

Rationality and the Bayesian Paradigm: An Integrative Note*

Itzhak Gilboa[†]

January 2014

Abstract

Keywords: rationality, probability, reasoning

It is argued that, contrary to a rather prevalent view within economic theory, rationality does not imply Bayesianism. The note begins by defining these terms and justifying the choice of these definitions, proceeds to survey the main justification for this prevalent view, and concludes by highlighting its weaknesses.

1 Introduction

The Bayesian approach holds that uncertainty should be quantified by probabilities. In decision and game theory, this approach has been coupled with expected utility maximization, thereby endowing probabilities with concrete behavioral meaning. The resulting model, suggesting that decision makers

*This note presents and integrates ideas that have, for the most part, appeared in other publications. Most of these publications are co-authored. Moreover, many of the ideas presented have been discussed with my teacher and close colleague David Schmeidler over the past 30 years. There is no claim to originality of any material included herein, and should some of it be new, it is also not clear whom it should be attributed to. However, the co-authors of the other publications and, in particular, David Schmeidler should not be held responsible for all of the views, or for the particular formulation of any position presented here. ISF Grant 396/10, ERC Grant 269754, and financial support from the Foerder Institute for Research in Economics are gratefully acknowledged.

[†]Tel-Aviv University, and HEC, Paris. tzachigilboa@gmail.com

maximize expected utility relative to a potentially-subjective probability, has been supported by axiomatic treatment, most notably Savage (1954), and it has been embraced by economics as the standard of rationality.

The subjective expected utility (SEU) model is also the most widely used for describing the behavior of economic agents facing uncertainty. As a descriptive theory, it has come under attack with Ellsberg's (1961) famous mind-experiments and the experimental and empirical literature that followed.¹ Over the years, and in particular with the rise of behavioral economics, many economists have come to concede that the SEU model may not provide a perfect description of the way decision makers behave, and that, for more accurate descriptive theories, one may expand one's horizon and seek alternative, typically more general models. Yet, this concession on the descriptive front is seldom accompanied by any flexibility on normative issues. It appears that most economic theorists who are willing to consider alternative descriptive models still hold that from a normative point of view such models are unsatisfactory. They hold that people may indeed be irrational in a variety of ways, including violating Savage's axioms, but that *rational* people should not behave so wildly, and that rationality implies Bayesianism.

The present note challenges this view. To this end, we begin by defining the terms "rationality" and "Bayesianism". The definition of rationality is unorthodox, and calls for justification, whereas the definition of Bayesianism is rather standard. After clarifying the way these two terms are used, we will proceed to examine the view that the former necessitates the latter.

¹Allais (1953) as well as Kahneman and Tversky (1979) also attacked the descriptive validity of expected utility maximization, but their focus has been the linearity of the evaluation functional with respect to given probabilities, rather than the existence of subjective probabilities.

2 Rationality

2.1 A Basic Definition²

Philosophers of the 18th and 19th century did not shy away from making statements about the substantive meaning of rationality. They expressed views about what “Rational Man” should think and do, on issues that are often a matter of value judgment. The rise of neoclassical mathematical economics in the early 20th century, influenced by logical positivism, could be viewed as taking a step back, and reducing the concept of rationality to consistency. Rationality started to be defined as behaving in a way that is sufficiently coherence to allow certain formal representation, such as utility maximization, expected utility maximization, and the like. No longer did rationality say what the utility function should be; rather, rationality was taken to be tantamount to having such a utility function, with, at most, minor restrictions such as monotonicity or concavity. In a sense, rationality ceased to be a matter of content, and became a matter of form.

Towards the end of the 20th century, partly due to attacks based on psychological findings, economics started questioning the axioms that defined rationality. While the vast majority of theoretical, experimental, and empirical studies were concerned with the descriptive validity of the classical model, normative issues inevitably surfaced. Orthodox economists often refused to accept violations of the standard model, arguing that it is unreasonable to assume that economic agents are irrational. Sooner or later, the essence of rationality was questioned. Thus, whereas in the beginning of the 20th century it was accepted that rational people may seek different goals, as long as they are all consistent, by the end of the century it was no longer obvious which notion of consistency should be used to define rationality.

In light of such debates, one may take an additional step back from the ideal notion of objective rationality, and admit that not only goals may be

²Based on Gilboa (1994) and Gilboa and Schmeidler (2001, Chapter 2.5, pp. 17-19).

subjective, but also the notion of consistency that is expected of the pursuit of these goals. Just as people may vary in valuing wealth or human life, they may vary in the way they cherish transitivity or some other decision-theoretic axiom. According to this view, a mode of behavior is *irrational* for a decision maker, if, when the latter is exposed to the analysis of her choices, she feels uneasy or embarrassed by them.³ The analysis used for the test of rationality should not include new factual information. Clearly, one may regret a decision post-hoc, given new facts, though the decision might have been rather reasonable given the information available at the time it was taken. In fact, one shouldn't be embarrassed by analysis that relies upon new information: one would be expected to be embarrassed only if one feels that one *could have known better*. By a similar token, one would typically not be embarrassed by analysis that one could not have carried out due to its computational complexity, the degree of imagination it requires, and so forth. Finally, *rationality* is defined by the negation of irrationality. Thus, a decision is rational for a decision maker if analysis of the decision, which could have been carried out by the decision maker at the time of decision, does not make one regret it.

This definition has several features that would strike many theorists as severe weaknesses. First, as opposed to behavioral axioms such as Savage's, this definition makes use of non-behavioral data. It does not suffice to know how a person behaves in order to determine whether they are rational. Rather, we need to find out whether they are embarrassed by their behavior. It is not clear how one can measure this embarrassment or unease, whether the expression of such emotions can be manipulated, and so forth.

Second, according to this definition the choice of axioms that define rationality cannot be made by decision theorists proving theorems on a whiteboard. Rather, the selection of axioms that constitute rationality becomes a subjective and empirical question: some people may be embarrassed by

³The term "embarrassed" in this context was suggested by Amos Tversky.

violating an axiom such as transitivity, while others may not. Moreover, the degree to which people are embarrassed by violation of an axiom may differ across domains, depend on the stakes involved, and so forth. Empirical research should determine what rationality consists of, and we should expect the answer to depend on people's education, culture, and so forth.

Third, this definition also makes rationality non-monotonic in intelligence: suppose that two people make identical decisions, and that they violate a certain axiom. They are then exposed to the analysis of their decisions. One is bright enough to understand the logic of the axiom while the other isn't. The bright one, by feeling embarrassed, would admit that she had been irrational. By contrast, if the less intelligent person fails to see the logic of the axiom, he won't be embarrassed by violating it, and will be considered rational.

In view of these weaknesses, why should we adopt this definition? Its main advantage is that it is potentially useful for the discourse of decision sciences. The current state of the art is marked by a conflict between the remarkable intellectual edifice of the rational choice paradigm and the vast body of experimental findings about violations of choice theoretic principles. On the one hand the field can boast decision and game theory, social choice, and economic theory at large, together constituting an awe-inspiring method of thinking, a general, coherent, and elegant way of conceptualizing choices and social interactions. On the other hand, decision sciences also offer convincing proofs that the elegant may not be realistic, and that people often fail to behave in accordance with the tenets of rational choice. What should we do about this conflict? How does one bridge the gap between the two bodies of literature?

One possibility is to bring theory closer to reality, that is, to modify rational choice theories by incorporating experimental findings from psychology and related fields, in the hope of making the elegant theories descriptively more accurate. This is the direction taken by behavioral economics. Another

possibility is to bring reality closer theory: if people behave “irrationally”, rather than simply documenting it, and perhaps allowing others to take advantage of these “irrational” modes of behavior, theorists can try to change the world they live in by preaching rational choice theories as normative standards.

Which possibility should one choose? It is claimed that the definition of rationality suggested here offers the appropriate test for guiding us in this choice. If it is the case that most people who violate the theory are embarrassed by realizing how they behave, that is, if it is irrational for them to violate the theory, then it makes sense to teach the theory to them, and to hope that they will make better decisions in the future, according to their own judgment. If, by contrast, most people seem to be unperturbed by their violation of the theory, that is, it is rational for them to violate it, then there’s little hope for the theory to be successful as a normative one, and we should accept people’s behavior as a fact that’s here to stay, and that should be incorporated into our descriptive theories to improve their accuracy.

Clearly, the distinction between “rational” and “irrational” behavior is bound to be fuzzy, subjective, context- and culture-dependent. Yet, it may still be useful and insightful. Consider, for example, framing effects (Tversky and Kahneman, 1974, 1981). Intuition, as well as casual observations in classrooms suggest that framing effects are highly irrational phenomena. I have taught many classes in which students exhibited such effects, and none of them tried to defend their choices. Thus, as an empirical hypothesis, I would venture the conjecture that framing effects would be irrational for most people, and that, correspondingly, people may become less prone to such effects as a result of education. Moreover, documenting such effects and discussing them in popular publications may well result in a reduction in their frequency, as people who are exposed to the analysis of these effects will be more immune to them.

By contrast, consider the fact that people do not play Chess optimally.

Suppose that Chess is played by rules that make it finite.⁴ In this case it is known since Zermelo (1913, see Schwalbea and Walkerb, 2001) that one of the following has to hold: (i) White has a strategy that wins against any strategy of Black; (ii) Black has a strategy that wins against any strategy of White; or (iii) each of White and Black has a strategy that guarantees it at least a draw against any strategy of the opponent. Pure logic suffices to determine which of the three possibilities is the case, and, consequently, also how to play Chess optimally (where at least in cases (i) and (ii) the meaning of “optimal play” is unambiguous). Yet, there is no way to figure out which of the three possibilities obtains, as there is no known algorithm that solves the problem and that is also practicable. Thus, people who do not find the optimal strategies in Chess violate a basic assumption of economic theory, namely, that agents know all tautologies. But most of these people would not be embarrassed by this failure. Indeed, had people known all tautologies mathematics departments would have been redundant. When people are shown difficult mathematical proofs they need not feel bad about failing to see the proofs a priori. Similarly, most people would not be troubled by the fact that they couldn’t find the optimal way to play Chess, where no one other human, nor any computer managed to perform this task.

It follows that, according to this definition of rationality, it is probably irrational to fall prey to framing effects but it is not irrational to fail to play Chess optimally. Relatedly, when we confront people with their behavior, exhibiting framing effects, they may well learn to behave differently in future decisions. By contrast, a person who is shown Zermelo’s theorem will still not be able to play Chess optimally in his future games. We may dub such a person “irrational”, but the term will then become hardly useful: all people will be irrational in this sense, and none would be able to become more rational as a result of our preaching. By contrast, the proposed definition of rationality is supposedly more useful: it will make some non-trivial distinc-

⁴Say, that a third repetition of a position is automatically declared a draw.

tions between rational and irrational modes of behavior, and it will use the term “rational” for those modes of behavior that are likely to be observed even after our theories become known to the agents they discuss.

2.2 Objectivity and Subjectivity⁵

The definition of rationality suggested above was refined to distinguish between two notions, called objective and subjective rationality. To define these, it might be useful to clearly define the terms “objective” and “subjective”, which are also laden with different meanings and interpretations.

There is a notion of objectivity that reduces to agreement among subjective concepts. According to this view, in the absence of direct access to anything that is independent of the knower, there could be no other meaning to the term “objective” apart from the equality of the corresponding “subjective” notions of different knowers. Thus, “objectivity” is no more than a nickname for “converging subjective”, not much more than intersubjectivity.

A personal note: my tendencies from early age have been subjectivist and even solipsistic. I never understood metaphysics, and had serious troubles understanding what is meant by objectivity. I have always felt more comfortable with views of the world that allowed any concept to be reduced to theories that attempt to describe, explain, and organize sense data, intuitions, feelings, and other pieces of observations that the mind has to deal with. Over the years philosophers often told me that I confound ontology with epistemology, and I came to realize that I am incapable of understanding metaphysics.⁶ Given this background, I found it very appealing that objectivity should not be anything beyond the agreement of the subjective.

⁵Based on Gilboa and Schmeidler (2001, Chapter 2.7, pp. 20-22) and Gilboa (2009, 13.1 pp. 138-9).

⁶That I claim to be metaphysically challenged should not be taken to suggest that all I do not understand in “metaphysical nonsense”. Just as I fear I will never have a perfect pitch, nor be able to analyze wine intelligently, I may recognize the fact that others can have meaningful metaphysical discussions that I am inherently incapable of following.

However, when David Schmeidler⁷ and I were working on our first book, he convinced me that this definition is lacking. His example was along the following lines: suppose that two people, A and B, are standing in a room and trying to estimate its width. A’s subjective assessment is 10 meters and so is B’s, so that the objective width, according to this definition, is 10 meters. Next compare this scenario to an identical one, apart from the fact that there is a meter stretched out, which shows that the width is indeed 10 meters. Will we not feel that the two scenarios differ?

Assuming that there is a difference between the two, how can it be reconciled with a subjectivist, or solipsistic viewpoint? Can it be captured without reference to an external truth, or to any measurement that is independent of the observer? I maintain that it can. Let us take, for the sake of the argument, a die-hard solipsistic view. I am the conscious speaker, trapped inside my own mind, and I’m trying to explain data to myself, consisting of my recollection of sense data, mental events, and so on. Such a speaking self would start theorizing, thereby postulating the existence of “objects” that are no more than theoretical terms that help her explain her sense data succinctly. As any scientist, a better explanation at a lower complexity cost is to be preferred.⁸ Thus one is led to believe that the sun “exists” where this existence, a la Quine (1948), only means “I, the reasoner, chose to work in a formal model where there exists an object called ‘sun’, which emanates light every day”. For such a reasoner any other person in their society, as well as their own past and future selves, are but figments of their imagination – theoretical concepts that were postulated in order to better organize observations. No member of their society should take offence at having their existence questioned. There’s nothing personal about that. This notion of

⁷David Schmeidler was my master’s and PhD adviser in the ’80s. Since I graduated we remained close colleagues and friends.

⁸I tend to believe that the need to explain data, as well as our preference for simplicity, are evolutionarily selected, and, as far as our discussion is concerned, should be taken as given. On the evolutionary advantages of the preference for simplicity, see Gilboa and Samuelson (2012).

existence is more or less the only one that the reasoner can relate to.⁹

This solipsistic view is not crucial to the argument. It is brought here only to emphasize the fact that the following definition of objectivity still makes no reference to anything that exists beyond a single person's mind. The point is that, back in the room width example, when a meter is present, both people in the room feel differently about their concurrence of opinions. To make this concrete, each person can ask herself what would happen if yet another person were to walk in and be asked about the width of the room. In the first case, the third person may or may not come up with the same subjective assessment. In the second, she is much more likely to do so. And the same would apply to future selves of the two people involved. Hence, even if the two of them are the only people on earth, one may still tell the two scenarios apart by their beliefs about their future observations.

If we say that in the second scenario the measurement is objective, or, perhaps, more objective than in the first, we offer a definition of objectivity that is second-order subjective: it is still based on a single person's beliefs, but these beliefs are not only about the observations in question, but also about the expressed beliefs of others regarding these observations. Thus, a person may still reside in her own mind, being totally solipsistic. But, in order to explain her sense data, she has come to postulate the existence of other people around her, who behave in a consistent enough manner to be attached identities, personalities, wishes, beliefs, etc. The person then asks herself, how much agreement would there be among the observed manifestations of these other people's assessments?

To make this scenario more concrete, assume that the two people who assess the room's width, A and B, are told that ten other people are going

⁹Does the reasoner assume that she herself exists at the time of thinking? This will depend on the model, of course. I believe that in the most intuitive models this question would have no meaning. However, it is also legitimate to argue that the first-person pronoun exists by virtue of being used, and, indeed, prefixing every statement (along the lines of "I observe..." "I think...") – and in this sense its existence is of a different kind than the existence that she may or may not bestow upon others.

to be randomly sampled from the population they live in. These ten people would be paid to participate in an experiment where they are asked to assess the room's width and their payoff will depend on the accuracy of their assessments. A and B are asked, in turn, to assess the standard deviation of the answers given by the ten randomly selected people. If I were, say, person A, and a meter were stretched along the room, I would be willing to bet on this standard deviation being zero. Of course, I might lose the bet. Some people may not be able to read a meter, or may be confused, annoyed by monetary incentives, insane, or what not. Indeed, if a thousand people were involved, I'd probably bet on some of them uttering a statement other than "10 meters". Yet, with ten people I'd probably feel that it is safer to bet on all of them stating the same assessment of 10 meters than on some of them deviating from it. By contrast, in the absence of a meter (that is, in the first scenario), I'd find it highly unlikely that ten people would be asked and, without knowing each other's assessment, would come up with the same number exactly. In other words, I would bet that a positive standard deviation in the sample is more likely than a zero one.

This example suggests that objectivity can be defined in a way that goes beyond the mere, perhaps coincidental, concurrence of subjective assessments, without reference to any external truth that lies outside the reasoner's mind. It also highlights the fact that this definition of objectivity would be quantitative rather than qualitative, and that it will be culture-dependent. I tend to view these as advantages of the definition of objectivity, mirroring the same features of the definition of rationality discussed above.

These definitions have gradations, and, indeed, so do the usages of the corresponding terms in everyday life. Similarly, the fact that these definitions are context- and culture-dependent can also be viewed as an advantage: it makes them pertinent to social science. Specifically, if one considers theories that deal with potential social change, one should better take into account the society that is the audience to which these theories are addressed.

2.3 Objective and Subjective Rationality¹⁰

We discussed a definition of rationality according to which a decision maker behaves rationally if she is not embarrassed by the analysis of her choices.¹¹ This definition is tightly related to our ability to convince a person that her mode of decision making should be changed, but it does not say what the decision maker would do instead. Hence it is rather weak: it judges whether a decision is rational given that it has been made (or about to be made); but it does not specify which decision would be rational without a pre-existing bias towards one of them.

Thus, one can seek a more demanding notion of rationality, that would not only test the robustness of a given, incumbent decision. Following the same context, viewing rationality as part of a rhetorical game or a hypothetical debate, it is natural to suggest that the stronger notion be that one be able to convince decision makers that the “rational” decision is the correct one, irrespective of their a priori tendencies. We thus refine the notion of rationality as follows: a decision is *subjectively rational* for a decision maker if she cannot be convinced that this decision is wrong; a decision is *objectively rational* for a decision maker if she can be convinced that this decision is right.¹² For a choice to be subjectively rational, it should be defensible once made; to be objectively rational, it needs to be able to beat other possible choices.

Clearly, these two definitions are of the same nature and suffer from the same drawbacks and advantages as the previous one: they rely on debates, so that they are not purely behavioral; they are distinctions of degree, not of kind; they would depend on the society and the context to which they

¹⁰Based on Gilboa (2009, 13.2, pp. 139-141, and 17.7, pp. 166-8).

¹¹The decision maker in question need not be a person – it can be an organization or a computer program. All that’s required for the definition to make sense is that we be able to converse with the decision maker, and that it be able to express embarrassment.

¹²These definitions appeared in Gilboa (2009) and used in the context of decision under uncertainty in Gilboa, Maccheroni, Marinacci, and Schmeidler (2010).

apply; and they allow less intelligent people to be more rational than more intelligent ones.

Objective rationality is reminiscent of Habermas's (1981) notion of rationality. Both highlight the need to convince others, and thereby make rationality dependent on culture and context. However, several differences exist. Objective rationality as defined here does not necessitate reasons, and can thereby apply to axiomatic principles. For example, transitivity of preferences may be compelling as an argument without supporting it by presumably more basic reasons (which then need to be assumed as primitive or axiomatic). Also, objective rationality does not presuppose that the society it applies to is democratic, allows free speech, and so forth. Thus, objective rationality is a significantly weaker concept than communicative rationality. It is, however, more pragmatic: it is a litmus test of what a society might accept which can be applied to non-democratic societies, with or without extensive reasoning.

Consider the health hazards of smoking and of mobile phones. At the beginning of the second decade of the 21st century, it is probably safe to say that smoking has been proven to be dangerous to one's health. Assuming that an individual wishes to maximize her life expectancy, it is objectively rational for her not to smoke. The meaning of this statement is that, in our society, a "reasonable person" could be convinced that smoking is a bad idea if one's only goal is maximizing the duration of one's life. Or, relatedly, one can take the statement to mean that, should we draw a person at random from this society, and expose her to all the evidence, research, and reasoning available, she will most likely be convinced not to smoke, given the goal above. This, in particular, should entail (for the reasonable person) that, if she did decide to smoke, she would feel awkward about this choice and may wish to change it. In other words, objective rationality should imply subjective rationality: if the decision maker finds that it is right to make a certain decision, she shouldn't also find it wrong to make that decision at the same

time. By contrast, the health effects of mobile phones are still not agreed upon. Therefore, given the same goal of maximizing one’s life expectancy, objectively rationality does not imply that one should avoid mobile phones. Nor does objective rationality imply that one may use them, should one have secondary goals that mobile phones might facilitate. Objective rationality is therefore silent on the issue of mobile phones. As no decision in this problem is objectively rational, decision makers may make either decision and enjoy subjective rationality. They cannot prove to others that their decisions were right, but others can’t prove to them that these decisions were wrong.

The above suggests modeling decision making by two relations, say $(\succ^*, \hat{\succ})$, where the first denotes objectively rational decisions, and the second – subjectively rational ones.¹³ The relation \succ^* is supposed to capture all preference instances that a reasonable person would be convinced of (adopting the decision maker’s goals). As such, we should expect it to be incomplete, that is, to leave many pairs of alternative decisions unranked.¹⁴ On the other hand, the relation $\hat{\succ}$ describes the decision maker’s actual behavior, and, as such, it would typically be complete: eventually, a decision is always made, and this decision should be brought forth and described in the model.¹⁵

When we consider decision theoretic axioms, we would typically interpret them differently when applied to objective and to subjective rationality. Consider transitivity for example. To test for subjective rationality, we should ask the decision maker, “Won’t you feel awkward about choosing f over g , g over h , and then h over f ? Or, if someone were to observe you making these choices, declaring strict preferences along the way, will you not be embar-

¹³See Gilboa, Maccheroni, Marinacci, and Schmeidler (2010), where this approach has been applied to decision under uncertainty.

¹⁴For models of incomplete preferences under risk and uncertainty, see the pioneering works by Aumann (1962) and Bewley (2002, working paper dating back to 1986) as well as the more recent Dubra, Maccheroni, and Ok (2004), Galaabaatar and Karni (2013), and others.

¹⁵As the standard argument goes, it is impossible not to decide: if such an option exists, it should be spelled out as one of the options in the problem. That is, deciding not to decide, or choose a “wait and see” option should be part of the model.

rassed? Were you a political leader, will you want to see headlines reporting that you chose alternatives cyclically?” By contrast, transitivity of \succsim^* is differently read: here complete preferences are not yet given, and one attempts to construct them. The statement $f \succsim^* g$ means that there is a “proof”, involving data and analysis, statistics and mathematics, that convinces the reasonable decision maker that f is at least as good as g . And so is the case with g and h if $g \succsim^* h$. Transitivity will now imply that there is such a proof that f is at least as good as h . Indeed, transitivity *provides* us with such a proof: we may take the proof that $f \succsim^* g$ and concatenate it with the proof that $g \succsim^* h$ and then, resorting to transitivity as an inference rule of sorts, we present the proof that $f \succsim^* h$.

More generally, the interpretation of axioms for the two relations is rather different: for the subjective rationality relation, $\hat{\succsim}$, we may assume that the relation is complete and given, and consistency axioms (such as transitivity) are used to verify that the preference order, that is, the totality of preference instances, does not look incoherent or ridiculous. For the objective rationality relation, \succsim^* , the consistency axioms are used to construct new instances of preference from known ones. This suggests that, in general, these two relations may satisfy different axioms. Specifically, as objective rationality starts out with preference instances that can be “proved”, it relies on relatively sound foundations, and may be required to satisfy rather demanding consistency axioms. By contrast, subjective rationality is required to be complete, and may include many preference instances that are almost arbitrary. Hence, it may be suggested that only weaker notions of consistency should be applied to this relation.

In Gilboa, Maccheroni, Marinacci, and Schmeidler (2010), this approach has been applied to decision under uncertainty. Basic axioms, such as transitivity, are imposed on both \succsim^* and $\hat{\succsim}$. Beyond these, the former relation is allowed to be incomplete, while required to satisfy the stringest consistency axioms (including the full strength of Anscombe-Aumann’s [1963] Indepen-

dence axiom). By contrast, the latter relation is required to be complete, but to satisfy only weaker consistency axioms (C-Independence and Uncertainty Aversion). That objective rationality should imply subjective rationality is captured by the axioms, $\succsim^* \subset \hat{\succsim}$, and additional axioms may be imposed to relate the two relations.

3 Bayesianism

3.1 The Bayesian Approach

The Bayesian approach holds that uncertainty should be quantified by probabilities. Specifically, it suggests that, in the absence of objective, agreed-upon probabilities, each person formulate her own probabilities, reflecting her subjective beliefs. One way to view the Bayesian approach holds that the laws of probability are tautologies when interpreted as facts about empirical frequencies; these laws can then be used as self-imposed constraints on the way one describes one’s beliefs also in the absence of empirical frequencies.

A fuller account of the Bayesian approach would start with the formulation of a state space, where each state (“of the world” or “of nature”¹⁶) provides a complete description of all that matters to the decision maker, and can be thought of as a truth table specifying the truth value of any proposition of interest. Thus, the world is assumed to be in precisely one of the possible states, and each state describes the entire history and future of the decision problem. On this state space one formulates a “prior”, that is, a probability measure describing one’s subjective beliefs before getting any information. At the arrival of new information, for example, the observation of the realization of some random variables, the prior probability is updated by Bayes’s rule to generate a “posterior”, which may, in turn, be the prior of the next period, to be updated again, and so forth. Within economics, people often read the term “Bayesian” as implying the maximization of expected

¹⁶The distinction between these terms is not crucial for our discussion.

utility relative to one's subjective beliefs. However, Bayesian approaches are used in statistics and in machine learning, where they are not necessarily related to decision making. Hence, we will not include any specific decision rule in the definition of the term.

The formulation of states of the world as mutually exclusive and jointly exhaustive list of the possible scenarios is hardly problematic. Admittedly, the construction of the state space might be a daunting computational task, and at times this complexity may render the approach impracticable. But, blithely ignoring computational difficulties, one can find little that is objectionable in the formulation of the state space. Moreover, there are cases in which this formulation is key to the resolution of troubling puzzles. (See the discussion of Newcombe's Paradox, Hempel's Paradox of confirmation, and Monty Hall's three-door problem in Gilboa, 2009, Chapter 11, pp. 113-122.)

Bayesian updating is generally considered to be the only rational approach to learning once one has a prior. The normative appeal of Bayes's formula has only rarely been challenged, partly because the formula says very little: it only suggests to ignore that which is known not to be the case. Following the empiricist principle of avoiding direct arguments with facts, Bayesian updating only suggests that probabilities be re-normalized to retain the convention that they sum up to unity.

By contrast, the second tenet of Bayesianism, namely that the uncertainty about the state space be represented by a probability measure, has been challenged since the early days of probability theory. (See Shafer, 1986, who finds non-Bayesian reasoning in the works of Jacques Bernoulli, 1713.) Indeed, while the first tenet – namely, the formulation of the state space – is merely a matter of definitions, and the third – Bayesian updating – is quite compelling, it is not at all clear how one should decide which prior probability to impose on the state space, and, consequently, whether it is rational to do so in the first place. The focus of our discussion is, indeed, whether rationality implies that such a specification of a prior probability be

made.

3.2 Bayesian and Classical Statistics¹⁷

There are situations where the prior probability can be determined by common practices, past experience, or mathematical convenience. There are also situations in which this prior probability is rather arbitrary, but as long as it is chosen in a sufficiently open-minded way (“diffused prior” or “uninformative prior”), this choice has little impact on predictions in the long run. Consider, for example, a textbook problem in statistics. A coin is to be flipped consecutively, generating a sequence of iid (independently and identically distributed) random variables. Each is a Bernoulli random variable,

$$X_i = \begin{cases} 0 & 1 - p \\ 1 & p \end{cases}$$

and one attempts to guess what p is based on a sample of X_1, \dots, X_n . The Classical approach to statistical inference holds that given each and every possible value of p , we have a well-defined probability model in which the X_i 's are iid. However, no probability statement can be made about p . The latter is an unknown parameter, not a random variable. For that reason, Classical statistics techniques has various terms that are derived, but differ from probability. For example, a confidence interval for p is an observation of a random variable, whose values are intervals, and that has a certain a priori probability of covering the true p – whatever that p is. After the random variable is observed, we retain the probability number, *that used to be its probability of covering p* , and call it the “confidence level” of the interval. However, this confidence level is not the probability that p is in the interval. This probability isn't defined, neither before nor after the sample is observed. Similarly, the notion of “significance” in hypotheses testing is derived from probabilities, but cannot be viewed as the probability that a

¹⁷Based on Gilboa, 1994, and Gilboa, 2009, Chapter 5.3, pp. 40-48.

hypothesis is true or false. Classical statisticians test hypotheses, but they do not formulate their beliefs about these hypotheses in terms of probabilities, neither before nor after the observations of the sample.

Apart from the theoretical complexity, where probabilities, confidence levels, and significance levels are floating around as separate notions, the Classical approach leads to various paradoxes.¹⁸ The Bayesian approach seems to be conceptually simpler and more coherent, allowing only one notion of quantification of beliefs, and describing all that is known and believed in the same language. In the example above, the Bayesian approach would require the specification of a prior probability over the values of p , treating the unknown parameter as a random variable. More generally, the Bayesian approach distinguishes between what is known and what isn't. Whether the unknowns are fixed parameters that the statistician does not know, or variables that are in some sense "inherently random" has no import: anything unknown is treated probabilistically. Once the unknown parameter p is a random variable with a given distribution (say, uniform, to keep the prior "diffused" and allow for learning of all possible values), the statistician has a joint distribution for the $(n + 1)$ random variables p and X_1, \dots, X_n . Having observed the latter n , she updates her probability distribution over p by Bayes's rule and thus gradually learns the unknown parameter.

Consider also the case of hypotheses testing. Adopting a court-case analogy that used to be popular in my youth, we may think of the hypothesis that is being tested, H_0 , as the default assumption that the defendant is not guilty, while the alternative, H_1 , is that he is. The nature of the exercise is that there is an attempt to "prove" H_1 beyond reasonable doubt. In the absence of such a proof, the hypothesis H_0 is not rejected (some even use the term "accepted"). However, this need not mean that the hypothesis is correct, or even believed to be correct. It is the null hypothesis by default. Failing to reject it does not imply that one believes it to be true.

¹⁸See, for example, De Groot (1975), pp. 400-401.

The Bayesian approach to this problem is, again, much simpler: one formulates a prior probability over the two hypotheses, or, to be precise, over all the unknowns (from which the probability that each hypothesis is true can be derived). One then observes the sample and updates the prior probabilities to posteriors according to Bayes's rule.

This example might also expose the main weakness of the Bayesian approach. Suppose that I am a defendant, and that the jury attempts to be Bayesian and to formulate a prior probability for the hypothesis that I am guilty. Assume that they ask my mother what their prior probability should be. She believes, or has an incentive to pretend to believe, that this probability is zero. Starting off with a zero prior, no amount of information would resurrect it to a positive posterior, and I will be found not guilty. Next assume that the same jury consults with someone else, who happens not to like me too much. This other person starts off with a very high probability for my guilt, and scant evidence suffices for my conviction. Considering these two extremes, I feel uneasy. I would not like my fate to be determined by someone's very subjective hunches. I rather live in a society where one's guilt is determined in a more objective way. Perfect objectivity is, of course, an unattainable ideal, but one may still strive to be more objective by ruling out subjective inputs.

This discussion suggests that Classical and Bayesian statistics need not be construed as two competing paradigms attempting to achieve the same goal. An alternative view holds that the two approaches are designed for different goals: Classical statistics deals with assertions that can be established in a society, with the understanding that people in this society have different beliefs, as well as different goals that may make them express different views (regardless of their true beliefs). In such a set-up, the question that Classical statistics attempts to answer is what can be objectively stated, that is, what beliefs can be attributed to society as a heterogenous whole. Bayesian statistics, on the other hand, deals with the most accurate representation of

a single person's beliefs, and, relatedly, with choosing the best decision for her. Focusing on a single individual, this approach gladly embraces subjective inputs, intuition and hunches. It strives to make these coherent, using the probability model as a way to impose discipline on these intuitions, but it does not rule them out due to their subjectivity.

It is important to emphasize that the goals of the two approaches need not always be attained. Classical statistics cannot be perfectly objective, and it requires some statistical choices that are, in the final analysis, subjective. And the Bayesian approach is not always very successful at capturing intuition. However, the distinction suggested here is based on the *goals* of the two approaches, and these are claimed to be different.

To see the distinction between the two goals more clearly, consider two examples. First, assume that I'm a juror on a murder case. The defendant, Mr. A, is accused of having murdered his girlfriend. There is something very troubling about the behavior of the guy in court. He seems to be sneering and the tone of his voice makes me believe that he's a psychopath. However, considering all evidence, including a psychologist's expert testimony, Mr. A's lawyer makes a very convincing case that guilt has not been established beyond a reasonable doubt. Together with the other jurors, I find myself voting for acquittal. As I come home, I see my 20-year-old girl dressing up. Asking her about her plans, she says she has a date with no other than Mr. A. No way, honey, I reply, nothing of the kind is going to happen.

Assuming that my daughter and I are a single decision maker (and that she heeds my advice), my decision about the date differs from my voting decision in court. In court, I try to be objective. I attempt to put aside all my hunches and the distrust I feel for the guy, and base my vote on what I believe can be said to be objectively established. However, when I make a decision for myself (or my daughter), I do wish to use all sources of information, including my intuition, even if it cannot be proved. I do not owe anyone a detailed report of the reasoning behind my decision, or the

evidence I rely on, and I'm perfectly justified in following hunches.

To consider another example, suppose that I suffer from a disease that does not have any FDA-approved medication yet. I hear that a large and very renowned drug company has developed a medication, which has been submitted to the FDA for approval, and which I may use as part of a clinical trial. My reasoning about this decision is quite sophisticated and it involves many intuitive assessments: I know that the firm has very good reputation; I take into account that, should the FDA fail to approve the drug, this would be a precedence for the firm, and would tarnish its reputation; this, in turn, may result in a decline of its market value. All in all, I think that the medication is probably safe; moreover, my beliefs about the safety of the medication are more or less the same as they would be had the FDA already approved it. Thus, I decide to go ahead and join the clinical trial. Yet, I would not want to see the FDA following the same reasoning and approving the medication without testing it. The FDA is supposed to be society's official seal, and it should be based on as objective data as possible, rather than on reasoning as above. Again, there is no inconsistency between the FDA conducting the presumably-objective test (Classical statistics) and my own decision to use the medication based on subjective assessment (Bayesian statistics).

To conclude, the Bayesian approach is, for better and worse, about subjectivity. A major advantage of this approach is that it allows one to express certain subjective inputs that are shunned by Classical statistics. A corresponding drawback is that in many cases that Bayesian approach cannot do without such subjective inputs. As a result, the Bayesian approach does not seem to be too closely related to objective rationality as defined above. It is the notion of subjective rationality that is likely to be related to Bayesianism. Thus, after having better defined these concepts we can sharpen the question and ask whether subjective rationality necessitates the Bayesian approach.

4 Does Subjective Rationality Imply Bayesianism?

4.1 History and the Axiomatic Approach

Subjective probabilities are as old as probabilities in general. Blaise Pascal, who is credited as one of the forefathers of the notion of probability and of expectation in the context of games of chance, was also the person who suggested the famous wager, according to which one should choose to run a lifestyle that ends up in faith, because faith has a higher expected utility than lack thereof, as long as the probability of God's existence is positive.¹⁹ Clearly, the probability that God exists is a subjective one, and cannot be related to relative frequencies. Thus, Pascal was a pioneer in defining and using objective probabilities, also ushered subjective probabilities: in his wager he suggested using the mathematical machinery of probability theory, developed to deal with objective probabilities, as a way to sort out one's vague intuitions. However, Pascal may not have been a devout Bayesian: his own argument also continues to allow for the possibility that one may not know the probability that God exists, and he continues to refine his argument for someone who cannot quantify this uncertainty probabilistically, that is, for the non-Bayesian.

Bayes (1763) used the notion of a prior to argue that God probably exists, a line of reasoning that is quite similar to the "Intelligent Design" argument: given the complexity of the universe we observe, it seems more likely that it was generated by a God as described in religious scriptures, than by pure chance: should such a God exist, the conditional probability of the emergence of the universe is 1; otherwise, this conditional probability is rather low. In other words, given the universe we observe, the likelihood function for God existing is much higher than for God not existing. However, to complete the argument and proceed from the likelihood function to conditional

¹⁹See Hacking (1975, pp. 63-67) and Connor (2006).

probabilities (of God's existence given the observations) one needs to have a probability over God's existence to begin with (that is, a prior).²⁰ Bayes used the prior beliefs of 50%-50%, a choice that has been criticized in the centuries that followed.²¹

The early 20th century witnessed the revival of the debate regarding the Bayesian approach. Some, like Keynes (1921, 1937) and Knight (1921) held that there are unknown probabilities, that is, uncertainties that cannot be quantified, while others held that all uncertainty should be quantified by probabilities. Among the latter two prominent thinkers, Ramsey (1926) and de Finetti (1931, 1937) suggested to ground subjective probability on one's willingness to bet. They outlined axiomatic approaches, according to which behavior (such as choices of bets) that is coherent must be equivalent to decision making based on a subjective probability.

Their contributions, coupled with von Neumann-Morgenstern's (1944) derivation of expected utility maximization in the presence of known probabilities ("risk") inspired Savage (1954) to offer one of the most remarkable mathematical results in the social sciences (philosophy included). Savage considered a decision maker facing uncertainty, who chooses between acts that are mappings from states of the world to outcomes. He suggested a few very compelling axioms on coherence of behavior, that are equivalent to the existence of (i) a utility function over outcomes; and (ii) a probability measure of the states of the world such that the decision maker's behavior is representable by the maximization of expected utility (in (i) relative to the probability in (ii)).

Savage's theorem is striking partly because, in an attempt to keep the model intuitive, he did not use any sophisticated mathematical structures in his primitives. His axioms do not resort to linear operations, convergence, or measurability. Correspondingly, his proof cannot apply known tools from

²⁰Indeed, a devout atheist who assigns a zero prior probability to God's existence will not be convinced by any amount of evidence that He is.

²¹See McGrayne (2011).

functional analysis, topology, or measure theory. Moreover, if we limit attention to two possible outcomes, Savage’s assumptions are simply that the decision maker can rank events (in terms of her willingness to bet on them) as suggested in de Finetti’s “qualitative probability” and that one more “technical” assumption is satisfied, guaranteeing continuity and archimedeanity of sorts.

Later on, Anscombe and Aumann (1963) suggested an alternative axiomatization of subjective expected utility maximization, which is closer in spirit to von Neumann and Morgenstern’s model, and which uses objective probabilities as part of the model, highlighting the intuition that one might calibrate one’s subjective beliefs by comparing events whose probabilities are not known to events whose probabilities are objectively given.

These axiomatic derivations should be credited with the fact that formal modeling in economics adopted SEU theory as its main tool for dealing with uncertainty. The axiomatic approach convinced economists that people and organizations who may not be able to calculate probabilities of certain events are still likely to behave *as if* they did attach probabilities to such events, as manifested by their decision making. Moreover, when interpreted normatively, the axioms seemed even more compelling: not only is SEU considered to be a good descriptive theory, it is considered to be the epitome, and often the definition of rationality.

4.2 Difficulties²²

4.2.1 What’s in a state?

Savage’s axioms appear to be very compelling, especially when one considers the states of the world to be well-defined, concrete outcomes of an experiment that can be repeated, such as drawing balls from an urn. But they need not be so convincing when the states have to be defined from the problem’s primitives. For example, Gibbard and Harper (1978) point out that if

²²Based on Gilboa, Postlewaite (2012, 2009) and Gilboa (2009).

states are defined as functions from acts to consequences, Newcombe-style problems disappear. (See also Gilboa, 2009, 11.1, pp. 113-116). Indeed, such a definition of a state allows for all conceivable causal relationships and does not limit the decision maker by any a priori assumptions. However, this definition of states of the world renders the set of conceivable acts that are needed for Savage's model to be larger by two orders of magnitude than the set of acts that are actually available. (See Gilboa and Schmeidler, 1995, and Gilboa, 2009, Chapter 11.1.3 pp. 115-116.) This means that the choices that are assumed in Savage's model are hardly observable. Consequently, the normative appeal of the axioms, assuming a complete order over conceivable acts, is also tarnished. In a similar vein, Gilboa, Postlewaite, and Schmeidler (2009) point out that if one attempts to deal with Ellsberg's (1961) two-urn experiment, one needs to construct the state space and then the violation of Savage's axioms occurs over an event that may not be observable.

The notion of a "state of the world" varies across disciplines and applications. In Bayesian statistics, as described above, the state may be the unknown parameter, coupled with the observations of the experiment. Should one ask, where would I get a prior over the state space, the answer might well be, experience. If the entire statistical problem has been encountered in the past, one may have some idea about a reasonable prior belief, or at least a class of distributions that one may select a prior from. However, in economic theory additional information was introduced into the notion of a state. Harsanyi (1967, 1968a,b) made path-breaking contributions to game theory by showing how games of incomplete information may be reduced to standard games: by adding a chance move at the beginning of the game, uncertainty about players' utilities and beliefs, or "types", can be dealt with as any other uncertainty. In this theoretical construction, states of the world are expanded to include the types of the players, which are partly determined before the players are born. Aumann (1976, 1987) further extended the states to describe all that is possibly relevant – the players' knowledge,

their information partitions, etc.

These extensions of the notion of a state of the world are very elegant, and are sometimes necessary to deal with conceptual problems. Moreover, one may always define the state space in a way that each state would provide answers to all relevant questions. But defining a prior over such a state space becomes a very challenging task, especially if one wishes this definition not to be arbitrary. The more informative are the states, the larger is the space, and, at the same time, the less information one has for the formation of a prior. If, for example, one starts with a parameter of a coin, p , one has to form a prior over the interval $[0, 1]$ and one may hope to have observed problems with coins that would provide a hint about the selection of an appropriate prior in the problem at hand. But if these past problems are now part of the description of a state, there are many more states, as each describes an entire sequence of inference problems. The prior should now be defined at the beginning of time, before the first of these problems has been encountered. Worse still, the choice of the prior over this larger space has, by definition, no information to rely on: should any such information exist, it should be incorporated into the model, requiring the “true” prior to be defined on an even larger state space.

4.2.2 Big questions

Gilboa, Postlewaite, and Schmeidler (2012) discuss Savage’s axioms in detail and attempt to show that rationality need not entail these axioms. It is not worth repeating these arguments here. Let us only comment on two examples that highlight the arbitrariness of Bayesian priors.

Suppose that two individuals, A and B, disagree about the probability of an event, and assign to it probabilities 0.6 and 0.4, respectively. Let us now ask individual A, “If you’re so certain that the probability is 0.6, why can’t you convince B of the same estimate? Or, if B holds the estimate 0.4, and you can’t convince her that she’s wrong, why are you so sure of your 0.6?”

That is, we have already agreed that the estimate of 0.6 can't be objectively rational, as B isn't convinced by it. It is still subjectively rational to hold the belief 0.6, as there is no objective proof that it is wrong. However, A might come to ask herself, do I feel comfortable with an estimate that I cannot justify?

Next suppose that you're walking around campus and you see a sign about a talk titled "Arbodytes and Cyclophines". You are asked about the probability that all Arbodytes are Cyclophines. You have never heard these terms before, so you find it hard to judge what the answer is. Using some social intelligence you presume that there is a field of knowledge where these are known terms and that someone on campus might know the answer. Yet, you have no idea whether this someone would be in the department of microbiology, mathematics, or ancient civilization.

If rationality entails Bayesianism, you will not be rational unless you give a number $p \in [0, 1]$ that is the probability that Arbodytes are a subset of Cyclophines. You may just have a gut feeling that $p = 0.54$. Or you may first have probabilities p_i that the talk is given in department i , and then a conditional probability q_i that Arbodytes are Cyclophines given that the relevant department is i . The latter approach has more structure, but this also means more unknown probabilities to estimate. Whichever way you approach the problem, the probabilities you are asked to name are inevitably very arbitrary.

Some people attempt to deal with this problem using Laplace's Principle of Indifference. Knowing nothing, they say, I'd put a probability of 50% of Arbodytes being a subset of Cyclophines. But then, what about Cyclophines being a subset of Arbodytes? Or the two being disjoint? Logically independent? What about meta-Arbodytes being pseudo-Cyclophines?²³

²³For a fuller discussion of the Principle of Indifference, see Gilboa (2009, Chapter 3, pp. 14-19).

4.3 Arbitrariness and Subjective Rationality

It follows that the Bayesian approach is supported by very elegant axiomatic derivations, but that it forces one to make arbitrary choices. Especially when the states of the world are defined, as is often the case in economic theory, as complete description of history, priors have to be chosen without any compelling justification.

In these situations, a particular prior probability cannot be objectively rational. Is it subjectively rational, then? In the absence of a compelling justification for any probability, every choice seems to be subjectively rational. However, when subjective rationality is applied to the Bayesian paradigm itself, the answer becomes less obvious. While the decision maker cannot be convinced that any particular probability is a wrong choice, she may feel awkward about her very choosing a particular number (or measure) without any basis.

The Bayesian approach is quite successful at representing knowledge, but rather poor when it comes to representing ignorance. When I try to say, within the Bayesian language, “I do not know”, the model asks me, “How much do you not know? Do you not know to degree 0.6 or to degree 0.7?” I simply don’t have an utterance that means “I don’t have the foggiest idea.”²⁴

It is therefore not at all obvious that rationality – even subjective rationality – suggests that we select one prior out of all possible ones. Applying the notion of subjective rationality to the choice of a paradigm, we might find that the gap between subjective and objective rationality is not as large as it may first appear. Many would probably agree that they would feel more comfortable with a choice of a paradigm that can represent ignorance as well

²⁴As Keynes (1937) wrote, “By ‘uncertain’ knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty...The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence...About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know.”

as knowledge.

5 Conclusion

There are many questions to which reason, data, and science provide answers. These are the cases where Classical statistics will be able to establish claims as more or less proven; where objective rationality will dictate specific choices and beliefs; where deviating from these beliefs would be embarrassingly irrational. But life also confronts us with many questions for which the answers cannot be easily obtained with any objectivity or scientific confidence. Alas, many of the more important problems are of this nature. We do not know how to objectively assess probabilities of wars and financial crises, and we even have problems with prediction of hurricanes and tsunamis, as well as with the assessment of global warming. At the personal level, individuals can estimate the probability of, say, their car being stolen, but not necessarily of their success in a given career, or in a marriage. It could be nice to remain silent on these issues, entertain no beliefs and make no predictions that are not objectively rational. Unfortunately, often this is not an option. Life imposes decisions on us, and, by action or inaction, we make choices.

What is the rational thing to do in these situations? How should one complete one's preference relation, answer the unanswered questions, go beyond objective rationality and act in an uncertain world? The Bayesian approach offers one possible answer. It suggests that one way to obtain subjective rationality is to select a prior probability, use it for belief formation and decision making, and update it in the face of new information. However, it is but one approach, which may not be the best choice for all people in all problems. The selection of a prior is highly arbitrary. This is evident in the Arbodytes-Cyclophines example above, as well as in any problem that starts at the very beginning of time, before any information has been gathered. Hence, the Bayesian approach may not be the most subjectively rational

one for many people: while decisions will be coherent given the selection of the prior, the arbitrariness of the latter renders the entire decision process exposed to criticism.

Decision theory has made progress in suggesting alternatives to the Bayesian approach, which are also axiomatically-based, general-purpose models of decision making. The pioneering work, Schmeidler (1989), uses the notion of Choquet (1954) integration with respect to probabilities that are not necessarily additive. A related model, Gilboa and Schmeidler (1989), uses maxmin expected utility with respect to a set of probabilities, reminiscent of Classical statistics, where a collection of distributions is considered possible, without a prior probability over them. Klibanoff, Marinacci, and Mukerji (2005) and Nau (2006) offered a model involving second-order probabilities (the “smooth” model), whereas Maccheroni, Marinacci, and Rustichini (2006) suggested “variational prevalences”, and many other models have been and are being developed. While these models are mostly motivated by the need to provide better descriptive theories, many of them can also be interpreted normatively: when objective rationality leaves us with incomplete preferences (as in Bewley, 2002), completing these preferences confronts us with a trade-off between the choice of a paradigm and the choice of a theory within it. The more elegant, Bayesian paradigm requires arbitrary choices. Some decision makers may prefer non-Bayesian paradigms, satisfying only weaker axioms but requiring less arbitrary choices, as their standard for subjectively rational decisions.

6 References

Allais, M. (1953), “Le Comportement de L’Homme Rationnel devant le Risque: critique des Postulats et Axiomes de l’Ecole Americaine”, *Econometrica*, **21**: 503-546.

Aumann, R. J. (1962), “Utility Theory without the Completeness Axiom”, *Econometrica*, **30**: 445-462.

——— (1976), “Agreeing to Disagree”, *The Annals of Statistics*, **4**: 1236-9.

——— (1987), “Correlated Equilibrium as an Expression of Bayesian Rationality”, *Econometrica*, **55**: 1-18.

Bayes, T. (1763), “An Essay towards solving a Problem in the Doctrine of Chances.” Communicated by Mr. Price, *Philosophical Transactions of the Royal Society of London* **53**, 370–418.

Bernoulli, J. (1713), *Ars Conjectandi*. 1713.

Bewley, T. (2002), “Knightian Decision Theory: Part I”, *Decisions in Economics and Finance*, 25: 79-110. (Working paper, 1986).

Choquet, G. (1953-4), “Theory of Capacities,” *Annales de l’Institut Fourier* **5**: (Grenoble), 131-295.

Connor, J. A. (2006), *Pascal’s Wager: The Man Who Played Dice with God*. New York: HarperCollins.

de Finetti, B. (1937), “La Prevision: Ses Lois Logiques, Ses Sources Subjectives”, *Annales de l’Institut Henri Poincare*, **7**, 1-68.

DeGroot, M. H. (1975), *Probability and Statistics*. Reading, MA: Addison-Wesley.

Dubra, J., F. Maccheroni, and E. A. Ok (2004) “Expected Utility Theory without the Completeness Axiom”, *Journal of Economic Theory*, **115**: 118-133.

- Ellsberg, D. (1961), "Risk, Ambiguity and the Savage Axioms", *Quarterly Journal of Economics*, **75**: 643-669.
- Fishburn, P.C. (1970) *Utility Theory for Decision Making*. John Wiley and Sons, 1970.
- Galaabaatar, T. and E. Karni (2013), "Subjective Expected Utility with Incomplete Preferences", *Econometrica*, **81**: 255–284.
- Gibbard, A. and W. L. Harper (1978). "Counterfactuals and Two Kinds of Expected Utility." *Foundations and Applications of Decision Theory 1*: 125-162.
- Gilboa, I., (1991), "Rationality and Ascriptive Science", unpublished manuscript.
- (1994), "Teaching Statistics: A Letter to Colleagues", unpublished manuscript.
- (2009), *Theory of Decision under Uncertainty*. Econometric Society Monograph Series Vol. No. 45. Cambridge: Cambridge University Press.
- Gilboa, I., F. Maccheroni, M. Marinacci, and D. Schmeidler (2010), "Objective and Subjective Rationality in a Multiple Prior Model", *Econometrica*, **78**: 755-770.
- Gilboa, I., A. Postlewaite, and D. Schmeidler (2009), "Is It Always Rational to Satisfy Savage's Axioms?", *Economics and Philosophy*, **25**: 285-296.
- (2012), "Rationality of Belief", *Synthese*, **187**: 11-31.
- Gilboa, I., and L. Samuelson (2012), "Subjectivity in Inductive Inference", *Theoretical Economics*, **7**: 183–215.
- Gilboa, I. and D. Schmeidler (1989), "Maxmin Expected Utility with a Non-Unique Prior", *Journal of Mathematical Economics*, **18**: 141-153.
- (1995) "Case-Based Decision Theory", *The Quarterly Journal of Economics*, **110**: 605-639.

- (2001), *A Theory of Case-Based Decisions*. Cambridge: Cambridge University Press.
- Habermas, J. (1981), *Theory of Communicative Action, Vol I: Reason and the Rationalization of Society*. English edition, 1984, Boston: Beacon.
- Hacking, I. (1975), *The Emergence of Probability*. Cambridge: Cambridge University Press.
- Harsanyi, J. (1967), “Games of Incomplete Information Played by ‘Bayesian’ Players. Part I: The Basic Model”, *Management Science*, **14**: 159-182.
- (1968a), “Games of Incomplete Information Played by ‘Bayesian’ Players. Part II: Bayesian Equilibrium Points”, *Management Science*, **14**: 320–334.
- (1968b), “Games of Incomplete Information Played by ‘Bayesian’ Players. Part III: The Basic Probability Distribution of the Game”, *Management Science*, **14**: 486–502.
- Kahneman, D. and A. Tversky (1979), “Prospect Theory: An Analysis of Decision Under Risk,” *Econometrica*, **47**: 263-291.
- Keynes, J. M. (1921), *A Treatise on Probability*. London: MacMillan and Co.
- (1937), *The Quarterly Journal of Economics*, from *The Collected Writings of John Maynard Keynes*, **14**: 109-123.
- Klibanoff, P., M. Marinacci, and S. Mukerji (2005), “A Smooth Model of Decision Making under Ambiguity,” *Econometrica*, **73**, 1849-1892.
- Knight, F. H. (1921), *Risk, Uncertainty, and Profit*. Boston, New York: Houghton Mifflin.
- Maccheroni, F., M. Marinacci, and A. Rustichini (2006), “Ambiguity Aversion, Robustness, and the Variational Representation of Preferences,” *Econometrica*, **74**: 1447-1498.

- McGrayne, S. B. (2011), *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. New Haven: Yale University Press.
- Quine (1948), "On What There Is", *Review of Metaphysics* (1948). Reprinted in 1953, in *From a Logical Point of View*. Harvard University Press.
- Ramsey, F. P. (1931), "Truth and Probability", in *The Foundation of Mathematics and Other Logical Essays*. New York, Harcourt, Brace and Co.
- Savage, L. J. (1954), *The Foundations of Statistics*. New York: John Wiley and Sons. (Second addition in 1972, Dover)
- Schmeidler, D. (1989), "Subjective Probability and Expected Utility without Additivity", *Econometrica*, **57**: 571-587.
- Schwalbea, U., and P. Walkerb (2001), "Zermelo and the Early History of Game Theory", *Games and Economic Behavior*, **34**: 123–137.
- Shafer, G. (1986), "Savage Revisited", *Statistical Science*, **1**: 463-486.
- Tversky, A., and D. Kahneman (1974), "Judgment under uncertainty: Heuristics and biases", *Science*, **185**: 1124-1131.
- (1981), "The Framing of Decisions and the Psychology of Choice", *Science*, **211**: 453-458.
- von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior*. Princeton, N.J.: Princeton University Press.
- Wittgenstein, L. (1922), *Tractatus Logico Philosophicus*. London, Routledge and Kegan Paul.