principle are logically independent. This is an important point, but one that a surprisingly small number of philosophers recognize.

I have pointed out two relative disagreements with Holtug's arguments. But these two points do not undermine the comprehensive theory of justice that he puts forward. Holtug's *Persons, Interests, and Justice* is a magnificent book. It exhibits a most thorough analysis of the most difficult problems in contemporary moral philosophy. His vigorous project deserves the highest acclaim. Any moral philosopher, actual or possible, ought to examine this book carefully.

Iwao Hirose
McGill University

doi:10.1017/S0266267112000077

Rational Choice, Itzhak Gilboa, MIT Press, 2010, xv + 158 pages.

Itzhak Gilboa's book aims at introducing rational choice theory to a readership without prior knowledge of the field. It presents the fundamental ideas and concepts, not, or very little, the mathematics behind them. Gilboa gives priority to intuitive explanations and illustrative examples. The mathematical details are relegated to the online Appendix.

Rational Choice Theory encompasses decision theory, game theory and social choice theory, as practiced today not only in theoretical economics, but also in computer science, logic and philosophy. One of the main assets of the book is in fact to show how close the concerns in rational choice theory are to those in many areas of philosophy.

This is a very good book, but one that should be 'Rated PG', or rather LG for 'lecturer's guidance'. The highly pedagogical introduction to the key concepts of Rational Choice Theory touches a lot of fundamental questions in philosophy, but there are not enough references for the student or the newcomer to see the points of contact clearly. This review tries to bring some of them into light.

Individual Decision Making: The first part of the book touches issues classically pertaining to Decision Theory. Chapter 1 introduces the basic concepts, the most crucial one being, unsurprisingly, rationality. Gilboa's claim that rationality, as understood in rational choice theory, is also accepted by 'most psychologist and behavioral decision theorists', is refreshingly controversial, and to a large extent correct, although it doesn't quite do justice to the literature on 'ecological rationality' (cf. Gigerenzer and Selten 2002). His own view on rationality is subjective and dynamic: 'a mode of behavior is rational for a given person if [he] feels comfortable with it, and is not embarrassed by [read here: would not be willing to

change] it, even when it is analyzed for him.' There is for course a lot to be said here in connection with the rich literature on rationality in meta-ethics (see for instance Way 2010 and references therein), especially on the distinction between state and process rationality (see the remarks in Kolodny 2005). The section on uncertainty makes another interesting bridge to action theory, via the notion of *attempts* (the classical reference here is O'Shaughnessy 1973), although without any references to this literature. The last section of the chapter claims that rational choice theory is, in fact, not a theory but rather a *paradigm*, that is 'a system of thought, a way of organizing the world in our mind'. Seen as such, Gilboa claims, rational choice theory is much less prone to refutation by empirical evidence. This is quite an interesting perspective, that runs throughout the whole book, and would deserve more thoughts from the perspective of philosophy of science.

Not that many authors would dare to introduce utility maximization through a dialogue. Gilboa does it in Chapter 2, and he does that well. Besides managing to introduce the basic idea *and* representation theorems (informally, of course), the dialogue defuses widespread misconceptions of utility and maximization, and stresses its understanding in terms of coherence, viz. transitivity and completeness, of choices. Gilboa then sketches the three main views on how to interpret utility maximization: the normative, descriptive and what he calls the 'meta-scientific' interpretation.

The exposition of the normative view is rather weak. One is left with the impression that representation theorems are not much more than rhetoric devices, helpful to 'convince decision makers [. . .] that we would actually like to behave in accordance with a particular decision model'. This is probably as strong as one can get given Gilboa's subjective view on rationality, but this only shows the weakness of this view, from a normative perspective. The discussion in this subsection can nevertheless serve as a good entry point to the extensive current philosophical literature on the normativity of rationality.

The presentation of the descriptive interpretation touches the perennial question of the role of theories in the social sciences, but also the more recent one of the role of intuitions in mathematical modelling.

By the meta-scientific interpretation Gilboa means that 'theoretical terms' regarding preferences get 'mapped' to other terms referring to choices. The idea is not new, and this is one of the very few instances in the book where philosophy of science is explicitly mentioned – although without references for further reading. The last two sections, on Measurement and Disutility, contain an excellent introduction and discussion of the question of uniqueness of utility functions.

Chapter 3 reads as a short postscript to the previous one, discussing constrained maximization. The exposition is clear and pedagogical,

but it goes rather quickly over two important issues: choices under unresolved conflicts (cf. Levi 1986, for instance), i.e. constraints that are not jointly consistent, and the role of mathematical models in social-scientific enquiry.

The presentation of maximization of *expected* utility in Chapter 4 extends on the themes developed earlier. The chapter opens with a honest discussion of what can be simply rephrased as the 'why be rational?' question, taking 'rational' as maximizing expected utility. Gilboa's answer is classical: at face value, there is no clear normative reason for being an expected utility maximizer, except in case you consider the classical decision-theoretic axioms on preferences to be normatively appealing. He offers no argument for the normative character of these axioms, though. The choice of including a section on prospect theory could suggest some form of scepticism on that issue, but the discussion here remains, rightly, at the descriptive level.

Chapter 5 covers an impressive number of much-discussed questions in epistemology and philosophy of science: the interpretation of probability, the use of probabilistic models to represent beliefs, as well as the difference between statistical and causal inference. It opens again with a dialogue, although less convincing than the one in Chapter 2 – but the topic is more controversial, and the protagonists themselves are struggling to find a common ground. The presentation of the notion of relative frequencies contains interesting remarks regarding the role of idealizations in the social sciences, and the discussion of subjective probabilities does well at making the connection with the normative issue of coherence in attitudes. The section on 'Statistical Pitfalls' is a gem of common errors debunking. A must-read for students. Finally, the discussion of statistical vs. causal inference rightly highlights the pragmatic relevance of both types of reasoning.

Group decisions: In chapters 6 and 7, Gilboa leaves individual decision making and starts discussing group-related notions. Chapter 6 is on attitudes aggregation. He starts with summation of utilities, not because of its intrinsic interest it seems, but rather because it allows him to illustrate some of the canonical issues for social choice: the problem of non-uniqueness of utility measures when the latter are lifted to groups, manipulability of social preferences, and interpersonal comparisons of utilities. The first two are well illustrated by a simple example, and the second, together with the Gibbard–Satterthwaite's theorem, are treated in greater detail later. Classical problems for aggregation are introduced through the non-less classical Condorcet Paradox, leading to a very pedagogical presentation of Arrow's Theorem. The discussion of the Unanimity Axiom is a bit too quick at concluding that 'it is not clear on what grounds [the choice of unanimously dis-preferred option by a social planner] would be justified'. Cases like the Prisoner's Dilemma

come immediately into mind here, and they are in fact implicitly discussed at the end of the chapter. The other Arrow conditions are admirably well explained, and Gilboa even manages to give the reader the gist of the proof of the impossibility theorem. The end of the chapter turns to the notion of Pareto efficiency, and here Gilboa does again very well at debunking common misconceptions about this notion, and at honestly showing its limitation. In particular, the emphasis on the non-completeness of Paretian orderings should be very useful for newcomers. Finally, it is worth noting that the very last paragraph, on aggregation in the face of uncertainty, touches on important issues for the philosophy of collective agency, especially on the possible tension between individually and collectively desirable outcomes.

The chapter on games (7) starts with an extended discussion of the Prisoner's Dilemma. Gilboa's view on the 'solution' of that game is classical, with a twist. As far as individual rationality is concerned, he holds that defecting is the only rational move. Furthermore, he stresses that attempting to find a solution *to this very game* by changing the individual payoffs misses the point. But here Gilboa takes this classical view one step further, in maintaining that this view doesn't dismiss transformations of the game as uninteresting or some form of cheats. Quite the contrary, he holds that the real import of the Prisoner's Dilemma is precisely to unveil the importance of the 'design of social situations' for achieving socially desirable outcomes. In cases where individual and group interest conflict, it is 'very dangerous to assume that altruism will take care of the problem'. Changing the game is the way to go or, in more trademarked terms, solutions to such conflict lies in mechanism design (cf. Myerson 2008 for a recent survey) or social software (cf. Parikh 2002) – although none of the bodies of literature on these topics are mentioned in the section. All in all, Gilboa's view of the Prisoner's Dilemma has the merit of taking conflicts between individual and group interest seriously, while being compatible with many solutions incorporating group-oriented types of reasoning.

The discussion of the Prisoner's Dilemma is not only meant to present Gilboa's own views on the topic, it also does well at introducing two solution concepts for games: strict and weak dominance. Gilboa's explanation hints at the epistemic characterization of the former (for more references on epistemic characterizations of solution concepts, see Brandenburger 2007), and at the equivalence between the later and best response against full support probability distributions, but again explicit references to these results are nowhere to be found in this section – some of them come only later, in the discussion of rationalizability. Gilboa's presentation of these solution concepts could nevertheless serve as a good starting point to the extensive literature on social norms (see references in Bicchieri and Muldoon 2011) and even contains a few paragraphs on the

Categorical Imperative, quite an unusual, but very pleasant encounter in a book on Rational Choice Theory.

The section on Nash equilibrium also reaches out for a number of fundamental issues. As before, the epistemic characterization of this solution concept (see again references in Brandenburger 2007) is hinted at, but without references. The discussion is however very helpful in laying down notions such as repeated games, mixed strategies, both in their objective and epistemic interpretations, as well as the infamous equilibrium selection problem. The collection of examples provided towards the end of the chapter is most illuminating, and the discussion of commitments in games takes a classical perspective in game theory (Schelling 1960) but one that would deserve more attention in the philosophical literature.

The chapter finishes with a short section on common knowledge, this time including the classical references on the topic – Lewis, Aumann, and further work in epistemic logic – together with a short excursion in extensive games.

Spin offs: The last part of the book is called 'Rationality and Emotions', but in fact it is more an exploration of various issues that arise once the conceptual machinery presented in the previous chapters is in place.

Chapter 8, on free markets, shows well the connection between the foundational issues introduced in the previous chapters and more classical results in economics. Once again Gilboa uses a dialogue to set the stage. This time he makes the characters debate the pros and cons of globalization. The lesson here is rather 'keep reading', a wink at the numerous misconceptions around on the topic. The first welfare theorem is well explained, and the extended discussion that follows, on the assumptions and limitations of the result, reiterates important points made earlier regarding the meaning of Pareto efficiency. Unfortunately, the presentation of the rationality assumption stays at the descriptive level, and so misses what would have been a good opportunity to discuss the normative character of this notion.

Chapter 9 is a short excursion into the theme of emotions, essentially aimed at refuting the idea that emotions and rationality exclude each other. Gilboa does this by drawing from evolutionary models and from what are by now textbook examples: parents' unconditional love for their children, and anger in retaliation scenarios. This last example is explicitly connected with earlier discussion of extensive games, but has broader implications, notably in conjunction with the discussion on the 'power of commitment' in Chapter 6.

The tone of the penultimate chapter (10) is rather different from the other ones, as it reads as a focused attack on the Adaptation Level Theory of well-being. Gilboa's first criticism is methodological, regarding the difficulty of interpreting peoples' own reports on their preferences and

well-being. He puts forward, as an alternative, the behavioral take on preferences that underlies the classical decision-theoretic approaches. As such the debate stays in known territories, but is nevertheless worthwhile in raising fundamental issues for the philosophy of science. His second criticism bears on the normative conclusions drawn from Adaptation Level Theory, namely that we should 'step out the hedonic treadmill'. Gilboa rightly points out that this recommendation is not unlike the one of cooperating in the Prisoner's Dilemma, and goes on to conclude that this might not be recommendable after all. Notably absent from this discussion, unfortunately, is the lesson that Gilboa himself draws from the Prisoner's Dilemma in Chapter 6: that maybe we could get out of the treadmill by clever 'design of social situations'.

All in all, Gilboa's book manages to get through an impressive amount of material, in a way that is both accessible and entertaining. As such it is highly recommendable for an undergraduate course on Rational Choice Theory, with the 'Lecturer's Guidance' proviso that I mentioned in the introduction.

The most challenging claim, philosophically speaking, is probably the one that Rational Choice Theory is *not* a theory but rather a paradigm, 'a way of organizing the world in our mind'. Whether this is or will prove true depends on how successful is the Rational Choice Paradigm at doing just that, organizing the world in *our* mind – the mind of people from a broader audience, not only of economists, trained in the discipline. If anything, such a well-written, accessible book on the topic is probably the best way to make this claim self-fulfilling.

Olivier Roy
Ludwig-Maximilians-Universität München

REFERENCES

Bicchieri, C. and R. Muldoon 2011. Social norms. *The Stanford Encyclopedia of Philosophy* (Spring 2011 Edition), ed. E.N. Zalta. http://plato.stanford.edu/

Brandenburger, A. 2007. The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory* 35: 465–492.

Gigerenzer, G. and R. Selten, eds. 2002. *Bounded Rationality; The Adaptive Toolbox*. Cambridge, MA: MIT Press.

Kolodny, N. 2005. Why be rational? *Mind* 114: 455.

Levi, I. 1986. *Hard Choices*. Cambridge: Cambridge University Press.

Myerson, R.B. 2008. Mechanism design. In *The New Palgrave Dictionary of Economics Online*. <http://www.dictionaryofeconomics.com/dictionary>

O'Shaughnessy, B. 1973. Trying (as the mental 'pineal gland'). *Journal of Philosophy* 70: 365–386.

Parikh, R. 2002. Social software. *Synthese* 132: 187–211.

Schelling, T.C. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Way, J. 2010. The normativity of rationality. *Philosophy Compass*, 5: 1057–1068.