting views. For instance, Morrison downplays the distinction between theoretical and phenomenological models and argues that it is orthogonal to the issue of how realistic the representation of the model is (p. 63). She nonetheless stresses that models can offer substantive theoretical knowledge of the world. Cartwright offers a sophisticated view of how theories relate to

the world, one that makes the models carry the proper ontological commitments of theories, but stresses that this view is consistent with the thesis that theories are 'warranted by their empirical successes' (p. 259). But Suarez seems to differ. After introducing a distinction between 'degree of confirmation' and 'degree of confidence', he offers an anti-realist gloss of the connection between theories and models. Unlike Morrison (and perhaps Cartwright), he seems to restrict the knowledge that models offer to 'the phenomena' (p. 195). Or take Boumans. Although he takes it to be the case that models have 'in-built' justification, he never raises the issue of whether this is genuine justification at all. He claims that his exemplary models 'were satisfactory to the model builders' (p. 95), without stepping up to the normative level. Klein and Morgan focus on learning from experimentation and manipulation of material or quasi-material models. One of the models examined by Morgan provides information concerning what policy makers should know and what they should be able to control in order for some intervention to be feasible and effective. Reuten's discussion of Marx's 'knife-edge' caricature similarly stresses the counterfactual (and mostly negative) nature of the knowledge provided by economic models. This diversity might seem natural. After all, one would expect that a functionalist account of models should be neutral about what kind of knowledge models offer. But since it is part of the functionalist account that 'models are both a means to and a source of knowledge' (p. 35), one is led to expect some general account of what sort of knowledge this is. Even if *MaM* aims to show how different models provide different kinds of knowledge in a truly pluralist vein, one might have expected some unity behind this pluralism.

This tendency to programmatically elude standard philosophical questions is sometimes frustrating, but in spite of that (or perhaps *by virtue* of it) the project undertaken in *MaM* is highly interesting and suggestive. The case studies are as detailed and realistic as those in the best relevant work in the philosophy of science. It opens up new ways of thinking about models and their relation to theories, and promises that further relevant work will cast more light on the central philosophical issues on which the project focuses. There is no doubt that all those who think about (or work with) models will learn a great deal from the book.

Francesco Guala

University of Exeter

and Stathis Psillos

University of Athens