Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 04-011

"Rationality of Belief
Or: Why Bayesianism is neither necessary nor sufficient for rationality"

by

Itzhak Gilboa, Andrew Postlewaite, and David Schmeidler

http://ssrn.com/abstract=523502

Rationality of Belief

Or: Why Bayesianism is neither necessary nor sufficient for rationality*

by

Itzhak Gilboa, Andrew Postlewaite, and David Schmeidler

March 2004

Abstract

Economic theory reduces the concept of rationality to internal consistency. The practice of economics, however, distinguishes between rational and irrational beliefs. There is therefore an interest in a theory of rational beliefs, and of the process by which beliefs are generated and justified. We argue that the Bayesian approach is unsatisfactory for this purpose, for several reasons. First, the Bayesian approach begins with a prior, and models only a very limited form of learning, namely, Bayesian updating. Thus, it is inherently incapable of describing the formation of prior beliefs. Second, there are many situations in which there is not sufficient information for an individual to generate a Bayesian prior. Third, this lack of information is even more acute when we address the beliefs that can be attributed to a society. We hold that one needs to explore other approaches to the representation of information and of beliefs, which may be helpful in describing the formation of Bayesian as well as non-Bayesian beliefs.

1. Rationality of Belief and Belief Formation

One of the hallmarks of the modern era is the belief in rationality. Many writers expected rationality to settle questions of faith, advance science, promote humanistic ideas, and bring peace on Earth. In particular, philosophers did not shy away from arguing what is rational and what is not, and to take a stance regarding what Rational Man should do, believe, and aspire to.

By contrast, economic theory in the 20th century took a much more modest, relativist, and even post-modern approach to rationality. No longer was there a pretense to know what Rational Man should think or do. Rather, rationality was reduced to various concepts of internal consistency. For example, rational choice under certainty became synonymous

---

* These notes were written to organize our thinking on the topic. We are aware of the fact that many of the arguments we raise here have been made by others. At this point we do not provide an exhaustive reference list, and we make no claim to originality. Our thinking on these issues was greatly influenced by discussions with many people. In particular, we wish to thank Marion Ledwig, Stephen Morris, and Peter Wakker for comments and references.

with constrained maximization of a so-called utility function. By and large, economic theory does not attempt to judge which utility functions make sense, reflect worthy goals, or lead to beneficial outcomes. In essence, any utility function would suffice for an agent to be dubbed rational. More precisely, utility functions might be required to satisfy some mathematical properties such as continuity, monotonicity, or quasi-concavity. But these do not impose any substantive constraints on the subjective tastes of the economic agents involved. Defined solely on the abstract mathematical structure, these properties may be viewed as restricting the form of preferences, but not their content.

This minimalist requirement has two main justifications. The first is the desire to avoid murky and potentially endless philosophical discussions regarding the "true" nature of rationality. The second is that such a weak requirement does not exclude from the economic discussion more modes of behavior than are absolutely necessary. As a result, the theory is rather general and rational choice theory has indeed been applied in a variety of contexts, for a wide variety of utility functions.

A similarly minimalist definition was applied to the concept of belief. In an attempt to avoid the question of what it is rational to believe, as well as not to rule out possibly strange beliefs, the theory has adopted a definition of rational beliefs that is also based on internal consistency alone. Specifically, anyone who satisfies Savage's (1954) axioms, and behaves as if they entertain a prior probability over a state space, will be considered a rational decision maker under uncertainty, and may be viewed as having rational beliefs.

Such a relativist notion of rationality of belief is hardly intuitive. If John were to believe that he is the current King of France, and to take decisions in accordance with this view, he would pass the rationality test. Yet, it seems clearly irrational to entertain such beliefs despite evidence to the contrary. Similarly, ardently believing that the sun will not rise tomorrow would hardly qualify as rational by any intuitive sense of the word. In everyday parlance we make such distinctions between rational and irrational beliefs, and indeed, we confine people to institutions who hold beliefs of the sort that they are the King of France. Yet, decision theory is silent on this issue.[1]

As in the case of rationality of utility, reducing rationality of belief to internal consistency allows the theory of rational agents to apply rather widely. But this generality may be costly. First, when we refuse to address the question of what beliefs are rational, we may not notice certain regularities in the beliefs entertained by economic agents. Second, by restricting attention to the coherence of beliefs, one evades the question of the generation

---

[1] The notion of equilibrium in economics and in game theory may be viewed as an implicit definition of rational beliefs. That is, rational beliefs are those that coincide with equilibrium behavior. However, such a definition does not enlighten us about the process by which rational beliefs come into being.

of beliefs. Indeed, economic theory offers no account of the belief formation process. Beliefs are supposedly derived from observed behavior, but there is no description of how beliefs arose in the first place.

We believe that economic theory might benefit from a theory of belief formation, and, relatedly, a classification of beliefs according to their rationality. A theory of belief formation may suggest a systematic way of predicting which beliefs agents might hold in various environments. It may also delineate the scope of competing models for representation of beliefs. Rationality of beliefs may serve as a refinement tool.[2]

We argue that the Bayesian paradigm is lacking in that it precludes a theory of belief formation. Further, we argue that the Bayesian paradigm cannot be viewed as a definition of rational beliefs. Thus, both rationality of beliefs and belief formation suggest that we look beyond the Bayesian model for other types of representations of information and belief.

For many economists, the concept of rational beliefs is equated with behavior that might be viewed as being governed by Bayesian beliefs. We take issue with this point of view. First, we argue that rationality requires more than behavior that is consistent with a Bayesian prior. Second, we find that sometimes such behavior is too high a standard for any reasonable definition of rationality. Thus, we claim that behaving in accordance with the Bayesian model is neither sufficient nor necessary for rationality.

We begin by stating more clearly what we refer to by the term "Bayesianism". We then proceed to argue that, from a cognitive point of view, Bayesianism has several limitations as a methodology for the representation of beliefs. We next address the behavioral derivation of Bayesianism and explain why we find it neither a tenable theoretical argument for Bayesianism, nor a viable procedure for the elicitation of Bayesian beliefs. Finally, we conclude by arguing that Bayesianism is not a compelling definition of rationality.

2. What is Bayesianism?

The Bayesian paradigm is dominant, but precisely what does it mean to be Bayesian?

There are at least three tenets that are sometimes understood by this term:

<u>The first tenet of Bayesianism</u>: Whenever a fact is not known, one should have probabilistic beliefs about it. These beliefs may be given by a single probability measure defined over a state space in which every state resolves all relevant uncertainty. The

[2] In a sense, Cho-Kreps (1987) "intuitive criterion" and related belief-based refinements are among the few exceptions in which economic theory does dare to rule out certain beliefs on the grounds of irrationality. For a discussion of this literature, see Mailath, Okuno-Fujiwara, and Postlewaite (1993).

notion of relevance in this statement hints that one may satisfy the first tenet with respect to certain problems but not necessarily with respect to others. But the stronger form of the first tenet, which is quite popular in modern economic theory, ignores this qualification and assumes that one must be Bayesian with respect to anything one can conceive of.

The second tenet of Bayesianism: In light of new information, the Bayesian prior should be updated to a posterior according to Bayes's law.

The third tenet of Bayesianism: When facing a decision problem, one should maximize expected utility with respect to one's Bayesian beliefs (incorporating all information that one has gathered).

In statistics, computer science, artificial intelligence, and related fields, Bayesianism typically means only the first two tenets. In economics, by contrast, it is often coupled with the third tenet, which matches to the Bayesian approach a decision theory with clear behavioral implications. Conversely, these behavioral implications can be the basis of an axiomatization of the Bayesian approach, coupled with expected utility maximization, as in Ramsey (1931), de Finetti (1937), and Savage (1954). One may, however, also provide behavioral axiomatizations of the Bayesian approach with other decision theories, as in Machina and Schmeidler (1992). Both types of axiomatization could also be used, in principle, for the elicitation of subjective probabilities based on behavior data.

3. Limitations of Bayesianism

We hold that the Bayesian approach has several limitations, because of which it cannot be the only methodology for the representation of beliefs. In this sub-section we list three distinct (though somewhat related) reasons that the Bayesian paradigm may be found lacking for certain purposes.

3.1 Formation of beliefs

The Bayesian paradigm assumes a prior, and begins the process of learning with that prior. As such, it does not deal with the process of belief formation, and does not address the question of the rationality of prior beliefs. Bayesian learning means nothing more than the updating of a given prior. It does not offer any theory, explanation, or insight into the process by which prior beliefs are formed. Hence, we need to look beyond Bayesianism to cope with the questions of rationality of (prior) beliefs and of belief formation.

Why would belief formation be of interest to economists? Which beliefs agents entertain, in any given environment, is an empirical question. A good economic theorist would have a good guess regarding the beliefs that should be attributed to agents in her model. This guess can be tested directly, whether rigorously or intuitively, or indirectly, by

contrasting the model's predictions with observations. In any case, what does economic theory stand to gain from a theory of belief formation?

We believe that a formal, general theory of belief formation might benefit economic theory in several ways. First, the process of abstraction and generalization may always suggest new insights. If one understands why agents' beliefs in a given, successful model are appropriate, one may use a theory of belief formation to find similarly appropriate beliefs in other models. Second, it may be useful to have a theory that allows one to rank beliefs, attributed to agents within models, according to their justification, rationality, and degree of reasonability. Finally, a theory of belief formation may help us relate beliefs off the equilibrium path to beliefs at equilibrium. Currently, the theory offers neither explanation, nor systematic prediction of the way agents form beliefs once their original beliefs have been refuted.[3] If we were to confront the question of how beliefs come into being in the first place, we might be able to predict how beliefs change as a result of observations that are inconsistent with original beliefs. Thus one may attempt to apply a belief formation theory to the problem of backward induction play in complete information games. We argue that one might predict what beliefs one might generate by deviating from the backward induction path, and that this prediction suggests that the deviation might not be profitable.

3.2 Insufficient information

All three tenets of Bayesianism have come under attack, especially as descriptive theories. The most famous critique of Bayesian principles in economics is probably the descriptive failure of the third tenet, namely that people maximize expected utility when they are equipped with probabilistic beliefs. This critique started with Allais's famous example, and continued with the behavioral deviations from EUT documented by Kahneman and Tversky (Kahneman and Tversky, 1979). Much of the literature referred to as behavioral economics is based on this critique.[4] The second tenet of Bayesianism, namely, that beliefs should be updated in accordance with Bayes's law, is almost unassailable from a normative viewpoint.[5] But it has also been shown to be descriptively lacking (Tversky and Kahneman, 1974). Finally, the first tenet, namely, that one has probabilistic beliefs over anything uncertain, as been shown by Ellsberg (1961) to be an inaccurate description of people's behavior.

---

[3] Blume, Brandenburger, and Dekel (1991) provide an axiomatic derivation of beliefs on as well as off equilibrium paths. But their theory does not restrict off equilibrium beliefs in light of the prior beliefs, or the fact that these have been refuted.

[4] The term "behavioral" in "behavioral economics" may be misleading. In this usage, it refers to the study of actual behavior, as opposed to normative theory (as in "behavioral decision theory"). But "behavioral" should not be understood as focusing on behavior as opposed to cognition or to emotions. Neo-classical economics is behavioral in that it considers only behavior as valid data. Behavioral economics is less "behavioral" in this sense.

[5] Still, complexity considerations might render Bayesian updating a much more daunting task than it would seem in a basic probability class.

The first tenet is perhaps the weakest point of Bayesianism. There are many problems involving uncertainty, where there is simply not enough information to sensibly generate probabilistic beliefs. In these problems one may expect people to exhibit behavior that cannot be summarized by a single probability measure. Moreover, when information is scarce, one may also reject the Bayesian approach on normative grounds, as did Knight (1921). This normative failure is related to the limitation discussed above: the Bayesian paradigm does not offer a theory of (prior) belief generation. It follows that, even if one were convinced that one would like to be Bayesian, the Bayesian approach does not provide the self-help tools that would help one become Bayesian if one isn't.

This normative failure stimulated the work of Schmeidler (1989) and Gilboa and Schmeidler (1989). While the non-additive Choquet expected utility model (CEU, Schmeidler, 1989) and the multiple prior (maxmin) model (MMEU, Gilboa and Schmeidler, 1989) can be used to explain Ellsberg's paradox (1961), they were not motivated by descriptive arguments, but rather, they were motivated by the argument that the Bayesian approach is too restrictive to faithfully represent the information one has.

Consider the following example (Schmeidler, 1989). You are faced with two coins, each of which is about to be tossed. The first coin is yours. You have tossed it, say, 1000 times, and it has come up Heads 500 times, and Tails 500 times. The second coin is presented to you by someone else, and you know nothing about it. Let us refer to the first coin as the "known" one, and to the second as the "unknown" coin. Asked to assign probabilities to known coin coming up Heads or Tails, it is only natural to estimate 50% for each, as these are the empirical frequencies gathered over a sizeable database. When confronted with the same question regarding the unknown coin, however, no information is available, and relative frequencies do not offer any help in the estimation of probabilities. But the first tenet of Bayesianism demands that both sides of the unknown coin be assigned probabilities, and that these probabilities add up to 1. Symmetry suggests that these probabilities be 50% for each side. Hence, you end up assigning the same probability estimates to the two sides of the unknown coin as you did for the two sides of the known coin. Yet, the two 50%-50% distributions *feel* rather different. In the case of the known coin, the distribution is based on a good deal of information that leads to a symmetric assessment. In the case of the unknown coin, by contrast, the same estimates are based on the absence of information. The Bayesian approach does not allow us to distinguish between symmetry that is based on information and symmetry that is based on lack thereof.

One may embed this example in a decision problem, and predict choices as found by Ellsberg's in his two-urn example. But it is important that the example above does not involve decision making. The point of departure of Schmeidler (1989) is not the descriptive failures of subjective EUT. Rather, it is what one might call a sense of *cognitive unease* with the manner that the Bayesian paradigm deals with absence of information. This cognitive unease points also to the normative failure of the Bayesian

approach in this example. Even if one wished to become Bayesian, and even if one were willing to change one's choices so as to conform to the Bayesian paradigm, one must ignore the amount of information that was used in the generation of prior beliefs.

Ellsberg's paradoxes and Schmeidler's two-coin example are simple illustrations of the distinction between "risk" and "uncertainty" in Knight's terms. These examples can also be used to show that Savage's axioms (specifically, P2) may fail as descriptive theories. But both examples are simple in a way that might be misleading. These examples exhibit enough symmetries to suggest a natural prior via Laplace's "principle of insufficient reason". If one honestly wished to become Bayesian, one could easily assign 50% probability to each color in Ellsberg's two-urn experiment, and, similarly, 50% to each side of the unknown coin in Schmeidler's example. In both cases, the 50%-50% distribution is the only prior that respects the symmetry in the problem, and it is therefore a natural candidate for one's beliefs. Hence, considering these examples in isolation, one might conclude that, cognitive unease aside, it is fairly easy to become Bayesian even if one was not born Bayesian.

This conclusion would be wrong. Most decision problems encountered in real life do not possess sufficient symmetries for the principle of insufficient reason to uniquely identify a prior. Consider, for example, the uncertainty about an impending war. One cannot seriously suggest that the relative frequency of wars in the past may serve as a good estimate of the probability of a war at the present. The world changes in such a way that the occurrence of wars cannot be viewed as repeated identical and independent repetitions of the same experiment. Thus, the question of war is an example of uncertainty, rather than of risk. Applying Laplace's principle of insufficient reason would suggest that war has 50% probability. But such a claim is preposterous. We know enough about war and peace to rule out any possible symmetry between them. Indeed, we can reason at length about the likelihood of war, to have sufficient reason to reject the principle of insufficient reason.[6]

To conclude, a major failure of the Bayesian approach is that many real-life problems do not offer sufficient information to suggest a prior probability. In a small fraction of these problems there are symmetries that suggest a unique prior based on the principle of insufficient reason. But the vast majority of decision problems encountered by economic agents fall into a gray area, where there is too much information to arbitrarily adopt a symmetric prior, yet too little information to justifiably adopt a statistically-based prior.

[6] A related difficulty with the principle of insufficient reason is that it depends on the partition of the states of the world. In the absence of obvious symmetries, various partitions might be considered, leading to very different probability assessments.

3.3 The beliefs of society[7]

Bayesian statistics and classical statistics have often been viewed as competing paradigms. We suggest the view that these methodologies are designed to solve radically different problems. This point of view illustrates another weakness of the Bayesian approach.

Many would agree that Bayesian statistics is theoretically simpler and conceptually more coherent than is classical statistics. Whereas classical statistics resorts to concepts such as "confidence" and "significance", that are distinct from probability, Bayesian statistics only uses the concept of probability. The Bayesian approach needs to distinguish neither between objective and subjective sources of uncertainty, nor between quantified and unquantified uncertainty. Yet, in every day life, as well as in scientific contexts, classical statistics appears to be more popular. The reason has to do with the subjectivity inherent in the Bayesian approach.

Many textbooks illustrate the idea of hypothesis testing with a court case metaphor: the null hypothesis is that the defendant is innocent. The alternative is that the defendant is guilty. When we test the null hypothesis, we may find that we reject it, and the court then proclaims that the defendant was found guilty. In other words, if no sufficient evidence of guilt is supplied, the null hypothesis is maintained as a default. The court does not generally proclaim that the defendant is innocent, only that the evidence to the contrary leaves a reasonable doubt.

This is a theoretically cumbersome structure. The null hypothesis and the alternative are not treated symmetrically. A hypothesis may be rejected but not validated. The set of proclamations that may be attributed to the court fails to be closed under negation. Indeed, the logic of court decision and of hypothesis testing alike would have been much simpler if one were to adopt a Bayesian view. In this case, one would start out with a prior probability regarding the defendant's guilt, as well as on obtaining any piece of evidence conditional on guilt and on innocence, and, given the evidence actually presented, one would only have to update one's prior. If the posterior probability of guilt is high enough, one may proclaim the defendant guilty, and the same applies to innocence.

But who is this "one"? If the defendant's mother is to be asked, she may start out with a zero prior probability of guilt, and no amount of evidence would convince her that her child is actually guilty. On the other hand, the judge may have a prior according to which the defendant is almost surely guilty before even considering the evidence. But most people would prefer to live in a society where their innocence (as defined by the legal system) does not depend on the judge's prior. In other words, the Bayesian approach is

[7] This sub-section is based on Gilboa (1994).

inappropriate for this problem since it is inherently subjective, whereas the court system aspires to objectivity.

There are situations where social institutions are called upon to make public claims. Whether society should make a certain claim will often be a topic of debate. People would normally entertain different beliefs regarding the question at hand. Moreover, people who have different goals may have an incentive to argue that they have different beliefs. In light of these tensions, social institutions would tend to adopt methodologies that are akin to classical statistics in their quest for objectivity. Such tools may be able to make certain claims, and may have to remain silent on many others.

Individual decision makers are still free to draw on their intuition and subjective hunches, whatever the claims made or not made by society. For example, assume that a man is accused on having murdered his wife. Mary follows the case. She feels uneasy about the defendant, and finds, in fact, that quite likely that he is the murderer. She is willing to admit, however, that the evidence in the case is rather weak. It is therefore perfectly coherent for Mary to wish that the defendant be found not guilty, according to objectively available evidence, and yet to decide that she would not date the defendant. Similarly, for the question "Can society make a certain claim?" one would like to use tools of classical statistics, aiming at objectivity and shunning intuition. However, for the question, "Do I believe in this claim?" one would prefer Bayesian tools, attempting to incorporate all intuition and subjective impressions, and result in a probability assessment.

Another domain in which the Bayesian approach is precluded because of its subjectivity is science. Consider a hypothetical case of Dr. Strangelove who develops a new medication for AIDS. If Dr. Strangelove were to publish his Bayesian analysis of his medication, he could point out that, given his prior beliefs and the evidence, his posterior belief that his drug is safe and effective is very high. But this posterior may be mostly due to a prior that practically assumes this conclusion. Instead, Dr. Strangelove is asked to put his conclusions to the tests of classical statistics, and to show that, based on objective measure, his new medication is indeed safe and effective.

As in the case of Mary and her potential date, one may distinguish between social statements and individual decision making. Assume, for instance, that a respectable drug company intends to market Dr. Strangelove's medication. The company has applied to the FDA for approval of the new medication. This approval process is pending. That is, society has not yet endorsed the claims of the scientist and the company. Next consider John who suffers from AIDS. John fears that he is dying and he wishes to take the new medication. He reasons that the company would not risk its reputation and would not submit the medication for approval unless it was truly helpful. It is therefore reasonable for John to take the medication if it is available, even though most people, perhaps John

included, would prefer the FDA to complete its testing and approve the new medication independently of the reputation of the company producing it.

Observe that in this discussion we offer a possible view of what Bayesian and classical statistics *attempt* to achieve, not necessarily what they *do* achieve. Classical statistics attempts to be objective, but it is well known that it can be manipulated in various ways. More generally, many authors question the possibility, and indeed the very notion of objective science. On the other hand, Bayesian statistics attempts to capture intuition, but it may not always succeed in this goal. Indeed, the examples discussed in the previous sub-section suggest that not all intuition can indeed be captured in a Bayesian model. Our claim is that classical statistics may be viewed as aspiring to objectivity, whereas Bayesian statistics aspires to a unified treatment of the subjective and the objective.

To conclude, there is another reason to be interested in non-Bayesian paradigms of information and belief representation. Even if each individual has enough information for the generation of a Bayesian prior, individuals often differ in their beliefs. Recognizing this fact, society usually will choose to endorse statements only according to a well-defined protocol, such as those offered by classical statistics. Society will choose to remain silent on many issues. Thus, the set of statements that society will endorse typically does not suffice for a generation of a prior.

4. Behavioral derivations of Bayesianism

The limitations of the Bayesian approach mentioned above are cognitive in spirit: they deal with the degree to which a mathematical model captures our intuition when reasoning about uncertainty. The standard approach in economics, however, would find these cognitive limitations hardly relevant. Rather, following the revealed preference paradigm, most economists would suggest that decisions are eventually being made, and if these decisions satisfy (say) Savage's axioms, then expected utility maximization with respect to a Bayesian prior is an inevitable conclusion. Further, one's subjective probability may even be elicited from one's observed choices. Finally, Savage's axioms appear very compelling. It therefore appears that any deviation from subjective EU maximization violates a certain canon of rationality.

We devote this sub-section to a criticism of this argument. We first draw a distinction between two possible interpretations of a preference relation: raw preferences and reasoned choice. We then address each interpretation and argue that, with this interpretation in mind, one cannot derive the Bayesian approach from Savage's axioms. That is, it is misleading to argue that the axioms necessitate the existence of a prior, and eliciting the prior from behavior is not a viable alternative. Finally, we conclude by arguing that Savage's axioms are neither sufficient nor necessary for rationality.

4.1 Raw preferences and reasoned choice

We need to address both the descriptive and the normative interpretation of Savage's axioms.[8] But it may be more efficient to split the discussion along slightly different lines. Consider a binary relation representing preferences, or choices, as in Savage's theory. This relation can be interpreted in (at least) two ways. First, it might reflect *raw preferences*, namely, the decision maker's almost-instinctive tendency to prefer one alternative over another. Second, the same binary relation might model *reasoned choice*, namely, choice that was arrived at by a process of reasoning. Roughly, the decision maker exhibits raw preferences if she first acts, then observes her own act and stops to think about it. The decision maker is involved in reasoned choice if she first thinks, then decides how to act.

The process of reasoning may be modeled by logic. A reason to, say, prefer alternative f over g is akin to a proof, where from certain assumptions one derives the conclusion that f should be chosen over g. With different sets of assumptions one may generate reasons for and, simultaneously, against certain choices.

A descriptive interpretation of preferences in Savage's model may be either one of raw preferences or of reasoned choice. When describing reality, one has to cope with the fact that in certain decision problems the decision maker acts first, and thinks later (if at all), whereas in others she may reason her way to a decision. By contrast, a normative interpretation of Savage's theory deals with reasoned choice: if one attempts to convince a decision maker to change their decision, they normally provide reasons to do so.

The concept of "reasoned choice" may also be extended to beliefs. One might distinguish among three levels of rationality of beliefs. The lowest degree of rationality is attributed to beliefs that are contrary to evidence, as in the case of a person who insists that he is the King of France. The highest degree of rationality would be reserved to belief that is justifiable by evidence, or to *reasoned belief*. In between one may find beliefs that are neither justified not contradicted by evidence.[9]

We started by observing that the commonly accepted Bayesian approach reduces rationality to internal consistency, allowing beliefs that are contrary to evidence. The Bayesian approach, however, does not necessitate such plainly irrational beliefs. By contrast, the Bayesian approach does require that beliefs be stated even in the absence of evidence. Thus, it insists on specification of beliefs beyond reasoned belief.

---

[8] The same discussion can be conducted in the context of any other behavioral axiomatization of Bayesianism. We choose to refer to Savage as his is, justifiably, the most well-known axiomatization.
[9] Similarly, rationality might be applied to absence of beliefs. For instance, if John insists that he does not believe that the Earth is flat, nor that it is round, this agnosticism may be viewed as irrational.

4.2 Derivation of Bayesianism from raw preferences

We argued that in many problems there is insufficient information for the generation of a (unique) prior probability measure. One might wonder whether raw preferences do not provide this information indirectly. After all, as long as Savage's axioms are satisfied, one may elicit from preferences a unique probability measure. It would therefore appear that in order to have a prior all one needs to know is one's own preferences. There are several related reasons for which we find this conclusion unwarranted.

Unreasoned preferences need not satisfy Savage's axioms

One may argue that reasoned preferences should adopt EUT as a normative theory, and attempt to satisfy its axioms. We comment on the appropriateness of the axioms for reasoned choice below. By contrast, raw preferences need not satisfy Savage's axioms. Indeed, there is plenty of evidence that in certain situations EUT is consistently being violated by many decision makers. Hence, using raw preferences as the starting point from which beliefs would emerge is a dubious endeavor.

Beliefs precede choices

Following logical positivism, economic theory in the 20th century opted for a behavioral approach, according to which intentional concepts such as "utility" and "belief" were considered theoretical constructs that need to be derived from observations. The revealed preference paradigm further holds that legitimate observations for economics can only be observed choice behavior, ruling out other sources of information such as introspection and subjective reports of preferences, likelihood judgments, and the like. The derivations of subjective EUT by Ramsey, de Finetti, and Savage fit this approach.[10]

The behavioral derivations of utility and of belief can be interpreted in several ways, depending on the degree of awareness of the decision maker. A maximalist interpretation would hold that the decision maker has direct access to a utility function and to a probability measure, and that the theory actually describes the decision making process. A minimalist interpretation, by contrast, would take no stance on the question of the decision maker's awareness. Rather, it would suggest that, for an outside observer, it suffices that a decision maker who satisfies a certain set of axioms can be described as if she were following a certain decision rule. Another possible interpretation is that the decision maker does not have direct access to her utility function and probability measure, but that she does have access to her own preferences, and that she can consequently derive her utility and probability from her choices, as could an external observer. It is this interpretation that we discuss here.

[10] De Finetti and Savage also discussed qualitative probabilities, that may be viewed as observable "at least as likely as" judgments. Thus, their works can also be interpreted as deriving probabilities based on subjective reports, rather than on actually observable choice.

Using an axiomatization to learn something about oneself is often rather plausible. Consider the case of a utility function. Consumers are generally expected to satisfy the axioms of completeness and transitivity without being aware of their own utility function. They are not even supposed to be aware of their preferences in any cognitive sense. Rather, it suffices that they know observed choice. Indeed, a consumer may find that she doesn't like a certain brand of juice by discovering that she has bought it but never got to consume it. In this sense, "preference needs no inference": choices are being made by consumers without them having to go through complicated mental procedures such as maximization of utility functions.

In certain restricted examples, a similar interpretation might hold also for beliefs. A driver who buckles up only when she drives on a highway might infer that she assigns a higher probability to the event of a deadly accident when driving on the highway as compared to city driving. Such a choice could perhaps be made without an explicit decision making process that estimates probabilities.

But in many decision situations, this interpretation is hardly valid. To consider an extreme example, assume that Mary is faced with the uncertainty about the truth of a mathematical conjecture. If Mary were Bayesian, she would have an exact probability p that the conjecture is true. Since she may not have enough information to generate such a prior, the present approach suggests that Mary introspect and figure out her preferences for various bets on the correctness of this conjecture. Thus, Mary is expected to ask herself, say, if she prefers to bet $100 on the correctness or on the falsehood of the conjecture. But Mary will soon find out that she does not have the foggiest idea about her own preferences in this case. Her only way to decide what her preferences are is to start thinking whether the mathematical conjecture is likely to be true.

To consider a more relevant example, assume now that Bob is asked what the probability of a nuclear war in Asia is in the next five years. Bob does not have enough statistical evidence to generate a prior based on past frequencies. If he were to adopt the approach suggested here, he should be asking himself whether he prefers to get an outcome x if such a war erupts, and an outcome y otherwise, to, say, a sure outcome z. Again, Bob cannot be expected to have well-defined accessible preferences over such choices. Rather, he would have to stop and ponder. Only after he assesses the probability of war can he meaningfully answer these preference questions.

One might argue that Bob's decisions implicitly define his beliefs even if Bob himself is not aware of these beliefs. For example, Bob's trading in the stock market (or lack thereof) indirectly reveals his preferences over bets involving such a war. Indeed, if Bob did bother to think about the problem and if he came up with a prior probability, the latter may be reflected in his investment choices. But if Bob has not made a conscious estimation of probabilities, there is no reason to expect his choices to satisfy Savage's axioms.

To conclude, in situations where there is not sufficient information to generate a prior, we may not assume that decision makers know all their relevant preferences. In fact, there are many such situations in which only reasoned choice is possible, and raw preferences simply do not exist.

The unobservability of preferences

The notion of "observability", in decision and in economic theory alike, allows some freedom of interpretation. Most theories in economics and in related disciplines have paradigmatic applications, which leave little room for various interpretations. For example, it is convenient to think of Savage's axioms in the context of bets on a color of a ball drawn at random from an urn. In this context, the states of the world are clearly defined by the balls in the urn. Choices made contingent on the color of the ball drawn can be thought of as direct observations of a preference relation in a Savage-type model. But most economic applications of subjective EUT do not have such a clearly defined state space. Worse still, it is often not clear the state space the decision maker has in mind.

Assume that Mary is considering quitting her job and taking another job offer. She presumably considers her potential promotion in the new firm. This decision problem involves uncertainty and can be couched in a Savage-type model. A state of the world in such a model should specify the values of all variables that Mary deems relevant. For instance, a state should be informative enough to determine Mary's salary in a year's time. Mary might also consider the expertise she might acquire on the job, the people she is likely to work with, the economic stability of her employer, her own job security, and a number of other variables, for any time horizon she might plan for. Moreover, the set of states of the world should also allow all possible causal relationships between Mary's actions and these variables. For example, Mary should take into account that with one employer her promotion is practically guaranteed, while with another it depends on her effort level. The nature of the causal relationship between effort and promotion is also subject to uncertainty, and should therefore also be specified by each state of the world.

These considerations give rise to two difficulties. First, the state space becomes large and complicated, and with it – the set of conceivable acts defined on this state space in Savage's mode. Second, it is not at all clear which of the above uncertainties are taken into account by Mary. We may observe Mary's decision regarding the new job offer, but not the states of the world that she entertains in her mind. It follows that, having observed Mary's choice, we may construct many different Savage-type models in which her choice is modeled as a preference between two acts. But in each such model there will be many other pairs of acts, the choice between which was not observed. Clearly, if the states of the world themselves are unobservable, one cannot hope to observe a complete binary relation between all the acts defined on these states.

A possible solution to the second problem would be to define an exhaustive state space, within which one may embed every conceivable state space that the decision maker has in mind. But such a solution aggravates the first problem. In fact, it renders most pairwise choices inherently unobservable. To see this, imagine that one defines the set of outcomes to include all conceivable consequences, over any time horizon. One then proceeds to define states as all possible functions from acts to outcomes. This would result in an exhaustive, canonical state space. Next, one must define all the conceivable acts (Savage, 1954): all possible functions from states to outcomes. Over this set of conceivable outcomes one assumes that a complete binary relation is observable, and that the observed choices would satisfy Savage's axioms. But such a relation cannot be observable even in principle. In this states are functions from actual acts to outcomes, and conceivable acts are functions from these states to the same set of outcomes. Thus the set of conceivable acts is by two orders of magnitudes larger than the set of acts that are actually available in the problem. This implies that the vast majority of the binary choices assumed in Savage's model are not observable, even in principle.

Savage's theory has a clear observable meaning in experiments involving simple set-ups such as balls drawn out of urns. But in many economic applications of EUT the state space is not directly observable, and hence Savage's behavioral axioms do not have a clear observable meaning. In particular, observing actual behavior does not contain enough information for the elicitation of a prior over the state space.[11]

4.3 Derivation of Bayesianism from reasoned choice

<u>Reasoned choice need not be complete</u>

The completeness axiom is typically justified by necessity: one must make a decision, and whatever one chooses will be viewed as the preferred act. This argument seems to apply to observed preferences.[12] Indeed, if one defines preference by observations, the completeness axiom is almost tautological.[13] But when we consider reasoned choice, there does not seem to be a compelling argument for completeness.

To see this point more clearly, consider first the case of transitivity. If there is a reason to prefer f to g, and if there is a reason to prefer g to h, then these two reasons may be combined to provide a reason to prefer f to h. The transitivity axiom may actually be

[11] In the construction mentioned above it is also not clear whether Savage's axioms are satisfied by actual decision makers, since we can never observe more than a fraction of the pairwise choices referred to by the axioms.

[12] As mentioned above, even with observed preferences this argument is questionable. In many applications of Savage's theorem one may not assume that all acts are possible candidates for observable choice.

[13] Completeness is not quite a tautology because it is often taken to mean that preferences within each pair of acts will be the same in repeated experiments of the same decision problem. However, as opposed to the other axioms, completeness cannot be refuted by single observations of distinct binary choices.

viewed as a reasoning axiom, providing an argument, or a justification for a certain preference. It can similarly be used to infer certain preferences from others. Similarly, Savage's axioms P2, P3, and P4 can be viewed as templates for reasoning that a certain act should be preferred to another.[14] The same cannot be said of the completeness axiom. The completeness axioms states that (reasoned) choice should be defined between any two acts *f* and *g*, but it provides no help in finding reasons to prefer *f* to *g* or vice versa.

If one views Savage's axioms as conditions on raw preferences, the completeness axiom may be mentioned as a half-axiom barely worth mentioning. Completeness in this set-up is one out of two requirements in P1, and it barely calls for elaboration. But if the Savage axioms are viewed as conditions for reasoned choice, the completeness axiom plays an altogether different role: it is contrasted with all the rest. The completeness axiom defines the question, namely, what are the reasoned preferences between pairs of acts, and all the rest are part of the answer, that is, potential reasons that may come to bear in determining preferences between particular pairs of acts.

Computational complexity

When decision problems do not present themselves to the decision maker with a clearly defined state space, the generation of all relevant states involves insurmountable computational difficulties. For example, assume that we are given *m* variables that may serve as predictors for another variable. We prefer to have a subset of the predictors that yields a good fit with relatively few variables. For each subset of the predictors there is a state of the world in which this subset is the most appropriate set of predictors, say, the set that obtains the highest adjusted R^2. Thus, there are exponentially many states of the world, relative to the size of the database. It is therefore far from trivial to list all these states, let alone the entire decision matrix. In Aragones, Gilboa, Postlewaite, and Schmeidler (2003) we show that it is NP-Hard to determine whether, for a given database, and a given *k*, there are *k* predictors that obtain a pre-specified level of R^2.

To take another example, consider contract theory. It is typically assumed that a contract specifies outcomes given states of the world. But the language in which contracts are stated involves clauses and conditional statements, rather than specific states. While it is possible to define a state of the world as a truth table, assigning truth value to each relevant proposition, the number of states is exponential in the size of the contract. Indeed, in Gilboa, Postlewaite, and Schmeidler (2003) we show that, given a contract in its clause form, it is NP-Hard to determine the most basic questions about it, such as whether it is complete, consistent, or whether it has a positive expected value for a given party (even if probabilities of events are given).

[14] To some degree, continuity axioms such as P7 and even P6 can also be viewed as reasoning axioms. But for the purposes of the present discussion it is best to focus on the axioms that are more fundamental from a conceptual viewpoint.

These considerations suggest that the Savage axioms may be a too demanding standard of rationality. A normative theory may not be useful if it requires agents to solve NP-Hard problems. We should not assume that economic agents could solve problems that computer scientists cannot solve.

Other difficulties

The elicitation of beliefs from reasoned choice encounters two additional problems that were discussed also for raw preferences. First, the observability problem means, for reasoned choice, that defining one's preference relations requires a large degree of hypothetical reasoning. Second, reasoned choice might contradict other axioms of Savage, beyond the completeness axiom. While these axioms are themselves reasons for a particular preference, they may generate conflicting preferences. The only algorithm that would guarantee a resolution of these conflicts in a way that satisfies the axioms would be to select a probability and a utility function and to make decision in accordance with EUT. This, however, brings us back to the task of specifying a prior directly, rather than deriving it from choices.

4.4 Conclusion

To conclude, we hold that Savage's axioms cannot in general serve as a standard of rationality. We argued above that the axioms are too strong, namely, that they are not necessary conditions for rationality to hold. In Section I we claimed that, in other ways, Savage's axioms are too weak and that they do not constitute a sufficient condition for rationality. This claim deserves elaboration.

We do not believe that criteria of internal consistency can serve as a satisfactory definition of rationality. Assume that John makes decisions in a way that violates some of Savage's axioms. His inconsistent choices are being presented to him, together with the result, namely, that he has failed the rationality test. Assume further that John is eager to obtain the highly coveted title, "a rational decision maker". He asks a decision theorist how he can do this. The theorist replies, "Oh, that's easy. You just pick a prior probability and a utility function, and from now on all you do is maximize the expectation of this utility with respect to this prior. I can help you with the calculations if you wish. If you still find this hard, however, let me mention that this is the *only* way in which you can pass Savage's test. There is a theorem to this effect. So being rational is precisely as easy (and as hard) as picking a utility-probability pair and maximizing the integral."

Assume that John accepts the advice, and arbitrarily selects a utility function and a probability measure. He then dutifully maximizes expected utility. Next time he is tested for rationality, his preferences are shown to satisfy Savage's axioms. John gets to be accredited as "rational decision maker under uncertainty".

Would we endorse this accreditation? Probably not automatically. It is after all possible that John is taking decisions in accordance with the belief that he is the King of France. There is, however, plenty of evidence that he is not. We would typically expect a rational decision maker to base her beliefs on some evidence, or at least to make sure that these beliefs do not contradict overwhelming evidence. It follows that Savage's consistency axioms, as well as any other set of axioms that confine themselves to internal consistency, are insufficient for a definition of rationality. Rather, an axiomatic approach to rational beliefs would require, inter alia, a treatment of the relationship between beliefs and evidence.

Gilboa and Schmeidler (2003) offer a model in which likelihood rankings are parameterized by a database of cases. But even this model says nothing about the relationship between a past case and future predictions, and allows counter-intuitive predictions (as long as they are aggregated in accordance with a set of axioms). An intuitive definition of rational beliefs requires additional structure that would allow one to state which specific predictions are more likely than others, given specific data. Such statements are bound to be controversial, and they will be open to criticism along the lines of Goodman's "grue-bleen" paradox.[15] One should expect that the truthfulness of such statements will not be judged based on a-priori reasoning alone, but also based on empirical investigations.

Economic theory seems to be living in interesting times. Over several decades, economics incorporated decision making under uncertainty into the full range of economic problems. Overwhelmingly, the specific model of decision making has been Savage's model. While this enterprise has led to many important insights, many economists have become increasingly aware of the limitations of this model and engaged in dialogues with disciplines such as psychology and computer science in search for alternatives or augmentations to the Savage model to address these limitations. It is too early to predict which of the methodologies offered by these fellow disciplines will enter the canon of economic thought. But there is reason to hope that this search will provide economic theory with a notion of belief and a theory of decision making that will be better justified and hence more rational than the idealization offered by the Bayesian approach.

[15] The "grue-bleen" paradox shows that certain predicates are natural, while others are not, in a way that can hardly be justified based solely on logical or a-priori reasoning. In particular, "green" and "blue" are natural predicates. By contrast "grue", which means "green until 2010 and blue thereafter", and "bleen", which means "blue until 2010 and green thereafter", are not. The choice of "natural" predicates also has implications to the choice of "natural" theories and to the intuitive selection of methods of induction. All scientific predictions are based on implicit assumptions regarding the correct choice or predicates, and a theory of belief formation should probably not aspire to be free of these assumptions.

References

Aragones, E., I. Gilboa, A. Postlewaite, and D. Schmeidler (2003), "Fact-Free Learning", mimeo.

Blume, L., A. Brandenburger, and E. Dekel (1991), "Lexicographic Probabilities and Choice Under Uncertainty", Econometrica, 59: 61-79.

Cho, I.-K. and D. M. Kreps (1987), "Signaling Games and Stable Equilibria", The Quarterly Journal of Economics, 102:179-222.

de Finetti, B. (1937), "La Prevision: Ses Lois Logiques, Ses Sources Subjectives", Annales de l'Institute Henri Poincare, 7: 1-68.

Ellsberg, D. (1961), "Risk, Ambiguity and the Savage Axioms", Quarterly Journal of Economics, 75: 643-669.

Gilboa, I. (1994), "Teaching Statistics: A Letter to Colleagues", mimeo.

Gilboa, I., and D. Schmeidler (1989), "Maxmin Expected Utility with a Non-Unique Prior", Journal of Mathematical Economics, 18, 141-153.

Gilboa, I., and D. Schmeidler (2003), "Inductive Inference: An Axiomatic Approach", Econometrica, 71: 1-26.

Gilboa, I., A. Postlewaite, and D. Schmeidler (2003), "Contracts are Complex", mimeo.

Kahneman, D. and A. Tversky (1979), "Prospect Theory: An Analysis of Decision Under Risk," Econometrica, 47: 263-291.

Knight, F. H. (1921), Risk, Uncertainty, and Profit. Boston, New York: Houghton Mifflin.

Machina, M. J. and D. Schmeidler (1992), "A More Robust Definition of Subjective Probability", Econometrica, 60: 745-780.

Mailath, G., M. Okuno-Fujiwara, and A. Postlewaite, (1993), "On Belief Based Refinements in Signaling Games," Journal of Economic Theory, 60: 241-276

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.

Schmeidler, D. (1989), "Subjective Probability and Expected Utility without Additivity", Econometrica, 57: 571-587.

Tversky, A., & Kahneman, D. (1974), "Judgment under uncertainty: Heuristics and biases", Science, 185: 1124-1131.