Erasmus Journal for Philosophy and Economics,
Volume 7, Issue 2,
Autumn 2014, pp. XX-XX.
http://ejpe.org/pdf/7-2-int.pdf

Rational choice as a toolbox for the economist: an interview with Itzhak Gilboa

CATHERINE HERFELD
Ludwig-Maximilians-Universität München

Itzhak Gilboa (Tel Aviv, 1963) is currently professor of economics at the Eitan Berglas School of Economics at Tel-Aviv University and professor of economics and decision sciences at the Hautes Études Commerciales (HEC) in Paris. He earned undergraduate degrees in mathematics and in economics at Tel Aviv University, where he also obtained his MA and PhD in economics under the supervision of David Schmeidler. Before joining Tel Aviv University in 2004 and the HEC in 2008, Gilboa taught at the J. L. Kellogg Graduate School of Management at Northwestern University, the University of Pennsylvania, and Boston University.

Gilboa's main area of interest is decision-making under uncertainty, focusing on the definition of probability, notions of rationality, non-Bayesian decision models, and related issues. He has published broadly in areas such as decision and game theory, microeconomics, philosophy, social choice theory, and applied mathematics. He has written over 90 articles in these fields. Gilboa has furthermore written a textbook entitled *Rational choice* (Gilboa 2010a), in which he lays out what he takes to be the main toolbox for studying and improving human decision-making. Additional books include *A theory of case-based decisions* (Gilboa and Schmeidler 2001), *Theory of decision under uncertainty* (Gilboa 2009), *Making better decisions* (Gilboa 2010b), and *Case-based predictions* (Gilboa and Schmeidler 2012).

Professor Gilboa was interviewed by Catherine Herfeld at the department of economics of the University of Mainz (Germany) on July 13, 2013. In this interview, Gilboa lays out his perspective on the nature and purpose of the rational choice paradigm, discussing it in the context of recent philosophical questions about the advantages of axiomatization and its relation to empirical research, the usefulness

NOTE: Catherine Herfeld is currently a postdoctoral fellow at the Munich Center for Mathematical Philosophy, Germany. Before coming to Munich, she was a research fellow at the Center for the History of Political Economy (Duke University), while finishing her PhD at Witten/Herdecke University in the history and philosophy of economics. This interview is part of a larger project entitled 'Conversations on rational choice theory', which aims at critically discussing the different manifestations of rational choice theory and the ways in which they have been used in philosophy and economics. Contact email: <Catherine.Herfeld@lrz.uni-muenchen.de>

of unrealistic assumptions, the future of neuroeconomics, the status of economics as a science, and his view of truth.

CATHERINE HERFELD: ***Professor Gilboa, you are currently professor of economics and decision sciences at Hautes Études Commerciales de Paris. What, broadly speaking, are the decision sciences?***

ITZHAK GILBOA: 'Decision sciences' is a general term. As I understand it, 'decision sciences' refers to the field of decision in general. 'Decision sciences' encompasses decision theory, applied work, and experimental work. But the field of decision theory today is starting to undergo a process of 'disintegration'. I do not want this to sound bad. This happened to game theory about 15 years ago. Both in game theory and in decision theory there is a general paradigm that is extremely beautiful and extremely insightful, and which has a lot to say about almost everything. But this general paradigm is at some point exhausted, and you start having to commit yourself to a specific type of theory you work with. And then you might find out that the theory you work with is not as general as the paradigm.

Could you illustrate this view with an example?

For instance, consider game theory, which is not too far away from the field of decision sciences, where you have a general approach to human interaction. You can model a wide variety of situations: you identify players and strategies to begin with, and then you have some things to say about the interaction. For example, the concept of Nash equilibrium, or maybe even that of a perfect equilibrium, allows you to say something insightful about everything that can be modelled as interaction among decision makers. It could be the interaction between couples, like battle of the sexes; it could be the interaction between two countries; it could be the interaction between buyers and sellers in a market; or it could be the interaction between species, such as in hawk-dove games. Surprisingly, game theory can capture those different situations and can say something meaningful about each of them, which is fantastic. But at some point, when you start looking at refinements of, for example, Nash equilibrium, or dynamics that would or would not lead to Nash equilibrium, you would say: 'Wait a minute. The dynamics that would capture evolution are not the same sort of dynamics that are applicable to the market'.

So game theory constitutes a general paradigm, offering a set of theories and concepts that become modified for particular applications within this paradigm. In what way does this happen with decision theory?

I think that in decision theory we see something similar. The general approach, taking Leonard Savage's work as being the main achievement along those lines, could be used to think about any kind of problem under uncertainty (see Savage 1972 [1954]). We should identify states of the world, acts, and outcomes, and can use concepts such as probability and utility. As such, this general approach always has something meaningful to say about decision-making. But sometimes what it says is not enough. And when you start going beyond that, you might have to decide what exactly you want to apply this paradigm to, that is, whether what you understand as an 'application' is, for example, the situation of sitting down with a patient who has to decide whether to undergo surgery; or whether an 'application' is a purely theoretical model in applied theory. When you are sitting there with a patient who has to decide whether to undergo surgery or not, you have to estimate actual parameters; you have to take this general approach seriously. You also have to take it seriously when an application for you is pointing out to a colleague who is doing a search model in labour economics: 'Wait a minute. Maybe you do not get this sigma, let us think about uncertainty instead of risk'. Both are called applications, but they are both very different things. And it is not at all obvious that the same paradigm, or conceptual framework, provided by decision theory is going to be relevant for both kinds of application to the same degree. In short, there is something common about the paradigm that is relevant to everything, but if you actually want to do something concrete with it, then you might have to commit to the kind of application you have in mind in order to capture the specifics of this particular application. This is what decision sciences and game theory have in common.

As such, 'decision sciences' encompass many things. They include theoretical work and experimental work. Even within theory, one probably has to decide whether one is developing a theory to be used by theoretical economists, by empirical economists, or to actually make decisions about whether we should build nuclear plants, take a specific drug, or whatever.

Can we say that you understand the relation between the general paradigm of decision sciences and the theories formulated and applied within this paradigm along Lakatosian lines: we have a research program consisting of a hardcore (what you call the paradigm) that remains untouched, but allows for formulating specific, falsifiable theories to address a variety of concrete problems?

Yes, that is right. To a large extent I see such a process along Lakatosian lines. Yet this process is not always accompanied by sufficient self-reflection in the field. I think that decision theorists tend to think of decision sciences as providing a particular theory, not a general paradigm. As such, the distinction between the general paradigm and the theory is not always sufficiently clear.

This sounds similar to the warning you voiced in your article entitled "Questions in decision theory" (Gilboa 2010c), where you also talk about the recent "soul searching" occurring in the decision sciences. Could you say a bit more about how the paradigm and the theories exactly relate to each other and why you think the distinction between a paradigm and an application is crucial?

The theories are obtained from the paradigm by two main processes. First, there is a specification of terms. For example, when I think of a 'player' in a game, it can correspond to a person or to a nation in a given reality. I can, for instance, decide to model an interaction in which the U.S.A. is a player, or to take the same interaction and think of the President of the U.S.A. as one player, Congress as another, and so on. We are confronted by similar modelling choices when we think of terms such as 'state of the world', 'time period', 'strategy', 'outcome'. Hence, the same paradigm allows for a host of different theories, all compatible with it, for the very same real-life application. Second, there is a process of tweaking and generalizing a theory within the same general paradigm. For example, expected utility theory suggests that payoffs are aggregated by mathematical expectation, and someone, say Kahneman and Tversky (1979), may propose that probabilities that are close to either 0 or 1 are 'distorted'. In and of itself, this generalization can be viewed as a newer theory within the very same paradigm. Note, however, that other ideas of these two scholars broke from the standard theory in more dramatic ways.

Keeping the difference between the paradigm and the theory is then crucial in appraisal for example when we discuss 'theories', whether

they 'work' or 'fail', and so forth. We should be careful to distinguish between different interpretations of the same mathematical models. Often the models of rational choice can be interpreted both as theories (coupled with auxiliary assumptions) and as paradigms, and often the empirical failure of the former does not imply that of the latter. This is why we should keep them strictly separate from each other.

Your own work in decision theory mainly focuses on applications of the paradigm in epistemology—questions about belief revision, statistical decision-making, etc.—and in philosophy of science—taking the main goal-directed activity to be scientific inquiry. What are the potential uses of decision theory for the social sciences?

I think that decision and game theorists are often interested in one of the most fundamental problems in the social sciences: How do people think, and how should they be thinking? As this is also a major concern for philosophers, it is why I also feel that parts of philosophy are a social science, especially if we focus on the normative question of how people should be thinking. When we take a more descriptive interpretation, we are closer to psychology and to its applications in behavioural economics. In these applications, there is a focus on ways of thinking that might be simply erroneous and that are not very useful for philosophy of science or statistics. But when we take a normative approach, asking ourselves how rational agents should be thinking in social set-ups, we are basically asking the question that a statistician asks when she wonders what can be inferred from the data, or that a philosopher of science asks if he takes a normative approach. For example, the preference for simple theories is considered to be an important criterion for the selection of theories (ever since Ockham), and it is correspondingly an important criterion for 'model selection' in statistics. The two strands of the literature differ in many ways, but they are asking the same fundamental question. Therefore, it is not too surprising that similar ideas have been developed in these fields.

Decision sciences have experienced an enormous expansion in the last decades, especially in the social sciences. Why do you think the representation of individual decision-making became so important, and how much does a mathematical theory of decision-making really matter in the social sciences? Economics, for example, seems to be at

least equally concerned with understanding macro-phenomena as with explaining individual behaviour.

First, I am not sure what economics is concerned with, nor what it should be concerned with. For instance, if you are interested in predicting how many kids seek college education, or what it might take to change that number, you need decision theory more than macroeconomic theories. Second, if you are dealing with questions on the macro-level such as central bank policy or bank runs, you are interested in game theory and its decision theoretic foundations. This would certainly be the case if you are interested in, for example, the possibility of a war because of its effect on macro variables and on financial markets: whether a certain country will wage war on another is not a classical economic question, and it is one that decision sciences can help to analyze. In fact, it is quite difficult to talk about economics (or political science, for that matter), without the shadow of decision sciences hovering above your head. Consumers and firms, governments and politicians, traders and bankers do just that: they make decisions.

But let me draw the link between the issue of the importance of a decision theory and axiomatization in economics. I find that the choice of the word 'representation' in your question is quite revealing, and this is where decision theory might be more important than one would expect: when one writes a model in macroeconomics, finance, or labour economics, whether the work is theoretical or empirical, there is often a need to model decision makers. Actual parameters may be assumed or estimated, but one needs a general framework into which parameters can be plugged. And this is why representation becomes important. It is a bit of a paradigm, as the specific parameters, defining a theory, are not yet specified. For that reason, one cannot yet test the representation, at least not in its intended use. One can test something similar to it in an experiment, but this gives rise to external validity issues. As a result, axiomatic work becomes more important: it is a way to convince scientists, who have not yet developed their economic theories, that a particular paradigm may be more useful to adopt and plug into their models.

Your view on separating paradigms from theories has several implications, not only for the question of how we appraise decision theory, but also for the assessment of new fields such as neuroeconomics and experimental economics. Those new fields

provide new ways of addressing decision theoretic questions. But you also observe that the question emerges: "what would be the right mix of axioms and theorems, questionnaires and experimental, electrodes and fMRIs?" (Gilboa 2010c, 2). What do you think those new branches can contribute to the field of decision sciences?

We can observe that many more people are now interested in neurological research, which is a good thing. At least if we ignore the moral dilemmas posed by neuroeconomics (both in terms of animal studies and in terms of alternative medical uses of equipment), it is a wonderful thing that we can know more about the human brain while making decisions. There is much more opportunity to connect between psychologists, neuroscientists, economists, decision theorists, game theorists, etc. And so far, neuroeconomics has been very exciting and I think that some of the questions that neuroeconomists pose are worthwhile.

However, I suppose that at some point the communities of economics and neuroeconomics will disintegrate, or separate. Neuroeconomics as a community is probably going to flourish, not necessarily within economics, but maybe in psychology. It is just not obvious that neuroeconomics is the best use of resources if we want to solve economic problems. And I do not think that neuroeconomics makes economics any more promising than it was before there was neuroeconomics. For the time being, there seems to be a very large gap between what we know, what we can possibly measure, and what we need to know about the brain in order to deal with economic questions. I also think that more and more people are very sceptical of the reductionist idea underlying neuroeconomics. In short, neuroeconomics may be a very respectable field within psychology or neuroscience, but it is not necessarily changing the way we do economics. And I thus, in all likelihood, I think that these two fields of research—economics and neuroeconomics—will remain separate for decades to come.

But you also seem to think that new fields, such as neuroeconomics, provoke novel questions for theoretical decision theory and that we can try to use decision theory to formulate those problems in a better, more precise, way. How fruitful do you consider attempts of behavioural economists or neuroeconomists to axiomatize their findings to reach a higher level of precision in formulating those new problems?

I make a living out of axiomatization, so I should not say anything bad about it. But people in the field are often not sufficiently self-reflective. Let us take an extreme view against axiomatizing scientific theories: some people would ask why one would axiomatize a theory at all. A theory should primarily match the data and thus we should first see if it in fact does so. Proving a characterization theorem that shows the equivalence of one formulation of a theory to another formulation does not, by definition, prove anything about the data. The two formulations will be just as close to, or far away from, the data. So you have to ask yourself: 'Why am I doing this and what purpose does it serve?'

But from my point of view, it is more important to axiomatize paradigms than theories. With respect to the axiomatization of a paradigm, I can tell a coherent story of scientific development where axiomatization would play a major role: scientists are dealing with various problems and we could help them. For example, many economists are developing models. And the questions that we can address are: 'What models and theories should they be using?' and 'In which language should they be formulating their models, when they develop them?' Here, an axiomatization can help. And in such a way it could also help in behavioural economics and neuroeconomics. But I do not think that these fields have yet developed such a paradigm. To the extent the behavioural economics has a paradigm, it seems to be the same rational-choice paradigm of economic theory.

So, what exactly is the purpose of axiomatization in economics? And what is the role of the decision scientist in this context?

Imagine that, within their own discourse, some economic theorists cannot answer the question of which language they should use to develop a theory, or which paradigm they should use. This problem arises because they cannot yet compare the theory or the paradigm in question to the data because they have not gathered them yet. Post hoc economists could say: 'Ok, this paradigm was great; it has allowed us to develop all these theories and explained all those phenomena'. But before that, they cannot resort to the data to help them convince each other. What decision theorists can do is use an axiomatization basically as a rhetorical device that says: 'If you find these axioms reasonable, you should find the implication derived from those axioms reasonable'. In other words, they can try to convince the economist to use the mathematical results, which are sometimes useful.

Here comes fancy math, or at least surprising math, which shows you things that are not obvious. If you can imagine a very convincing but complicated proof that takes you from this set of axioms to this theorem, as for example Savage (1972 [1954]) or John von Neumann and Oskar Morgenstern (1947) did, then axiomatization is more powerful because it is not saying something that is obvious to see, but that is mathematically correct. But as mentioned before, I think that sometimes people in the field are not sufficiently self-reflective. While we should develop axiomatic theories, we should be more sensitive to the question, why we are doing that; why we play the game; whether we are really trying to be logical positivists; whether what we are trying to do is descriptive or normative; and what the role of axiomatic work is for realizing what we are trying to do.

What you seem to say is that axiomatization might be useful in economics when we do not yet have a theory, when we want to derive specific, maybe surprising, conclusions from a set of axioms, because these conclusions might be hard to reach without axiomatization. But we can then subject the conclusions to empirical testing. You also seem to suggest that once we have a theory, we should take seriously the idea that it should explain the data. In order to fulfil those two roles, to what extent should axioms be inspired by reality?

Yes, that is right! And I think that the important thing here is what we call 'intuitive' or 'natural', something that informs our axioms but is not necessarily related to a particular example or to a concrete empirical observation about how human beings in fact behave.

So, is it sufficient for a good theory of decision-making that the axioms appear intuitively or naturally plausible?

When you think about axiomatizing a paradigm and not a specific theory, a good decision 'model'—let us avoid here the word 'theory'—is one that will be relatively abstract and sufficiently general, so that I can use it to think about many specific theories. For instance, we can think in a rather general way that there is an outcome. I can specify what that means by thinking in terms of a particular example: you give me 100 Euros. That is an outcome. But I can think of other definitions of 'outcomes' that would include psychological and social payoffs as well, such as 'getting 100 Euros when my friend got 1000' or 'accepting 100 Euros when my friend was exploited by strategic weakness', and so on.

Clearly, this re-definition of an outcome can result in theories that belong to the same paradigm, but have different predictions. The same would be true of other conceptual aspects of a general paradigm, such as 'player' and 'strategy', 'state of the world' or 'time horizon'. Good models will typically need to be both abstract and convincing, where they can be convincing either because they are intuitive, or because they are mathematically derived from intuitive conditions (axioms). A model needs to be abstract to allow for a range of applications, for many specific theories, for us to feel that we have a general tool that can address many issues. It has to be intuitive for us to believe that these applications, often not yet developed, have a chance of making sense, explaining data well, and generating good predictions.

But maybe some sort of abstraction is sometimes to be trusted more than focusing on a particular experiment that I just observed. Sometimes the examples that a group of scientists starts off with, especially if they have a particular experiment in mind, may be a somewhat biased basis for generalization. There are situations in which, when we think about them in the abstract, we might get a better global view of what is going on in such situations rather than if we get into the details. If you focus on one particular example in greater detail to subsequently generalize, that might affect your generalization; things present in the example might look bigger and more important to you than they really are, and sometimes you might get a better idea if you zoom out. The standard examples are availability heuristics: you estimate the probability of an event and you split it down to a couple of events and get a bigger estimate. That is a case where thinking more is not necessarily thinking better, because you end up with something that might be a worse example. For instance, we can talk about the probability of your car being stolen and elaborating on various scenarios. And, when you give me an estimate at the end, this estimate might be worse than what you would have given me based upon your overall intuition. I think something similar can happen when we think about an abstract problem. Of course this has to do with the external validity of the experiment, and if I am interested in something that is extremely close to that experiment, then it might be fine. But if I start by looking at an experiment and then I switch over to talk about how people behave in markets, even when this experiment was conducted in social psychology, I might get this problem.

To make a long story short, sometimes we can overly generalize. So I would not insist so much on the idea that axiomatizations should be inspired by actual experiments, and on a very close relationship between empirical observation and axiomatization. Rather, as I said, I think axioms should be mostly generalizable and acceptable to us in an intuitive way.

So what role should examples play then and which examples do you consider useful?

We should not ignore examples, as it is fine to be motivated by an example. But I think that one must be able to understand the example as a sort of paradigmatic example, something that can be easily generalized so that one sees exactly what general point it makes.

Could you give a particular instance of such a paradigmatic example?

Sure! When David Schmeidler (1989) began with his research on probability and expected utility, he was not motivated by Daniel Ellsberg's (1961) experiments. He looked at the Bayesian theory and found that it is too limited to capture uncertainty, especially when one is in a condition of ignorance. Schmeidler gives the example of the coin that comes out of his pocket that he has tested many times and the coin that comes from someone else's pocket that he knows nothing about. I think the example is sort of paradigmatic. It is intuitively also quite convincing. It turned out to be very similar to Ellsberg's two-urn experiment, but it was a mere example of a very general difficulty. By contrast, when people looked at Ellsberg's experiment itself, I think that they had a tendency to develop theories that were much less generalizable.

In your own work, empirical observation and axiomatization are closely related. Even when using a theory for prescriptive purposes, you consider the descriptive dimension relevant. For example, in your article "Rationality of belief or: why Savage's axioms are neither necessary nor sufficient for rationality" (Gilboa, et al. 2012), you praise the flexibility of the rational choice paradigm compared to previous attempts in economics to conceptualize human behaviour. The notion of rationality that you refer to is basically defined as consistent choice. And obviously you use the axiomatic method to formulate this concept. But your definition of rationality is also

inspired by observed deviations from the rationality axioms. You stress that we should not think about rationality as detached from reality. Even when used as a normative concept, observing actual decision-making matters. Can you expand on your view about the usefulness of empirically inadequate axioms in both cases, i.e., for empirical and normative purposes?

I stress the practicality of a decision theory in my work. Axioms or axiomatic theories in the social sciences have double lives. You try to use the theories that rest upon a set of axioms as positive theories and if they do not work, you try to sell those theories as normative theories. But even for their normative use, theories should be practical. As Reinhart Selten once informally said, a normative theory that says you should run 100 meter in 4 seconds is not very useful, because you simply cannot do it. As such, the theory does not allow for a practical prescription. When you think about a normative decision theory, it is important to think about the practical behaviours that could be selected by decision makers. That is what I wanted to capture by the notion of rationality that I have articulated.

Could you give an example to introduce the idea of regret that characterizes your definition of rationality?

For example, is it rational to make calculation mistakes? To answer this question, I ask: 'Would you be embarrassed if I were to show you that you do not calculate correctly?' Well, if it were the case that the calculation was too complicated to be performed correctly, you would probably not be embarrassed. If it is impossible for any human being to calculate correctly, you could respond: 'How could anybody have known?' In that case, your behaviour is rational according to my definition. It is consistent, or robust to our analysis, in the sense that preaching to you to behave differently would be useless. For the sake of usefulness, we need to somehow place practicality into the picture when endorsing normative theories.

In your article entitled "Is it always rational to satisfy Savage's axioms?" (Gilboa, et al. 2009, 289) you write: "The question we should ask is not whether a particular decision is rational or not, but rather, whether a particular decision is more rational than another. And we should be prepared to have conflicts between the different demands of rationality. When such conflicts arise, compromises are called for.

Sometimes we may relax our demands of internal consistency; at other times we may lower our standards of justifications for choices. But the quest for a single set of rules that will universally define* the *rational choice is misguided". So you formulated a definition of rational choice that weakens the idea of a unique standard of rationality...

My definition of rationality started with this: asking what do people have in mind when they refer to something as 'rational'. But the best definition I came up with is in terms of what most people would be willing to accept as their decision making modes, as opposed to what they would like and could change.

Is this definition of a rational choice pragmatically useful for improving one's decision-making?

Yes. I think that we can play around with definitions to our heart's content, and judge them for elegance and beauty of results as we do in mathematics. But to the extent that we care about a particular definition—say, what is and what is not called 'rational', or, for that matter, 'scientific', and the like—we should ask what it is exactly that we care about in specific definitions and then choose them accordingly. Why do we bother to dub some modes of behaviour as 'rational' or 'irrational'? If it is only a matter of name calling, then it is not so clear that it is worth the effort. Rather, we need to think about what kind of discourse we have in mind that might be facilitated by a specific definition. Indeed, this boils down to a pragmatic position.

How does the mathematics enter this picture?

Part of what happens is that the way people choose is an issue of 'either-or'. Either a person makes decisions in a completely intuitive way, or she makes decisions in a supposedly rational, but at the same time highly mathematical way. People tend to view these two things, mathematics and rational choice, often as going hand in hand. People who are scared of mathematics often tend to not even listen to what insights there are behind the mathematical apparatus. Rather, they tend towards the other extreme, that is, they fully reject mathematical theories of decision-making. But one does not have to be scared of the mathematics. Good decisions, be that for individuals or for society as a whole, should involve some kind of dialogue between the theory and the

subjective input, be this an input that originates in emotions, intuitions, or something else.

And what do you take 'rational choice theory' to be in this context?

Well, it is a bit of everything. It is a toolbox and not everything in it is tightly related. It consists of a couple of useful things we find in decision and game theory, microeconomics, and so forth. Yet, they come from the same way of thinking, and they describe the state of the art in a way that is not too biased. I do think that these are very beautiful ideas that should be more publicly available.

You mention at some point in your textbook* Rational choice *(Gilboa 2010a) that you would like to live in a society where everybody knows about the tools in this book contains. Why?

Yes, I do indeed think that. People vote. People make decisions for themselves and for us, and they do it based on various pieces of information that they get. This information can be highly manipulable. For example, you hear that a certain percentage of inmates belong to a certain ethnic group and people in their minds begin to think that people belonging to this ethnic group must be criminals. In this example, people confuse the probability of A given B with the probability of B given A, a psychological phenomenon we understand very well. But we could teach people to become aware of this confusion and learn to avoid it.

Is improving decision-making your pedagogical aim when teaching with this textbook?

Yes, I think it is valuable to teach the tools in this book to everyone, including people who have no background in mathematics and no willingness to get into that. Ultimately, mathematics itself is not important for the public debate. It is often rather a sort of barrier to entry to many people involved in practical decision-making. What is important is to convey the basic messages and to have people participate in such public debates in a more educated way and especially to address the basic questions about what is feasible in achieving the desirable.

So, we should use the toolbox of rational choice as a normative instrument to make people behave rationally...

Yes, that is right. I think it could teach us to improve our reasoning and judgment, to think more critically about so-called experts, be it in politics, medicine, or whatever. Judging whether politicians are more or less successful could be done in a more rational way. I think I could even convince some of them that their way of making decisions can be improved (in their own eyes). I think it is essential to understand how to be rational in the context of economic and political questions, and to understand the nature of democracy in the context of the limitations of preference aggregation.

Another thing that often bothers me in the political domain is that people tend not to think in terms of what is feasible when they think about what is desirable. For example, in political debates, people sometimes reason by assuming that what they want is also possible. This would be considered flawed reasoning according to many rational choice models. Indeed, most people would not make this type of mistake in the context of, say, a financial investment. But when it comes to ideological questions, it is often a big no-no even to pose the question of feasibility.

In your account of rationality, you fuse several different dimensions of a theory of rational choice. You repeatedly talk about the trade-off between having a mathematically beautiful theory—one that people might not conform to—and a more descriptively accurate account of human behaviour. And, you use your toolbox of rational choice for prescribing the rational course of action. To make the mathematical theory more descriptively accurate, you can either change the theory—like behavioural economists do—or you can make people conform to the theory. You define rational choice as a choice where a decision maker does not want to change anything when confronted with a mathematical analysis of his or her behaviour; the decision maker might regret the choice in light of new information, but not as a result of a theoretical argument. This account of rational choice allows the study of how people in fact deviate from the prescribed rational choice and assumes that they would regret it. How exactly do you bring those different dimensions under one roof? How does this definition relate to what rationality is usually considered to be,

i.e., independent from subjective elements, arrived at by reason and rational calculation and serving as an ideal standard?

First of all, I do not have any bit of understanding of metaphysics, so I do not know whether anything exists, unless I know how to measure it. And that is partly why I appreciate axiomatic work, because it gives concrete meaning to things that can be very abstract. I can only understand metaphysical concepts when there are psychological manifestations thereof. For example, I can talk about free will, but I only refer to the psychological phenomenon that people seem to experience making choices and exercising free will. Whether they really make free choices, and what that would mean, are questions I do not fully understand. Taking this non-metaphysical stance, I do not know what exists out there that is 'independent from subjective elements', and I am not even trying to grasp it.

Second, I think of myself as a generally democratic, liberal person, and I do not think that I have the right to decide for people what to do with their lives, their children, their money, etc. I am in this sense, somewhat anti-elitist. If you ask me what we should do with taxpayers' money, what we should teach in schools, and so forth, I would give answers that I believe I can support based on these people's future well-being as perceived by them. For example, should we expose kids to Mahler's music at school? I would say 'Yes'. But I would say that not because I think Mahler's music is great, although it is, but because I think that, if you were to run experiments, many people would acquire the taste for it and think that it is the greatest music that exists. I am not sure if I can make the same case for Karlheinz Stockhausen. If we have a kind of music that remains something extraordinary only for a very small group of people, an elite, it is not clear why we should use taxpayers' money because it is not the case that most people would benefit from it.

My non-metaphysical stance and my democratic criterion form the background to how I address the question of rationality: I do not believe we have access to anything out there that determines 'true' rationality independently of human beings' judgment. And I do not think that a bunch of smart people should define what rationality is for the rest of humanity, whether the latter does or does not agree with it. I believe that, eventually, judgments of rational choice should go back to the people about whom we are talking, for whom we are making decisions, whose money we are spending. This is why my notion of rationality

is about explaining and convincing, and will eventually depend on the majority's view.

Yet, does your definition of a rational choice not presuppose that reasoning according to logical rules is of greater value than reasoning ignoring logic and that people would therefore regret violating the basic rules of logic?

Not necessarily: I am willing to subject logic to the same test. True, I think that most people would be convinced by logic, and, for example, be able to understand modus ponens, and feel bad about violating it. But this claim of mine is an empirical claim. If you show me ample evidence to the contrary, I will have to give up my faith in logic as a widely accepted form of reasoning. I will have to admit that the structures I like in reasoning are not necessarily shared by people in the society I live in. I hope I will not be caught saying that these structures have 'greater value' than other structures. Just as I hope not to be supercilious about the kind of music or literature I consume. At present, I do believe that many of the principles we preach will, given the exposure, be adopted by a large majority of people. But I should be ready to admit that I might be wrong about that.

Why, do you think, would people feel bad about violating the principles of rational choice?

I believe that we have immediate, affective responses to cognitive inputs such as logical reasoning. It is akin to, or maybe just a special case of, aesthetic judgments. Just as we can have positive or negative affective responses to a painting or a piece of music, we can have such responses to reasoning. I conjecture that, as an empirical claim, we are hardwired, by and large, to enjoy coherent reasoning and to be aversive to contradictions. I suppose that we will have to go to evolutionary psychology to answer this question. We can explain much of our aesthetic and even ethical judgments by evolutionary stories and I think the same applies to reasoning and even to decisions. One can argue, for example, that because cyclical preferences were dysfunctional, humans evolved to dislike them, or to feel uneasy about them.

How then are psychology and decision sciences exactly related?

As most social sciences, decision sciences have a descriptive and a normative side: they are about what reality is, but also about how we

can change it. Psychology feeds decision sciences with facts about how people actually behave, which is the reality that has to be taken into account. At times one has the chance to try to change this reality (say, if someone asks you for advice). But then again you need to know something about reality—that is, what can possibly be done, what we can expect humans to do.

And what are the implications of such a view for the potentials and limitations of rational choice theory?

I indeed do not think it is a theory. In most real life situations, it does not commit to any specific prediction. Rather it is a way of thinking that may, at least post hoc, explain a remarkable array of observations and phenomena. When viewing rational choice as a paradigm rather than a theory, it offers ways of thinking about decision problems, but it does not commit us to produce a single, well-defined answer in all cases. For descriptive and normative purposes alike, a paradigm may offer more than one prediction or recommendation, and one may need to use common sense or ad hoc considerations to choose among them. In short, while the rational choice approach is indeed limited as a theory, I think it is quite successful as a paradigm, as a way of organizing our thoughts, and as a way of testing and critiquing reasoning.

This relates to another prominent debate in philosophy of economics about the empirical limitations of economics as a discipline. There was a time in which philosophers like Alex Rosenberg did even call into question its status as a science (e.g., Rosenberg 1992). Critics of economics often referred to the axiomatic theory of rational choice as the main weakness of economic theory and believed that behaviourally or psychologically more accurate theories of individual agents would rescue economics from all its troubles. Is this still a legitimate criticism of economics and rational choice theory specifically?

Economics certainly has limitations as a science. However, we should not take this criticism too seriously for two reasons. First, one should not expect to be able to predict the behaviour of large, complex, and causally interrelated systems such as economies, polities, or societies. Even in the case of weather prediction, where the basic physical laws are well understood, prediction for more than ten days ahead is rather limited. In the social sciences we have two additional problems:

a) we do not have the basic laws, the counterpart of the flow equations; b) we are dealing with systems that respond to the predictions made about them. Thus, there are some fundamental limitations to the possibility of predicting the behaviour of economies and we should have realistic expectations about these limitations. Indeed, when we are dealing with smaller, isolated systems, that are causally independent of each other and can be experimented with, prediction is much easier.

Second, the failures of basic choice theory in psychological experiments are often exaggerated. Surely every axiom and every principle has counter-examples. The question is not whether a theory is perfectly correct, but whether it is incorrect in an important way for economic applications. Psychological experiments are selected by their ability to shed new light on the working of the human mind. They need not be a representative sample of economic decisions. I do not think we should be entrenched in defending our classical theories (as economists were some 20 years ago), but we should not get carried away to the other side, decide that the theory is completely wrong, and make predictions only on the basis of vague similarities between experimental situations and real-life economic decisions.

Furthermore, it remains unclear whether the empirical shortcomings of economic models are always to be sought in the rational choice foundations. Even if we are unhappy with a model's predictions, I would argue that the problem rests only sometimes in the foundations of rational choice theory and very seldom in the rational choice paradigm. Let us start with the rational choice paradigm: behavioural economics, by and large, retains the paradigm. In fact, it has been criticized precisely on these grounds, namely, that it does not do much beyond incorporating one more variable in the utility function. As for rational choice theory, there are many problems in the very assumption that we observe equilibria, that all agents share the same beliefs, or that beliefs can be represented by probabilities. All these assumptions are highly questionable and have little or nothing to do with rationality, as I understand it. But even if you think that the agents should be rational à la Savage, and care only about monetary payoffs, the assumptions that they all have the same prior probabilities, or that they play an equilibrium of the game cannot easily be derived from each agent's rationality, even when the terms are very broadly understood.

So, do economic models based upon rational choice foundations have epistemic value?

Yes! Let me stress two more points. First, the rational choice paradigm can also be used for making predictions by drawing analogies to models, and not only by applying general rules. This is a different view of scientific reasoning than the classical, Popperian one, and it is a way in which economic models can be useful without providing general rules that are empirically validated.

Second, an important role of economics is to criticize reasoning. Just as logic is a basic tool for such criticism, so is equilibrium analysis. According to this view, economics is not about making predictions, but only about finding flaws in reasoning by others (say, politicians). I think that there is little doubt that the rational choice paradigm has been very useful as an aid to such criticism.

But would economics not lose its empirical value if we took its role to be criticizing reasoning?

It is not the only way to understand or do economics. But suppose we do follow this line—economics can be very useful without being an empirical science. History is an example of a discipline that is broadly considered to be very useful. Yet, very few historians would venture to make empirical predictions as if they were scientifically based. Similarly, the standard view of philosophy is that it is very far from being an empirical science, yet that it is a good idea to study philosophy, and that, in some ways, the world will become a better place with philosophers. I believe that economists could justify the existence of their discipline in a similar way: focusing on criticism and helping society avoid major mistakes would be enough to justify the field and its costs to society.

Taking this issue one step further, the status of economics as a science has frequently been addressed in discussions about using rational choice theory in economic models, asking the question whether a model based on descriptively unrealistic assumptions can have any epistemic value and, if so, what kind of knowledge it generates. In your book review (Gilboa, forthcoming) of Mary Morgan's* The world in the model *(Morgan 2013) you highlight that one frequent defence of abstract economic models is that what matters for them to have epistemic value is not the realisticness of

assumptions, but the consistency of assumptions with reality. How much 'inaccuracy' can economists accept without jeopardizing the little if any empirical value that is still granted to economics, especially after the recent economic crisis?

First, I think that the economic crisis of 2008 is not a good example. As mentioned earlier, economists are no better equipped to predict financial crises than are physicists to predict tsunamis. This goes back to the issues of complexity of the system, the inability to test the system in isolation, and so on. There are problems that the last crisis highlighted—whether it is a matter of incentives or the belief in free markets (which might involve a major component of betting)—but it should be born in mind that unpredicted crises do not cast doubt on economics as such any more than unpredicted tsunamis cast doubt on physics.

Second, the easiest way to defend the position that economics has some value is to emphasize models as tools for criticism: models, even if they make assumptions that are generally implausible, can be very useful in testing the logic of claims being made in the public domain. And such criticism can be very useful and save us a lot of unnecessary suffering. Relatedly, assumptions that are implausible as general rules may still be very useful in constructing models that may be, to some degree, similar to reality.

However, I do believe that the lack of realisticness should be kept in mind. And when we see economists who truly believe the predictions of their models, we should be wary. It is wonderful to have models, as long as we acknowledge their limitations. Here starts one important task of the philosopher. Philosophers should not just endorse the use of unrealistic assumptions. They should ask: 'When and how do and should scientists use such assumptions despite their unrealism?', 'Why do scientists find unrealistic assumptions still useful?', 'When should we, philosophers, warn them that they have been going too far with the implications of these assumptions?'

So, when should philosophers warn economists?

This question has a theoretical and an empirical side. On the theoretical side, I could say that the answer depends on the model of philosophy of science that you apply to economics: do you think of it as a Popperian science, as a practice of reasoning by analogies, as a field of criticism, and the like. On the empirical side, I fear that I do not have a good

answer. My approach to philosophy of science as a social science requires that I restrict myself to its theories and keep silent on empirical questions. Just as I would not make empirical comments in, say, labour economics, I should not make them in philosophy of science. Hopefully there are empirical researchers who can give much better founded answers to these questions than I possibly could.

Philosophers and psychologists are often ignored by economists and decision theorists. Although behavioural economics has gained prominence in economics, psychologically informed decision theories, such as the research program defended by Gerd Gigerenzer (Gigerenzer, et al. 2011) have not had a considerable impact on economics. Why?

There may be several, perhaps related reasons for that. First, economists and decision theorists tend to be people who like beauty and elegance. Given a wonderful construction such as the Bayesian paradigm, it is just not fun to use other methods, which are less elegant, and whose inclusion would make the entire theory even messier. That is, one has to have a meta-theory, describing when one should use a Bayesian approach and when one should use other approaches. The whole thing may look rather ad hoc. Second, for many questions that are interesting to economists, the origin of beliefs as requested by the Bayesian paradigm may not matter that much. Thus, some economists ask: 'Why should I care? If I can capture the relevant aspects of behaviour by a model using probabilities, why should I bother to specify the process of generation of such probabilities? If I need to know the probabilities for empirical work, I will anyway have to measure them directly'. This line of reasoning also conforms to the 'black box' interpretation of choice theory—the revealed preference paradigm, on which most economists have been educated. I should mention that, while I personally take issue with this line of reasoning, it is not easy to make the case for the importance of the process, and it is particularly difficult to make an argument to convince economists that these foundational choices might lead to different predictions; partly because we are comparing paradigms, or languages, rather than specific theories.

Could you nevertheless try to sketch such an argument?

With pleasure! Suppose that we wish to predict economic behaviour after a financial crisis such as that of 2007-2008. Past examples are very

few, and it is hard to argue that we have observed behaviour under many similar circumstances. Worse still, these are causally intertwined: just because governments did little in 1929, they were prodded to do more in 2008. That is, we cannot use available data to make predictions as in standard statistics; the very fact that certain things happened changes the likelihood that they will happen again. So we cannot rely on the behaviour of the black box in the past, and pretend that we know enough about behaviour so as not to worry about cognition. And we have to ask: How do people think? Will they make predictions here by analogies or by rules? Will they end up using a probability measure, and if so, how will they find one? And, if not, what will they use instead? In short, when we have sufficient data on past cases that are similar and causally independent, we can say that how people think is a problem for psychologists, and we only care about their (economic) behaviour. But when we do not have enough such data, we have to roll up our sleeves and delve deeper into the decision making process.

A similar observation of ignoring new approaches can be made in the research on decision-making under uncertainty. There are several kinds of axiom systems and more generally rational choice theories that attempted to capture the idea of uncertainty. There is the Bayesian approach, J. M. Keynes's approach (Keynes 1921), and Isaac Levi's work. Together with David Schmeidler (Gilboa and Schmeidler 1989), you suggested an axiomatic foundation of the maxmin expected utility decision rule to address the problem of a non-unique prior for example, and thereby made an important contribution to decision making under uncertainty that takes the Knightian concept of uncertainty seriously (and also opposed many accounts that reduce uncertainty to risk for operational purposes). While those approaches have profoundly influenced each other, why do you think some of them failed to be influential in economics while others, like the Bayesian approach, were widely taken up?

Isaac Levi's work is mostly unknown to economists, but it also does not provide the axiomatization that is needed to convince economists that a particular paradigm is the one to use. So we are mostly left with the Bayesian paradigm and the alternatives proposed by the uncertainty theories, starting with Schmeidler's version of Choquet expected utility (Schmeidler 1989), the maxmin rule, and others.

Independently of whether we think of these new theories as part of a 'protective belt' of the same traditional research program or as a new research program, I think they do get sufficient attention in economics. When economists see phenomena that are difficult to reconcile with the Bayesian approach, at some point they are willing to look a bit beyond—and then it becomes advantageous to have a model that is a slight generalization, as opposed to a whole new approach, using a different language.

Surely, 'paradigms' or 'conceptual frameworks' such as Savage's are adhered to longer than are specific theories, precisely because these are paradigms within which new theories can be developed. But there comes a point where people are willing to look beyond the paradigm as well.

This process differs from Thomas Kuhn's in that no one expects a theory—or a paradigm—to be universal (see Kuhn 1962). So that the fact that we need to go beyond a certain paradigm to explain some phenomena does not mean that the paradigm should be discarded. If you will, you can also argue that this is the case in classical examples such as physics. Just as Newtonian physics is still the basic working tool for engineering, the Bayesian approach may well remain the basic workhorse of economic theory.

What are the most important unresolved questions in decision sciences today?

Maybe we should start with resolved ones. I fear I do not know of any. We still do not know how people make, and should make decisions, under risk as well as under uncertainty, in lab situations and in real life, in economic set-ups, or in others. We have some wonderful ideas constituting a fantastic paradigm, but we have very few concrete answers.

Yet, I think we gained a much better understanding of the questions. We learned to distinguish between, say, risk and uncertainty, groups and individuals, and so forth. But, as mentioned, I think we also need to distinguish between types of applications—say, a theoretical application where an economist plugs a representation into a formal model, or a practical one, where a patient decides whether to undergo an operation. Also, it is not clear that the same model would apply to people's decisions when they trade stocks as when they get married, purchase products or wage wars, when they consciously make decisions, or find out that a certain decision has simply occurred.

In a sense, I think that we have not quite resolved the question of what decision sciences are about, or what their questions are precisely. In my view, we need to realize that there are many different questions that need not share an answer. Once we realize that, we can start asking which of these questions have been resolved.

Do you expect decision sciences to progress?

It is possible that we suddenly see less axiomatic models, just like we saw tons of refinements of the Nash equilibrium in the 1980's and then, at some point, people lost interest in them. The research moved forward, or backward, or sideward. It is hard to tell whether it moved forward in a progressive way. There are always fads in the different disciplines. And although we now see a lot of general decision-theoretic models, it is possible that after a while people would still keep asking: 'What has decision theory done for us lately?' And if the answer is negative, then we might be seeing less of these models.

Are you after truth?

Not in a metaphysical sense of truth that exists outside—I do not understand what it means. So, I am willing to do only psychological metaphysics, which is along the lines of 'let us take a metaphysical question and consider its psychological manifestations'. Let me then reread 'truth', or translate the term to mean, a warm feeling of understanding, or the warm feeling that comes from understanding, coupled with the belief that I am not going to change my mind so soon. If that is truth, then yes, I am after that. I like to understand things, and mathematics allows me to do that because, once you check the proof, you rarely change your mind about it. You might change your mind more about things that cannot be mathematically proven or have not been proven yet. So in short, if the meaning of truth is psychological subjective truth, then yes, I am after it.

REFERENCES

Ellsberg, Daniel. 1961. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 75 (4): 643-669.

Gigerenzer, Gerd, Ralph Hertwig, and Thorsten Pachur (eds.). 2011. *Heuristics: the foundations of adaptive behavior*. Oxford: Oxford University Press.

Gilboa, Itzhak. 2009. *Theory of decision under uncertainty*. Cambridge: Cambridge University Press.

Gilboa, Itzhak. 2010a. *Rational choice*. Cambridge: MIT Press.

Gilboa, Itzhak. 2010b. *Making better decisions.* Chichester (UK): Wiley-Blackwell.
Gilboa, Itzhak. 2010c. Questions in decision theory. *Annual Reviews in Economics*, 2: 1-19.
Gilboa, Itzhak. *Forthcoming.* Book review: Mary S. Morgan's 'The world in the model: how economists work and think'. *Journal of Economic Methodology*.
Gilboa, Itzhak, Andrew Postlewaite, and David Schmeidler. 2009. Is it always rational to satisfy Savage's axioms? *Economics and Philosophy*, 25 (3): 285-296.
Gilboa, Itzhak, Andrew Postlewaite, and David Schmeidler. 2012. Rationality of belief or: why Savage's axioms are neither necessary nor sufficient for rationality. *Synthese*, 187 (1): 11-31.
Gilboa, Itzhak, and David Schmeidler. 1989. Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics*, 18 (2): 141-153.
Gilboa, Itzhak, and David Schmeidler. 2001. *A theory of case-based decisions*. Cambridge: Cambridge University Press.
Gilboa, Itzhak, and David Schmeidler. 2012. *Case-based predictions: an axiomatic approach to prediction, classification and statistical learning*. Singapore: World Scientific Publishing.
Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory. *Econometrica*, 47 (2): 263-292.
Keynes, John Maynard. 1921. *A treatise on probability*. London: Macmillan & Co.
Kuhn, Thomas S. 1962. *The structure of scientific revolutions*. Chicago: University of Chicago Press.
Morgan, Mary S. 2012. *The world in the model: how economists work and think*. Cambridge: Cambridge University Press.
Rosenberg, Alexander. 1992. If economics isn't a science, what is it? In *Readings in the philosophy of social science*, eds. Michael Martin, and Lee C. McIntyre. Cambridge: MIT Press, 661-674.
Savage, Leonard. 1972 [1954]. *The foundations of statistics.* New York: John Wiley Publications.
Schmeidler, David. 1989. Subjective probability and expected utility without additivity. *Econometrica*, 57 (3): 517-587.
von Neumann, John, and Oskar Morgenstern. 1947. *Theory of games and economic behavior* [2nd edition]. Princeton (NJ): Princeton University Press.

Itzhak Gilboa's Webpage: http://itzhakgilboa.weebly.com/