Ambiguity and the Bayesian Paradigm*

Itzhak Gilboa[†] and Massimo Marinacci[‡]

April 2011

Abstract

This is a survey of some of the recent decision-theoretic literature involving beliefs that cannot be quantified by a Bayesian prior. We discuss historical, philosophical, and axiomatic foundations of the Bayesian model, as well as of several alternative models recently proposed. The definition and comparison of ambiguity aversion and the updating of non-Bayesian beliefs are briefly discussed. Finally, several applications are mentioned to illustrate the way that ambiguity (or "Knightian uncertainty") can change the way we think about economic problems.

1 Introduction

1.1 Varying probability estimates

John and Lisa are offered additional insurance against the risk of a heart disease. They would like to know the probability of developing such a disease over the next ten years. The happy couple shares some key medical parameters: they are 70 years old, smoke, and never had a blood pressure problem. A few tests show that both have a total cholesterol level of 310 mg/dL, with HDL-C (good cholesterol) of 45 mg/dL, and that their systolic blood pressure is 130. Googling "heart disease risk calculator", they find several sites that allow them to calculate their risk. The results (May 2010) are:

	John	Lisa
Mayo Clinic	25%	11%
National Cholesterol Education Program	27%	21%
American Heart Association	25%	11%
Medical College of Wisconsin	53%	27%
University of Maryland Heart Center	50%	27%

*We thank Giulia Brancaccio, Simone Cerreia-Vioglio, Fabio Maccheroni, Andrew Postlewaite, Xiangyu Qu, and David Schmeidler for comments on earlier drafts of this survey. We are also grateful to many members of the "decision theory forum" for additional comments and references. Gilboa gratefully acknowledges the financial support of the Israel Science Foundation (grant 396/10) and of the European Reseach Council (advanced grant 269754), and Marinacci that of the European Reseach Council (advanced grant BRSCDP-TEA). An abridged version of this survey will appear in "Advances in Economics and Econometrics: Theory and Applications", Tenth World Congress of the Econometric Society (D. Acemoglu, M. Arellano, and E. Dekel, eds.), Cambridge University Press.

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

[‡]Università Bocconi, massimo.marinacci@unibocconi.it

The different calculators don't completely agree on the probability in question. Partly, the reason is that no two human bodies are identical, and therefore simple relative frequencies cannot be used as a definition of "the probability of a heart disease". Rather, the probabilities above are computed by more sophisticated techniques such as logistic regression and variants thereof. These techniques allow researchers to assess probabilities for different individuals depending on their characteristics. But the resulting numbers are not perfectly objective: the researchers have to choose the variables (predictors), the database, as well as the estimation technique itself. Consequently, the estimated probability is not unique. The estimates vary substantially: the highest for John is 100% higher than the lowest, whereas for Lisa the ratio is 5:2. Moreover, gender differently affects the estimated numbers across risk calculators. Clearly, different probability estimates in the range of 25%-50% and 11%-27% may result in different decisions.

Next consider Mary, who contemplates an investment in a beach resort. The profitability of this venture depends on global warming: the resort will not be very successful if, due to climate changes, the beach becomes much rainier than it is now, or if the oceans' level increases significantly. Mary wonders what the probabilities of these eventualities are.

For this problem the estimation of probabilities seems conceptually harder. We have but one globe to base our predictions on. The history of global warmings in the past is only indirectly attested to, and the number of past cases is not very large. Worse still, current conditions differ from past ones in significant ways, especially in the much-discussed human-generated conditions. Finally, it is not obvious that one can assume causal independence across different warmings. Due to all these difficulties, it is perhaps not very surprising that the estimates of the distribution of average temperature several years hence vary considerably across experts.

1.2 Does rationality necessitate probability?

Since the mid-20th century, economic theory is dominated by the Bayesian paradigm, which holds that any source of uncertainty can and should be quantified probabilistically.[1] According to this view, John and Lisa should have well-defined probabilities that they will develop a heart disease within the next ten years, as should Mary for the temperature distribution anywhere on the globe five years hence. But where should John, Lisa, or Mary get these probabilities from? If they are to consult experts, they will typically obtain different estimates. Which experts are they to believe? Should they compute an average of the experts' estimates, and, if so, how much weight should each expert have in this average?

The standard line of reasoning of the Bayesian approach is that, in the absence of objective

[1] As Cyert and DeGroot (1974) write on p. 524 "To the Bayesian, all uncertainty can be represented by probability distributions."

probabilities, the decision maker (DM, for short) should have her own, *subjective* probabilities, and that these probabilities should guide her decisions. Moreover, the remarkable axiomatic derivations of the Bayesian approach (culminating in Savage, 1954), show that axioms that appear very compelling necessitate that the DM behave as if she maximized expected utility relative to a certain probability measure, which is interpreted as her subjective probability. Thus, the axiomatic foundations basically say, “Even if you don’t know what the probabilities are, you should better adopt some probabilities and make decisions in accordance with them, as this is the only way to satisfy the axioms.”

There is a heated debate regarding the claim that rationality necessitates Bayesian beliefs. Knight (1921) and Keynes (1921, 1937) argued that not all sources of uncertainty can be probabilistically quantified. Knight suggested to distinguish between “risk”, referring to situations described by known or calculable probabilities, and “uncertainty”, where probabilities are neither given nor computable. 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.”

Gilboa, Postlewaite, and Schmeidler (2008, 2009, 2010) argue that the axiomatic foundations of the Bayesian approach are not as compelling as they seem, and that it may be irrational to follow this approach. In a nutshell, their argument is that the Bayesian approach is limited because of its inability to express ignorance: it requires that the agent express beliefs whenever asked, without being allowed to say “I don’t know”. Such an agent may provide arbitrary answers, which are likely to violate the axioms, or adopt a single probability and provide answers based on it. But such a choice would be arbitrary, and therefore a poor candidate for a rational mode of behavior.

The notion that rational individuals would select probabilities to serve as their subjective belief might be particularly odd in the context of a market. Consider John and Lisa’s problem again. Suppose that they receive a phone call from an agent who offers them an insurance policy. They find the policy too expensive. In the process of negotiation, the insurance agent quotes the Medical College of Wisconsin’s site and estimates the probability of John and Lisa developing a heart disease within ten years at 53% and 27%, respectively. John and Lisa, by contrast, looked up the site of the Mayo Clinic and concluded that these probabilities are only 25% and 11%. They decide not to buy the insurance, and the conversation ends by John, Lisa, and the agent politely agreeing to disagree. After John and Lisa hang up, they look up a few other sites and realize that there are estimates that indeed place the probability at around 50% and 27%. They might think, "How can we be so sure that 25% and 11% are indeed the right numbers? We are glad we didn't say 50% and 27% when talking to this agent, but, truth be told, we don't know where it is between 25% and 50% for John and between 11% and 27% for Lisa".

Axiomatic derivations such as Savage's may convince the DM that she ought to have a probability, but they do not tell her which probability it makes sense to adopt. If there are no additional guiding principles, an agent who picks a probability measure arbitrarily should ask herself, is it so rational to make weighty decisions based on my arbitrarily-chosen beliefs? If there are good reasons to support my beliefs, others should agree with me, and then the probabilities would be objective. If, however, the probabilities are subjective, and others have different probabilities, what makes me so committed to mine? Wouldn't it be more rational to admit that these beliefs were arbitrarily chosen, and that, in fact, I don't know the probabilities in question?

1.3 Modeling diverging opinions

Economic reality provides a host of examples of events for which there is no objective, agreed-upon probability. For example, internet betting sites suggest that people hold different beliefs over many events. The volume of trade in the stock market also raises a serious doubt about the assumption that all agents have a common probabilistic belief. Theoretical results such as the impossibility of agreeing to disagree (Aumann, 1976) and the no-trade theorem (Milgrom and Stokey, 1982) further question the agreement among beliefs.[2] But the most compelling evidence may be self-reported beliefs: people state assessments and opinions that cannot be reconciled with common probabilistic beliefs. As pointed out in the examples above, experts may provide different assessments of probabilities for such events as the onset of a heart disease or global warming. And it suffices to leaf through any daily newspaper to see that people disagree about the prediction of rates of growth and of inflation, about the success of products, the intentions of world leaders, and so forth.

In view of these disagreements, one may relax the assumption that agents are Bayesian, or defend it by assuming that different agents have different subjective beliefs. As mentioned above, we think that it is not entirely rational for an agent to hold on to particular probabilistic beliefs though she knows that other, equally reasonable DMs, entertain different beliefs. Relatedly, at a steady state one would not expect different agents to hold different beliefs.

We therefore hold that a potentially fruitful path is to relax the assumption that agents are Bayesian. This does not mean that agents never think in probabilistic terms or that they fail to perform probability calculations correctly. It only means that regarding *some* events the rational

[2] The no-trade theorem does not depend on there being a shared prior, but it does depend on the assumption that agents have probabilistic beliefs. See section 6 below.

agents are allowed to say "the probability of this event isn't precisely known". Indeed, in empirical work economists use classical statistics techniques, such as hypotheses tests and confidence intervals. These techniques are non-Bayesian, and reflect the researcher's state of knowledge, specifying the objectively given distributions and remaining silent on the unknown parameters. A non-Bayesian approach allows the agents in the model to be as aware of their own ignorance as are economists.

There are many uncertain situations for which one can assume that probabilities are known, or agreed upon. The most convincing examples are games of chance, involving fair dice and coins, or balls that are randomly drawn from urns. Unfortunately, these examples tend to pop up mostly in the context of gambling or probability courses. As one moves on to analyze more realistic examples, it becomes harder to justify the assumption that agents have known probabilities. Even examples that involve repetitions, such as insurance problems, already begin to compromise the notion of objective probability, as explained above. Worse still, many of the important events in our lives are never repeated in the same, or even in a similar way. This class includes wars and stock market crashes at the economy level, as well as losing one's job or getting a divorce at the individual level. For such events it is difficult to assign a single probability number. For theoretical purposes, it is convenient to assume that each event A has a probability p, shared by all agents. But as we argue below, such an assumption may lead us to wrong conclusions.

1.4 Outline

The rest of this paper is organized as follows. Section 2 discusses the history and background of the Bayesian approach. It highlights the fact that this approach has probably never been adopted with such religious zeal as it has within economic theory over the past 60 years. Section 3 describes several alternatives to the standard Bayesian model. It surveys only a few of these, attempting to show that much of the foundations and machinery of the standard model need not be discarded in order to deal with uncertainty. Section 4 surveys the notion of ambiguity aversion. The updating of non-Bayesian beliefs is discussed in Section 5. Section 6 briefly describes some applications of non-Bayesian models. The applications mentioned here are but a few examples of a growing literature. They serve to illustrate how non-Bayesian models may lead to different qualitative predictions than Bayesian ones. A few general comments are provided in Section 7.

2 History and background

2.1 Early pioneers

Decision theory was born as a twin brother of probability theory through the works of a few scholars in the 16th and 17th century, originally motivated by the study of games of chance. Among them the works of Christiaan Huygens (1629-1695) and Blaise Pascal (1623-1662) are particularly relevant. We begin with Pascal, whose footsteps Huygens followed.

Pascal (1670) Since its very early days, probability had two different interpretations: first, it captures the notion of *chance*, referring to relative frequencies of occurrences in experiments that are repeated under the same conditions. This includes the various games of chance that provided the motivation for the early development of the theory. Second, probability can capture the notion of *degree of belief*, even when no randomness is assumed, and when nothing remotely similar to a repeated experiment can be imagined.

It is this second interpretation that, over time, evolved into the Bayesian approach in both decision and probability theory. In this regard, Pascal is perhaps the most important pioneer of probability theory. Though he made early key contributions to the probabilistic modeling of games of chance, it is his famous wager that is mostly relevant here. Roughly at the same time that Descartes and Leibniz were attempting to prove that God existed,[3] Pascal changed the question from the proof of existence to the argument that it is worthwhile to believe in God, an option that he identified with the choice of a pious form of life based on the precepts of the Christian religion ("taking the holy water, having masses said ... [being] faithful, humble, grateful, generous, a sincere friend, truthful ... [without] those poisonous pleasures, glory and luxury;" Serie II in Pascal, 1670).[4] In so doing, he applied the mathematical machinery developed for objective probabilities in games of chance to the subjective question of God's existence, where no repeated experiment interpretation is possible. This led him to informally introduce several major ideas of modern decision theory.[5]

Specifically, Pascal considers the choice between two acts, a pious form of life and a more worldly one. He recognizes that their consequences depend on what today we would call two states of the world, that is, whether or not God exists ("God is, or He is not."). Thus, from his verbal description the following decision matrix can be inferred:

	God exists	God does not exist
Pious	Salvation ("Gain all")	Constrained Life
Worldly	"Error and misery"	Unconstrained Life

[3] Pascal was younger than Descartes (1596-1650) and older than Leibniz (1646-1716).

[4] According to Pascal a pious life would ultimately induce faith, "You would soon have faith, if you renounced pleasure" in his words. It is important, however, that Pascal did not assume that one can simply choose one's beliefs.

[5] Pascal did not finish his *Pensées*, which appeared in print in 1670, eight years after his death. The text that was left is notoriously hard to read since he only sketches his thoughts (here we use the 1910 English edition of W. F. Trotter). Our rendering of his argument crucially relies on Hacking (1975)'s interpretation (see Hacking, 1975, pp. 63-72, and Gilboa, 2009, pp. 38-40).

that is, by quantifying consequences (in utils) and beliefs,

	p	$1-p$
Pious	$+\infty$	a
Worldly	b	c

with $c > a$ and $b \in [-\infty, \infty)$.[6] The very formulation of the matrix, with its distinction between acts (the actual object of choice) and states (over which DMs have no influence), is a first breakthrough. Pascal observes that for an individual that does not value pleasurable sins, to be pious is what today we would call a weakly dominant alternative ("if you gain, you gain all; if you lose, you lose nothing"). But, even for individuals who value these sins, to be pious is a better alternative as long as they attach a positive, however small, probability to the existence of God ("a chance of gain against a finite number of chances of loss"). For, salvation is infinitely better than "error and misery" and so the expected utility of being pious is higher than that of the worldly alternative. The notions of a dominant strategy and of expected utility maximization first appear here, in an almost formal model of decision making. These are two additional major breakthroughs, dividends of the connection that Pascal made between decision problems and games of chance.

Importantly, in this argument Pascal seems to consider the possibility that the probability of God's existence may not be given a precise value. The argument is that, as long as this probability is within a certain range, say $p \in (0, \varepsilon)$ for some $\varepsilon > 0$, however small, one can determine that it is better to be pious because of the infinite difference between salvation and "error and misery".[7] For, $+\infty = (+\infty - b) / (c - a) > (1 - p) / p$ for each $p \in (0, \varepsilon)$. This argument is in the spirit of the multiple priors model a la Bewley, presented in Section 3.4.

Summing up, apart from its (widely debated) theological merits, Pascal's famous wager was an incredible leap in several ways and marks the birth of decision theory. Thanks to the connection between decision problems and games of chance that he envisioned, he sketched the basic ideas of this discipline, with the formalization of decision problems through decision matrices, as well as their resolution through weak dominance and expected utility maximization, possibly with a multiple priors twist.

The main point for our discussion of Pascal's wager is that the subjective interpretation of probabilities and their application as a tool to quantify beliefs showed up on the scene more or less as soon as did the objective interpretation and the application to games of chance. Further, as soon as the notion of subjective probability came on stage, it was accompanied by the possibility that

[6] Setting $b = -\infty$ is not an obvious implication of Pascal's text and is immaterial for his argument.

[7] "You would act stupidly ... by refusing to stake one life against three at a game in which out of an infinity of chances there is one for you, if there were an infinity of an infinitely happy life to gain."

this probability might not be known (see Shafer, 1986, for related remarks on Bernoulli, 1713, who introduced the law of large numbers).

Huygens (1657) In the wake of the early probability discussions of Fermat and Pascal, Huygens (1657) first clearly proposed expected values to evaluate games of fortune.[8] Unlike Pascal's grand theological stand, Huygens only dealt with games of fortune ("cards, dice, wagers, lotteries, etc." as reported in the 1714 English version). Nevertheless, he was well aware of the intellectual depth of his subject.[9]

Huygens' arguments are a bit obscure (at least for the modern reader; see Daston, 1995). His essay has, however, a few remarkable features from our perspective. First, he does not present the expected value criterion as an axiom; rather, he justifies its relevance by starting from more basic principles. In other words, Huygens did not claim that expected value was the appropriate criterion to value games of fortune; he tried to prove it. For this reason his essay is articulated in a sequence of mathematical propositions that establish the expected value criterion for more and more complicated games.[10] Huygens' propositions can be thus viewed as the very first decision-theoretic representation theorems, in which the relevance of a decision criterion is not viewed as self-evident, but needs to be justified through logical arguments based on first principles. In so doing Huygens opened up a research path that, through Bernoulli's 1738 masterpiece, culminated in the celebrated expected utility axiomatization of von Neumann and Morgenstern (1947).

A second remarkable feature of Huygens' essay is the basic principle, his "postulat" in the English version, which he based his analysis upon. We may call it the principle of equivalent games, in which he assumes that the values of games of chances should be derived through the value of equivalent fair games. In his words "... I use the fundamental principle that a person's lot or expectation to obtain something in a game of chance should be judged to be worth as much as an amount such that, if he had it, he could arrive again at a like lot or expectation contending under fair conditions, that is, in a game which works to no one's advantage."[11] Though his definition of a fair game is ambiguous, it is a class of games that he regarded to have self-evident values, possibly because of

[8] Huygens acknowledged the influence of Fermat and Pascal in the preface of his essay "Lest anyone give me undeserved glory of first discovering this matter, it should be known that this calculus was already discussed some time ago by some of the most outstanding mathematicians of all of France." (English translation here and in the rest of the section is from the 2005 edition of Bernoulli 1713 by Edith Dudley Sylla). More on the relationships between Fermat, Huygens, and Pascal can be found in Ore (1960).

[9] In the preface of his essay Huygens warns the reader not to consider frivolous its subject matter ("Quippe cum in re levi ac frivola operam collocasse videri alioqui possem"), which is instead beautiful and very subtle, comparable in depth to Diophantus' books and even more interesting since it does not just consider mere properties of numbers ("... sed pulchrae subtilissimaeque contemplationis fundamenta explicari. Et problemata quidem quae in hoc genere proponuntur, nihilo minus profundae indaginis visum iri confido, quam que Diophanti libris contenitur, voluptatis autem aliquanto plus habitura, cum non, sicut illa, in nuda numerorum consideratione terminentur.")

[10] The essay, which is about seventeen pages long, consists of fourteen propositions and five problems.

[11] The Latin text, pp. 521-522, is "Hoc autem utrobique utar fundamento: nimirum, in alae ludo tanti estimandam esse cuiusque sortem seu expectationem pervenire, aequa conditione certans." The final sentence "that is, ... advantage" only appeared in the Dutch version as "dat is, daer in niemandt verlies geboden werdt."

suitable symmetries in their structure (cf p. 25 of Daston, 1995, and p. 95 of Hacking, 1975). The meaning of "equivalent" is also unclear. But, despite all these obscurities, the idea of postulating the existence of benchmark games in order to use them to evaluate general games has played a prominent role in decision theory, in different guises. Ramsey's assumption of the existence of bets with equally likely outcomes (that he calls "ethically neutral") is an instance of this principle, as well as de Finetti's assumption of the existence of partitions of equally likely events (see axiom D.4 below). More recently, the central role that certainty equivalents play in many axiomatic derivations can be viewed as a later instance of Huygens' comparative principle of studying uncertain alternatives by means of benchmark alternatives with suitably simple structures.

Hacking (1975) makes some further observations on the relevance of Huygens' book for the history of subjective probability. We refer the interested reader to his book, with a warning on the difficulty of interpreting some of Huygens' arguments.

2.2 Subjective probabilities and the axiomatic approach

Modern decision theory, and in particular the way it models uncertainty, is the result of the pioneering contributions of a truly impressive array of scholars. Some of the finest minds of the first half of last century contributed to the formal modeling of human behavior. Among them, especially remarkable are the works of Frank Plumpton Ramsey (1903-1930) with his early insights on the relations between utilities and subjective probabilities, John von Neumann (1901-1957) and Oskar Morgenstern (1902-1977) with their classic axiomatization of expected utility presented in the 1947 edition of their famous game theory book, Bruno de Finetti (1906-1985) with his seminal contributions to subjective probability, and Leonard J. Savage (1917-1971), who – in an unparalleled conceptual and mathematical tour de force – integrated von Neumann-Morgenstern's derivation of expected utility with de Finetti's subjective probability.

For our purposes the contributions of de Finetti, Ramsey, and Savage are especially relevant since they shaped modern Bayesian thought and, through it, the modeling of uncertainty in economics. Next we briefly review their landmark contributions.

Ramsey (1926a) A main motivation of Ramsey (1926a) was Keynes (1921)'s logical approach to probability theory, in which the degrees of beliefs in different proposition were connected by necessary/objective relations, called probability relations. Skeptical regarding the existence of such relations, Ramsey argued that degrees of belief should be viewed and studied as subjective entities. To this end, he promoted the behavioral definition of subjective probability as willingness to bet, and claimed that if subjective probabilities, so defined, do not follow standard probability calculus, the individual will make incoherent decisions. These are two central ideas in the methodology of decision theory, which in about the same years were also advocated by Bruno de Finetti.

Specifically, the first tenet of Ramsey's approach is that the only sensible way to measure degrees of beliefs is not through introspection, but by considering them as a basis of action – "the kind of measurement of beliefs with which probability is concerned ... is a measurement of belief qua basis of action", p. 171. How strongly DMs believe in the occurrence of an event is revealed by how much they are willing to act on them (by "how far we should act on these beliefs" in his words). This first key insight leads Ramsey to consider betting behavior as a way to quantify beliefs on given events through the odds that DMs are willing to accept for bets on these events ("The old-established way of measuring a person's beliefs is to propose a bet, and see what are the lowest odds which he will accept. This method I regard as fundamentally sound" p. 172).

Very similar remarks can be found in de Finetti.[12] This elicitation/measurement of subjective probabilities through betting behavior – that is, through action based on them – as opposed to mere introspection, gives the theory an operational, empirical content and relates theoretical concepts to observations. This is in line with the preaching of the logical positivist, culminating in the Received View, first stated by Rudolf Carnap in the 1920s (see Carnap, 1923, and Suppe, 1977). de Finetti explicitly adopted the doctrine of Operationalism (see, e.g., the last chapter of his 1937 article), and saw the elicitation of subjective probabilities through betting behavior as methodologically akin to Vilfredo Pareto's ordinal utility theory based on the elicitation of indifference curves rather than on some psychological entities that could not measured (when data are assumed to be only a weak order over alternatives). Ramsey was motivated by similar methodological concerns, in a Pragmatist perspective,[13] and viewed this approach as akin to what was done in the physical sciences (see, Section 3 of his article).

Ramsey understood that the measurement of subjective probabilities through betting behavior might be inexact if it does not properly take into account the utility of money, which he viewed as subject to diminishing marginal utility and to what today we call risk attitudes ("...eagerness or reluctance to bet..." in his words, p. 172). To overcome this problem he outlined a theory in which "... a person's actions are completely determined by his desires and opinions" (p. 173), thus anticipating some of the main themes of Savage's analysis.

The view that the rules of standard probability calculus correspond to consistent betting behavior is the second main tenet of Ramsey's approach. By consistent betting behavior he meant behavior that was not subject to so-called "Dutch books". That is, consistency requires that the DMs do not

[12] See, e.g., de Finetti (1931) p. 302 "Anche nel linguaggio ordinario ... si esperime il grado di fiducia che abbiamo nel verificarsi di un dato evento mediante le condizioni a cui ci si potrebbe scommettere." ("Even in everyday language ... our degree of confidence in the occurrence of an event is expressed through the conditions at which we would bet on it.")

[13] Operationalism started with Bridgman (1927), after Ramsey's articles of 1926.

accept series of bets in which they will incur a sure loss. Ramsey observed that the probabilities elicited from betting behavior satisfy the rules of probability calculus if and only if this behavior is consistent. Remarkably, subjective probabilities measured via consistent betting behavior are thus characterized exactly by these rules. As Ramsey wrote on p. 183 "Having degrees of beliefs obeying the laws of probability implies a further measure of consistency, namely such a consistency between the odds acceptable on different proposition as shall prevent a book made against you."

Like the first tenet on the measurement of beliefs via betting behavior, also this second consistency tenet was made by de Finetti in very similar terms.[14] Together, the two tenets form the basis of de Finetti and Ramsey's approach to subjective probability. Their parallel, altogether independent, development of these key ideas is impressive. We close by remarking that the consistency tenet is an early instance of the weak, internal, notion of rationality adopted by decision theory, which only requires consistency of choices without any reference to their motives.[15] This consistency approach was part of a more general intellectual movement in the 1920s, with its most famous instance being Hilbert's program on the axiomatization of mathematics.[16]

de Finetti (1931) Bruno de Finetti, one of the greatest probabilists of the twentieth century, was a key figure in the development of the Bayesian approach. To the best of our knowledge, he was the first to promote the Bayesian approach as an all-encompassing method of reasoning about uncertainty, and he did so with a religious zeal. His two main papers in this regard are probably de Finetti (1931, 1937). In both papers he forcefully emphasized the two key ideas on subjective probabilities that we just discussed in relation with Ramsey's work.[17] Besides these methodological issues, the main contribution of the 1937 article is a complete presentation of his classic extension of the law of large numbers to deal with sequences of random variables that were merely exchangeable, an extension that he had developed in a series of papers in the 1930s. His theorem was a crucial step for Bayesian statistics because it provided the subjective foundation for the parametric statistical model, the basic framework of statistical inference that is based on a set of possible probabilistic models $\{P_\theta\}_{\theta\in\Theta}$ that may govern a given stochastic process, indexed through a parameter space Θ over which there is a

[14] See, e.g., de Finetti (1931) p. 305 "[Il decisore] è costretto a ... rispettare ... i teoremi del calcolo delle probabilità. Altrimenti egli pecca di *coerenza*, e perde *sicuramente*, purchè l'avversario sappia sfruttare il suo errore." ("[The DM] must ... follow ... the theorems of probability calculus. Otherwise he will not be *coherent*, and will lose *for sure*, provided the opponent knows how to exploit his mistake.") Emphasis as in the original.

[15] See Gilboa, Postlewaite, and Schmeidler (2010) and Gilboa, Maccheroni, Marinacci, and Schmeidler (2010) for a recent discussion of this issue.

[16] See Ramsey (1926b) for a critical account of Hilbert's program. As Weyl (1944) p. 550 remarked, in this program "Admittedly the question of truth is thus shifted into the question of consistency."

[17] Both de Finetti and Ramsey developed their seminal ideas on subjective probability when they were in their early twenties. They both made exceptional contributions to several fields. This is all the more remarkable for Ramsey, who in his short life was able to make important contribution in economics, mathematics, and philosophy (in the Arrow's (1980) words, "Frank P. Ramsey must be one of the most remarkable minds of all times, though the contributions to human knowledge of which he was capable were so severely rationed by the chance of an early death").

prior probability μ (see, e.g., Kreps, 1988, Schervish, 1995, and Cifarelli and Regazzini, 1996).

For the modelling of uncertainty in economics, however, the most important contribution of de Finetti is his 1931 article in Fundamenta Mathematicae, arguably the most innovative mathematics journal in the interwar years. The central part of that paper, in particular pages 320-324, are truly remarkable. In these few pages de Finetti first introduced a binary relation $\succsim$, a *qualitative probability*, over a collection Σ of events of a state space S, where $E \succsim E'$ is interpreted as "event E is at least as likely as event E'." de Finetti viewed this relation as a primitive ("supponiamo acquisita la nozione della relazione",[18] p. 320) and did not explicitly relate it to betting behavior. The connection was made explicit later on by Savage (1954), as we will see momentarily.

de Finetti proposed a few basic properties that $\succsim$ should satisfy. Specifically:

D.1 $\succsim$ is a weak order, that is, it is complete and transitive.

D.2 $S \succsim E \succsim \emptyset$ for all events E, with $S \succ \emptyset$.

D.3 Given any two events E and E',

$$E \succsim E' \Rightarrow E \cup E'' \succsim E' \cup E''$$

for all events E'' such that $E \cap E'' = E' \cap E'' = \emptyset$.

As de Finetti emphasized, these are natural properties to require on $\succsim$, with D.3 being the main conceptual property ("la proprietà essenziale" in his words). In modern terminology, D.3 is a basic independence assumption.

The central question that de Finetti raised was whether a qualitative probability that satisfies properties D.1-D.3 can be represented by a standard (finitely additive) probability measure P on Σ, that is, whether there exists a probability P such that, for every two events E, E',

$$E \succsim E' \Leftrightarrow P(E) \geq P(E'). \tag{1}$$

In this way, the qualitative probability relation, which is very appealing from a foundational standpoint, can be given a much broader modelling scope thanks to a numerical representation P to which standard probability theory can be applied.

de Finetti was only able to prove that there exists a probability P on Σ such that

$$E \succsim E' \Rightarrow P(E) \geq P(E') \tag{2}$$

provided the following "equidivisibility" condition holds:[19]

[18] "Let us assume as known the meaning of the relation."

[19] See Kreps (1988) p. 121 and Gilboa (2009) p. 111.

D.4 For each $n \geq 1$ there is a partition $\{E_i\}_{i=1}^n$ of equally likely events.

de Finetti's question remained open for several years and was finally answered in the negative by Kraft, Pratt, and Seidenberg (1959). They provided some ingenious counter-examples of qualitative probabilities on finite state spaces that satisfy axioms D.1-D.3, but that do not have a probability measure that satisfies (1) or even the weaker (2), as well as conditions that characterize a qualitative probability that can be represented by a probability measure.

Mathematically and operationally, de Finetti's approach brings to mind Laplace's classic approach to probability. Both make a reference to a partition $\{E_i\}_{i=1}^n$ of equally-likely events, and it is easy to see that de Finetti's assumptions D.1-D.3 imply that their quantitative probability is $P(E_i) = 1/n$ and $P\left(\bigcup_{i \in I} E_i\right) = |I|/n$ for each finite index set $I \subseteq \{1, ..., n\}$. In the proof of (2) these assessments are then used to evaluate the probability of more complicated events by suitably approximating them.

Focusing on the words, "... to a certain number of cases equally possible, that is to say, to such *as we may be equally undecided* about in regard to their existence", one may interpret Laplace himself as a subjectivist.[20] This is in line with Laplace's deterministic view of the world (cf. "Laplace's demon"), which suggests that the notion of probability results from our ignorance.[21] However, this interpretation of Laplace as a precursor of de Finetti is quite different from the way that the "Principle of Indifference" or "Principle of Insufficient Reason" is often read: for a subjectivist such as de Finetti, the subjective judgment, whether justified or not, is primitive. By contrast, the "Principle of Indifference" is taken to mean that whenever one cannot justify a preference for one event over the other, their probabilities should equal. This interpretation has a much stronger objective flavor. Unfortunately, it often yields incoherent assessments.[22]

The novelty of de Finetti (1931) was both methodological and scientific. Methodologically, it is one of very first articles that adopted the axiomatic method based on a binary relation $\succsim$ and its numerical representation derived from suitable axioms on $\succsim$.[23] Scientifically, he provided the first result that axiomatized subjective probability, thereby establishing one of the two pillars which Savage's great synthesis relied upon.[24]

We close with a few words on the works of Koopman (1940a, 1940b, 1941), who is best known for the Hilbert space formulation of classical mechanics that he pioneered along with von Neumann (see,

[20] See pages 6 and 7 of the 1951 English edition. As Hacking (1975) p. 122 remarks, the idea of defining probability through equipossibility was known since the early days of Probability Theory (for example, Hacking mentions a 1678 essay of Leibniz that relied on that idea).

[21] This does not mean that objective probabilities did not exist for him. Perhaps he just did not need this hypothesis...

[22] This was already pointed out by Bertrand (1907). See Gilboa (2009) and the end of Section 2.3 below.

[23] Frisch (1926) was the first article we are aware of that adopted a similar approach in economic theory.

[24] See, e.g., chapter 8 of Kreps, 1988.

e.g., Koopman and von Neumann, 1932). In these works, Koopman studied qualitative probabilities in a logical perspective a la Keynes, not as a basis for action as in de Finetti and Ramsey.[25] His logical approach was based on conditional probability comparisons $a/h \precsim b/k$ involving logical propositions a , b, h and k, which he read as "a on the presumption h is no more probable than b on the presumption k". Koopman interpreted these propositions as "empirical propositions", that is, scientific assertions (e.g., outcomes of physical or biological experiments) whose truth value can be verified through experimental observation, as prescribed by Bridgam's Operationalism.

Consistency conditions among conditional comparisons $a/h \precsim b/k$ play a central role in Koopman's logic approach. Through them he is able to derive numerical representations of the conditional comparative relation $\precsim$. Though Koopman was not interested in the decision theoretic side of probability, his analysis of qualitative probability has some formal overlaps with the works of de Finetti and Ramsey.

Savage (1954) de Finetti's derivation of subjective probability was conceptually complementary with von Neumann and Morgenstern's (vNM, 1947) derivation of expected utility maximization under risk, which assumed known numerical probability measures. The integration of de Finetti's subjective probability with vNM's expected utility was achieved by Savage's (1954) book, which derived subjective expected utility maximization when neither probabilities nor utilities were given.

Inspired by Wald (1950), in his book Savage introduced what is now the standard model of a Bayesian decision problem. There is an abstract space S of states of the world and an abstract set X of outcomes. Subsets of S are the events and Savage considered all them, that is, the power set 2^S. Preferences $\succsim$ are over acts, which are maps $f : S \to X$ from states to outcomes. For simplicity, we consider $\succsim$ on the collection $\mathcal{F}$ of simple acts, that is, acts that have a finite number of outcomes.

Savage considered six axioms on $\succsim$, which he called P1-P6. The first one, P.1, requires that $\succsim$ be transitive and complete. The other ones are more subtle and are discussed in detail in Gilboa (2009) pp. 97-105. For later reference here we briefly recall P.2, the celebrated Sure-Thing Principle,[26] which is especially important for our purposes. Given any pair of acts f and g, and any event E, let us write

$$fEg = \begin{cases} f(s) & \text{if } s \in E \\ g(s) & \text{if } s \notin E \end{cases}$$

In other words, fEg is the act equal to act f on E and to act g otherwise. Using this notation we can introduce the Sure-Thing Principle.

[25] As he writes at the beginning of his 1941 article, his purpose was "a general study probability regarded as a branch of intuitive logic".

[26] Though nowadays P.2 is commonly referred with this name, this was not Savage's original terminology (see Gilboa, 2009, p. 99).

P.2 SURE-THING PRINCIPLE: given any acts $f, g, h, h' \in \mathcal{F}$ and any event E, we have

$$fEh \succsim gEh \Longleftrightarrow fEh' \succsim gEh'.$$

The two comparisons in P.2 involve acts that only differ in their common parts h and h'. The Sure-Thing Principle requires that if act fEh is preferred to act gEh, no reversal in preference can occur if their common part h is replaced by a different, but still common, part h'. In other words, rankings of acts should be independent of common parts.

In his classic representation theorem, Savage proved that axioms P.1-P.6 are equivalent to the existence of a utility function $u : X \to \mathbb{R}$ and a (nonatomic[27]) probability measure P on the power set 2^S such that acts $f \in \mathcal{F}$ are ranked according to their subjective expected utility $\int_S u(f(s))\, dP(s)$.[28] Relative to the vNM derivation, the novelty was the subjective probability P, whose derivation was based on de Finetti (1931). However, the fact that Savage combined ideas of de Finetti and of vNM may be misleading: de Finetti used bets on numerical amounts, interpreted as measuring money, to derive probability. vNM used probabilities to derive a utility. But having neither numerical concept as primitive, Savage was coping with a problem that was conceptually and mathematically much more challenging.[29]

To see the gist of Savage's approach, observe that bets on events are binary acts xEy that pay x if event E obtains and y otherwise. If $x \succ y$,[30] then xEy can be viewed as a bet on event E. Through bets, Savage defines a qualitative probability $\succsim^*$ on 2^S by setting

$$E \succsim^* E' \Leftrightarrow xEy \succsim xE'y \tag{3}$$

when $x \succ y$. In the wake of de Finetti and Ramsey, Savage thus also assumes that DMs reveal their likelihood assessments over events through betting behavior: we can say that according to a DM event E is at least as likely as event E' if she finds a bet on E at least as attractive as a bet on E'.

Savage showed that P1-P6 imply that $\succsim^*$ satisfies de Finetti's axioms D.1-D.3, as well as a version of de Finetti's Laplacian condition D.4. As a result, Savage was able to prove that there is a nonatomic probability measure P that represents $\succsim^*$ as in (1).

[27] A probability P is nonatomic if, for each $P(A) > 0$ and each $\alpha \in (0,1)$, there is $B \subseteq A$ such that $P(B) = \alpha P(A)$. This property reflects equidivisibility assumptions a la D.4.

[28] Here and in the sequel we follow the tradition of de Finetti and Savage, and do not assume that P is necessarily countably additive. See Theorem 6 and its discussion for more on this issue.

[29] Savage's investigation apparently started with conversations with Herman Chernoff, who showed him that his 1951 minimax regret criterion did not satisfy Arrow's principle of irrelevant alternatives. As Chernoff says in Bather (1996) "I brought this to Savage's attention and, after arguing futilely for a little while that it proved how good his criterion was, he finally agreed that it was wrong. He felt then that perhaps we should be elaborating on de Finetti's Bayesian approach, which he had come across. (He was a voracious reader.)"

[30] An outcome x is indentified with the constant act, still denoted by x, that delivers x in all states. This naturally suggests a defintion of $\succsim$ on X, and makes expressions such as $x \succ y$ well defined.

The derivation of P completes the first part of Savage's derivation, based on de Finetti's subjective probability theory, the first pillar which he was building on. The second part of his derivation builds on the other pillar, von Neumann-Morgenstern's expected utility theory. This theory is based on a preference over the lotteries on outcomes, that is, over the collection $\Delta(X)$ of all probability measures $p : 2^X \to [0,1]$ on outcomes that are simple.[31] Through the subjective probability P, each act $f \in \mathcal{F}$ induces a probability measure $p_f \in \Delta(X)$ given by:

$$p_f(A) = P(s \in S : f(s) \in A), \qquad \forall A \subseteq X.$$

Thanks to the nonatomicity of P, Savage was able to prove that, given any $p \in \Delta(X)$, there is a suitable $f \in \mathcal{F}$ such that $p = p_f$. Therefore, the primitive preference $\succsim$ induces on $\Delta(X)$ a derived preference $\succsim^{**}$ given by

$$p_f \succsim^{**} p_g \Leftrightarrow f \succsim g.$$

Using P, the preference $\succsim^{**}$ reduces a choice problem under uncertainty to a choice problem under risk a la von Neumann-Morgenstern. In particular, Savage showed that P1-P6 imply that $\succsim^{**}$ is well defined[32] and satisfies von Neumann-Morgenstern's axioms. As a result, their classic representation theorem implies the existence of a utility function $u : X \to \mathbb{R}$ such that the lotteries p_f are ranked according to their expected utility $\sum_{x \in \operatorname{supp} p_f} u(x) p_f(x)$.

We can now put all pieces together. For, given any two simple acts f and g, we can write:

$$\begin{aligned} f &\succsim g \Leftrightarrow p_f \succsim^{**} p_g \Leftrightarrow \sum_{x \in \operatorname{supp} p_f} u(x) p_f(x) \geq \sum_{x \in \operatorname{supp} p_g} u(x) p_g(x) \\ &\Leftrightarrow \int_S u(f(s))\, dP(s) \geq \int_S u(g(s))\, dP(s), \end{aligned}$$

where the last equivalence follows from a simple change of variable.

Acts f and g are thus ranked through their subjective expected utilities, and this completes Savage's derivation.[33] From both conceptual and mathematical viewpoints, Savage's theorem remains an unparalleled gem in the literature. To this day, it is universally viewed as the most compelling reason to assume that rational choice necessitates Bayesian quantification of all uncertainty, that is, the reduction of uncertainty to risk.

2.3 Ellsberg paradox

The classic Bayesian theory culminating in Savage's opus represents beliefs probabilistically, but it does not capture the degree of confidence that DMs have in their own probabilistic assessments, a

[31] A probability measure $p : 2^X \to [0,1]$ is *simple* if it has a finite support, i.e., if there is a finite set, written $\operatorname{supp} p$, such that $p(\operatorname{supp} p) = 1$.

[32] In this regard a key step in Savage's derivation is to show that from his axioms it follows that $p_f = p_g$ implies $f \sim g$ (see, e.g., Theorem 14.3 of Fishburn, 1970).

[33] The Savage derivation is spelled out we refer the reader to Fishburn (1970), Kreps (1988), and Gilboa (2009). Recent extensions of Savage's result can be found in Kopylov (2007, 2010).

degree that depends on the quality of the information that DMs use in forming these assessments. The classic theory focused on how to measure beliefs, without providing a way to assess the quality of such measurements.

Savage himself was well aware of this issue. For example, on p. 57 of his book he writes "there seem to be some probability relations about which we feel relatively 'sure' as compared with others." But, as he remarked soon after "The notion of 'sure' and 'unsure' introduced here is vague, and my complaint is precisely that neither the theory of personal probability, as it is developed in this book, nor any other device known to me renders this notion less vague."

Ellsberg (1961) provided two stark thought experiments that showed how this limitation may lead many people to violate Savage's otherwise extremely compelling axioms, and to express preferences that are incompatible with any (single, additive) probability measure. Ellsberg argued that a situation in which probabilities are not known, which he referred to as *ambiguity*,[34] induces different decisions than situations of risk, namely, uncertainty with known probabilities.

Ellsberg's two urns Specifically, one of Ellsberg's experiments involves two urns, I and II, with 100 balls in each. The DM is told that:

(i) in both urns balls are either white or black;

(ii) in urn I there are 50 black and 50 white balls.

No information is given on the proportion of white and black balls in urn II.

The DM has to choose among the following 1 euro bets on the colors of a ball drawn from each urn:

1. bets IB and IW, which pay 1 euro if the ball drawn from urn I is black and white, respectively;
2. bets IIB and IIW, which pay 1 euro if the ball drawn from urn II is black and white, respectively.

Let us model the DM's choice among these bets in a Savage framework. The state space is $S = \{B, W\} \times \{B, W\}$, with the following four possible states:

1. state BB: in both urns a black ball is drawn;
2. state BW: in urn I a black ball is drawn and in urn II a white ball is drawn;
3. state WB: in urn I a white ball is drawn and in urn II a black ball is drawn;

[34] Today, the terms "ambiguity", "uncertainty" (as opposed to "risk"), and "Knightian uncertainty" are used interchangeably to describe the case of unknown probabilities.

4. state WW: in both urns a white ball is drawn.

The next table summarizes the DM's choice problem.

	BB	BW	WB	WW
IB	1	1	0	0
IW	0	0	1	1
IIB	1	0	1	0
IIW	0	1	0	1

Suppose that the DM ranks these bets according to the subjective expected utility criterion. First of all, normalize his utility function by setting $u(1) = 1$ and $u(0) = 0$. As to his subjective probability, in view of the information that the DM was given about urn I, it is natural to set

$$P(BB \cup BW) = P(WW \cup WB) = \frac{1}{2}. \tag{4}$$

On the other hand, due to symmetry it seems natural also to set

$$P(BB \cup WB) = P(BW \cup WW), \tag{5}$$

because the DM has no information on the colors' proportion in urn II and there is no reason to consider the draw of either color as more likely. But the symmetric evaluation (5) when combined with the additivity of P leads to:

$$P(BB \cup WB) = P(BW \cup WW) = \frac{1}{2}. \tag{6}$$

In other words, in both urns we end up with the same probabilities, despite the stark difference in the quality of the information on which the subjective probability assessments have been based. It is natural to think that the DM's degree of confidence in his probability assessment (4) is much higher than that in (6).

From (4) and (6) it follows that all four acts have the same subjective expected utility, equal to 1/2. The following indifference pattern

$$IB \sim IW \sim IIB \sim IIW$$

thus characterizes a subjective expected utility DM with subjective probability (4) and (6). But it is not obvious that all rational DMs will be indifferent among betting on urn I, on which they have solid information, and betting on urn II, on which they have much more limited information. The following ranking is introspectively much more plausible

$$IB \sim IW \succ IIB \sim IIW. \tag{7}$$

An overwhelming experimental evidence has confirmed that many subjects have this preference pattern in this choice problem. This pattern is, however, not compatible with Savage's axioms, in particular with his classic Sure-Thing Principle (see, e.g., Gilboa, 2009, p. 134). Hence, no subjective expected utility DM can exhibit the preference pattern (7).

Ellsberg's single urn Why the Sure-Thing Principle may be less compelling than one might think *prima facie* is, however, best illustrated by the second Ellsberg experiment. Suppose there is a single urn, with 90 balls. The DM is told that:

(i) in the urn balls are either red, yellow, or green;

(ii) there are 30 red balls.

No information is given on the proportion of yellow and green balls in the 60 balls that are not red.

Suppose that the DM has to choose among the following 1 euro bets on the colors of a ball drawn from the urn:

1. bets 1_R and 1_Y, which pay 1 euro if the ball drawn from the urn is red and yellow, respectively;
2. bet $1_{R\cup G}$, which pays 1 euro if the ball drawn from the urn is either red or green;
3. bet $1_{Y\cup G}$, which pays 1 euro if the ball drawn from the urn is either yellow or green.

The state space is $S = \{R, Y, G\}$, with the following three possible states

1. R: a red ball is drawn;
2. Y: a yellow ball is drawn;
3. G: a green ball is drawn.

The next table summarizes the DM's choice problem.

	R	Y	G
1_R	1	0	0
1_Y	0	1	0
$1_{R\cup G}$	1	0	1
$1_{Y\cup G}$	0	1	1

The DM has much better information on the event R and its complement $Y \cup G$ than on the events Y and G and their complements. As a result, it seems reasonable to expect that a DM would regard 1_R as a "safer" bet than 1_Y and, therefore, she would prefer to bet on R rather than on Y; that is,

$$1_R \succ 1_Y.$$

For the very same reason, when comparing bets on $R \cup G$ and on $Y \cup G$ it seems reasonable to expect that a DM would prefer to bet on the latter event because of the much better information that she has about it. That is,

$$1_{Y \cup G} \succ 1_{R \cup G}.$$

Summing up, the quality of the information on which the DM's beliefs are based should lead to the following preference pattern

$$1_R \succ 1_Y \quad \text{and} \quad 1_{Y \cup G} \succ 1_{R \cup G} \tag{8}$$

that is, that the DMs consistently prefers to bet on events on which they have superior information. Pattern (8), which has been indeed confirmed in a number of actual experiments that carried out Ellsberg's thought experiment, is not compatible with the Sure-Thing Principle. In fact, set $A = \{R, Y\}$. Bets 1_R and 1_Y are identical on A^c. According to the Sure-Thing Principle, changing their common value in A^c from 0 to 1 should not alter their ranking. But, these modified acts are the bets $1_{R \cup G}$ and $1_{Y \cup G}$, respectively. Hence, by Sure-Thing Principle

$$1_R \succsim 1_Y \Longleftrightarrow 1_{Y \cup G} \succsim 1_{R \cup G},$$

which is violated by pattern (8).

Ambiguity aversion The phenomenon illustrated by the patterns (7) and (8) is known as *uncertainty aversion*, or *ambiguity aversion*: people tend to prefer situations with known probabilities to unknown ones, to the extent that these can be compared. Clearly, one can have the opposite phenomenon, of uncertainty/ambiguity liking, when people exhibit the opposite preferences. While gambling is an important exception, it is commonly assumed that people who are not uncertainty neutral tend to be uncertainty averse, in a way that parallels the common assumptions about attitudes toward risk.

Ellsberg's experiments are extremely elegant and they pinpoint precisely which of Savage's axioms is violated by DMs who are not indifferent between betting on the two urns. But the elegance of these experiments is also misleading. Since they deal with balls and urns, and the information about the colors is completely symmetric, it is very tempting to adopt a probabilistic belief that would reflect this symmetry. Specifically, one may reason about the urn with unknown composition, "The number of red balls in it can be any number between 0 and 100. My information is completely symmetric, and there is no reason to believe that there are more red balls than black balls or vice versa. Hence, if I were to adopt a prior probability over the composition of the urn, from [0:100] to [100:0], I should choose a symmetric prior. That is, the probability that there are 3 red balls should be equal to the probability that there are 97 red balls, and so forth. In this case, the probability that a red ball is drawn out of the urn is precisely 50%, and I should no longer express preferences for the known probabilities." Relatedly, one may also use the unknown urn to generate a bet with objective probability of 50%: use an external chance device, which is known to be fair, and decide between betting on red or on black based on this device. If the DM has symmetric beliefs about the composition of the urn, she can thereby generate a bet that is equivalent to the bet on the urn with the known composition.

Based on such arguments, theorists often feel that there is no problem with subjective probabilities, at least as far as normative theories of choice are concerned. But this conclusion is wrong. In most real life examples there are no symmetries that allow the generation of risky bets. For example, suppose that Mary does not know what is the probability of the globe warming up by 4 degrees within the next ten years. She cannot assume that this probability is 50%, based on Laplace's Principle of Indifference (or "Principle of Insufficient Reason"). The two eventualities, "average temperature increases by 4 degrees or more" and "average temperature does not increase by 4 degrees" are not symmetric. Moreover, if Mary replaces 4 degrees by 5 degrees, she will obtain two similar events, but she cannot generally assign a 50%-50% probability to any pair of complementary events. Nor will a uniform distribution over the temperature scale be a rational method of assigning probabilities.[35] The fundamental difficulty is that in most real life problems there is too much information to apply the Principle of Indifference, yet too little information to single out a unique probability measure.[36] Global warming and stock market crashes, wars and elections, business ventures and career paths face us with uncertainty that is neither readily quantified nor easily dismissed by symmetry considerations.

2.4 Other disciplines

The Bayesian approach has proved useful in statistics, machine learning, philosophy of science, and other fields. In none of these fellow disciplines has it achieved the status of orthodoxy that it enjoys within economic theory. It is a respectable approach, providing fundamental insights and relishing conceptual coherence. It is worth pointing out, however, that in these disciplines the Bayesian approach is one among many. More importantly, in all of these disciplines the Bayesian approach is applied to a restricted state space, such as a space of parameters, whereas in economics it is often expected to apply also to a *grand state space*, whose elements describe anything that can possibly be

[35] Bertrand's (1907) early critique of the principle of indifference was made in the context of a continuous space. See also Gilboa (2009) and Gilboa, Postlewaite, and Schmeidler (2009).

[36] It is not entirely clear how one can justify the Principle of Indifference even in cases of ignorance. For example, Kass and Wasserman (1996) p. 1347 discuss the partition paradox and lack of parametric invariance, two closely related issues that arise with Laplace's Principle. Similar remarks from a Macroeconomics perspective can be found in Kocherlakota (2007) p. 357.

Based on a result by Henri Poincaré, Machina (2004) suggests a justification of the Laplace's Principle using a sequence of fine partitions of the state stace. This type of reasoning seems to underlie most convincing examples of random devices, such as tossing coins, spinning roulette wheels, and the like. It is tempting to suggest that this is the only compelling justification of the Principle of Indifference, and that this principle should not be invoked unless such a justification exists.

of interest.

Consider statistics first. The statistical inference problem is defined by a set of distributions, or data generating processes, out of which a subset of distributions has to be chosen. In parametric problems, the set of distributions is assumed to be known up to the specification of finitely many parameters. Classical statistics does not allow the specification of prior beliefs over these parameters. By contrast, Bayesian statistics demands that such beliefs be specified. Thus the Bayesian approach offers a richer language, within which the statistician can represent prior knowledge and intuition. Further, the Bayesian prior, updated to a posterior based on sampling, behaves in a much more coherent way than the techniques of classical statistics. (See, for example, Welch, 1939, also described in DeGroot, 1975, pp. 400-401.)

The main disadvantage of the Bayesian approach to statistics is its subjectivity: since the prior beliefs of the parameters is up to the statistician to choose, they will differ from one statistician to another. Admittedly, classical statistics cannot claim to be fully objective either, because the very formulation of the problem as well as the choice of statistics, tests, and significance levels leave room for the statistician's discretion. Yet, these are typically considered necessary evils, with objectivity remaining an accepted goal, whereas the Bayesian approach embraces subjective inputs unabashedly.[37] On the bright side, if a Bayesian statistician selects a sufficiently "diffused" or "uninformative" prior, she hopes not to rule out the true parameters a priori, and thereby to allow learning of objective truths in the long run, despite the initial reliance on subjective judgments.[38]

The Bayesian approach has a similar status in the related fields of computer science and machine learning.[39] On the one hand, it appears to be the most conceptually coherent model of inference. On the other, its conclusions depend on a priori biases. For example, the analysis of algorithms' complexity is typically conducted based on their worst case. The Bayesian alternative is often dismissed because of its dependence on the assumptions about the underlying distribution.

It is important to emphasize that in statistics and in computer science the state space, which is the subject of prior and posterior beliefs, tends to be a restricted space that does not grow with the data. For example, it can comprise of all combinations of values of finitely many parameters, which are held fixed throughout the sampling procedure. By contrast, the standard approach in economic theory suggests that the state of the world resolves all uncertainty, and thus describes everything that might be of relevance to the problem at hand, from the beginning of time until eternity. As a result, the state space that is often assumed in economics is much larger than in other disciplines.

[37] See Lewis (1980) and chapter 4 of van Frassen (1989) (and the references therein) for a discussion of the relations between "objectivity" and subjective probabilities from a philosophical standpoint.

[38] Kass and Wasserman (1996), Bayarri and Berger (2004), and Berger (2004) discuss uninformative priors and related "objective" issues in Bayesian statistics (according to Efron, 1986, some of these issues explain the relatively limited use of Bayesian methods in applied statistics).

[39] See Pearl (1986) and the ensuing literature on Bayesian networks.

Importantly, it increases with the size of the data.

Consider the following example. The year is 1900 and we are interested in forecasting wars over the next 200 years. For simplicity, assume that every year we either observe war (1) or no-war (0). Suppose that we believe that wars occur randomly, so that every year there might be a war with probability p, where consecutive occurrences are i.i.d. We do not know what the parameter p is. A Bayesian approach to the statistical problem would be to form a certain distribution over the values of p, say, $[0, 1]$, and to update it as new information comes in. If our basic assumption about the process is correct, that is, if we do observe i.i.d Bernoulli random variables, our posterior probability will converge to the true value of p with very high probability.

But the assumption that consecutive wars are independent is dubious. Some wars are horrific enough to make people want to avoid future wars; and some wars simply drain resources, thereby making future wars less likely; by contrast, some wars end in defeat or humiliation that induce revenge in future wars. Thus, the assumption that the occurrence of wars is an i.i.d process is highly doubtful. Moreover, it seems fair to say that we do not know the data generating process by which wars erupt. In such a case, the standard way to model the problem in economics is to define the grand state space, which is not the space of the parameter, $p \in [0, 1]$, but all the 2^{200} sequences of 0's and 1's of length 200. This state space is detailed enough to describe any eventuality, and any theory about the world can be stated as an event in this model, namely, as the subset of states that conform to the theory. However, assuming that we have a prior probability over the 2^{200} states is quite different from assuming that we have such a probability over a fixed parameter space such as $[0, 1]$. In particular, if the forecast horizon increases from 200 to 300, the parameter space $[0, 1]$ is fixed, whereas the grand state space grows from 2^{200} to 2^{300} states. Therefore, it makes sense to assume that Bayesian updating will converge to assign a high conditional probability to the true state in the first case, but it is less clear how learning could take place in the grand state space.

Relatedly, if one considers a restricted set of parameters, one may argue that the prior probability over this set is derived from past observations of similar problems, each with its own parameters, taken out of the same set. But when the grand state space is considered, and all past repetitions of the problem are already included in the description of each state, the prior probability should be specified on a rather large state space before any data were observed. With no observations at all, and a very large state space, the selection of a prior probability seems highly arbitrary.

In applications of the Bayesian approach in statistics, computer science, and machine learning, it is typically assumed that the basic structure of the process is known, and only a bounded number of parameters need to be learnt. Many non-parametric methods allow an infinitely dimensional parameter space, but one that does not grow with the number of observations. This approach is sufficient for many statistical inference and learning problems in which independent repetitions are allowed. But economics is often interested in events that do not repeat. Applying the Bayesian approach to these is harder to justify.

We are not fully aware of the origins of the application of the Bayesian approach to the grand state space. It is well known that de Finetti was a devout Bayesian. Savage, who followed his footsteps, was apparently much less religious in his Bayesian beliefs. Yet, he argued that a state of the world should "resolve all uncertainty" and, with a healthy degree of self-criticism, urged the reader to imagine that she had but one decision to be taken in her lifetime, and this is her choice of her strategy before being born. Harsanyi (1967, 1968) made a fundamental contribution to economics by showing how players' types should be viewed as part of the state of the world, and assumed that all unborn players start with a common prior over the grand state space that is thus generated. Aumann (1974, 1976, 1987) pushed this line further by assuming that all acts and all beliefs are fully specified in each and every state, while retaining the assumption that all players have a prior, and moreover, the same prior over the resulting state space.

Somewhere along recent history, with path-breaking contributions by de Finetti, Savage, Harsanyi, and Aumann, economic theory found itself with a state space that is much larger than anything that statisticians or computer scientists have in mind when they generate a prior probability. Surprisingly, the economic theory approach is even more idealized that the Bayesian approach in the philosophy of science. There is nothing wrong in formulating the grand state space as a canonical model within which claims can be embedded. But the assumption that one can have a prior probability over this space, or that this is the only rational way to think about it is questionable.

2.5 Summary

Since the mid-20th century economic theory has adopted a rather unique commitment to the Bayesian approach. By and large, the Bayesian approach is assumed to be the only rational way to describe knowledge and beliefs, and this holds irrespective of the state space under consideration. Importantly, economic theory clings to Bayesianism also when dealing with problems of unique nature, where nothing is known about the structure of the data generating process. Research in recent decades plainly shows that the Bayesian approach can be extremely fruitful even when applied to such unique problems. But it is also possible that the commitment to the Bayesian approach beclouds interesting findings and new insights.

The preceding discussion highlights our view that there is nothing irrational about violating the Bayesian doctrine in certain problems. As opposed to models of bounded rationality, psychological biases, or behavioral economics, the focus of this survey are models in which DMs may sometimes admit that they do not know what the probabilities they face are. Being able to admit ignorance is not a mistake. It is, we claim, more rational than to pretend that one knows what cannot be known.

Bounded rationality and behavioral economics models often focus on descriptive interpretations. At times, they would take a conditionally-normative approach, asking normative questions given certain constraints on the rationality of some individuals. Such models are important and useful. However, the models discussed here are different in that they are fully compatible with normative interpretations. When central bank executives consider monetary policies, and when leaders of a country make decisions about military actions, they will not make a mistake if they do not form Bayesian probabilities. On the contrary, they will be well advised to take into account those uncertainties that cannot be quantified.

3 Alternative models

3.1 The Anscombe-Aumann setup

Anscombe and Aumann (1963) developed a version of the subjective expected utility model of Savage that turned out to be especially well suited for subsequent extensions of the basic Bayesian decision model. For this reason, in this sub-section we present this important setup.

The basic feature of the Anscombe-Aumann (AA, for short) model is that acts map states into lotteries, that is, acts' consequences involve exogenous probabilities a la von Neumann-Morgenstern. This feature is important both conceptually and mathematically. We now turn to introduce the setting formally, in the version presented by Fishburn (1970).

The set of simple probabilities $\Delta(X)$ on some underlying space X of alternatives is the space of consequences considered by the AA model.[40] There is a space of states of the world S endowed with an event algebra Σ. The objects of choice are acts, which map states into lotteries. We denote by $\mathcal{F}$ the collection of all simple acts $f : S \to \Delta(X)$, that is, acts that are finitely valued and Σ-measurable.[41]

A key feature of $\Delta(X)$ is its convexity, which makes it possible to combine acts. Specifically, given any $\alpha \in [0, 1]$, set

$$(\alpha f + (1-\alpha) g)(s) = \alpha f(s) + (1-\alpha) g(s), \qquad \forall s \in S. \tag{9}$$

The mixed act $\alpha f + (1-\alpha) g$ delivers in each state s the compound lottery $\alpha f(s) + (1-\alpha) g(s)$. In other words, *ex post*, after the realization of state s, the DM obtains a risky outcome governed by the lottery $\alpha f(s) + (1-\alpha) g(s)$.[42]

The possibility of mixing acts is a key dividend of the assumption that $\Delta(X)$ is the consequence space, which gives the AA setting a vector structure that the Savage setting did not have. The

[40] Throughout the section we use interchangeably the terms lotteries and simple probabilities.

[41] Simple acts have the form $f = \sum_{i=1}^{n} p_i 1_{E_i}$, where $\{E_i\}_{i=1}^{n} \subseteq \Sigma$ is a partition of S and $\{p_i\}_{i=1}^{n} \subseteq \Delta(X)$ is a collection of lotteries.

[42] For this reason, mixing acts in this way is sometimes called "ex post randomization." For recent models with ex ante randomization, see Epstein, Marinacci, and Seo (2007), Ergin and Sarver (2009), and Seo (2009).

derivation of the subjective expected utility representation in the AA setting is based on this vector structure.

Risk preference The DM has a primitive preference $\succsim$ on $\mathcal{F}$. In turn, this preference induces a preference $\succsim_\Delta$ on lotteries by setting, for all $p, q \in \Delta(X)$,

$$p \succsim_\Delta q \Leftrightarrow f \succsim g,$$

where f and g are the constant acts such that $f(s) = p$ and $g(s) = q$ for all $s \in S$.

Constant acts are not affected by state uncertainty, only by the risk due to the lotteries' exogenous probabilities. For this reason, $\succsim_\Delta$ can be seen as the risk preference of the DM. This is an important conceptual implication of having $\Delta(X)$ as the consequence space. This richer consequence space mathematically delivers a most useful vector structure, while from a decision theoretic standpoint it enriches the setting with a risk preference that allows to consider the DMs' risk behavior separately. Differently put, the AA consequence space can be viewed as derived from an underlying consequence space X a la Savage, enriched by a lottery structure that allows to calibrate risk preferences.

Alternatively, one may view AA's model as an improved version of de Finetti's (1931, 1937) axiomatic derivation of expected value maximization with subjective probabilities. de Finetti assumed additivity or linearity in payoffs. This is a problematic assumption if payoffs are monetary, but it is more palatable if payoffs are probabilities of receiving a fixed desirable outcome. Replacing the payoffs in de Finetti's model by probabilities of outcomes, one obtains a model akin to AA's.

In a sense, the AA model is a hybrid between vNM's and Savage's. Mathematically it is akin to the former, as it starts with a vNM theorem on a particular mixture space, and imposes additional axioms to derive subjective probabilities. Conceptually, it is closer to Savage's model, as it derives probabilities from preferences. Many view this derivation as conceptually less satisfactory than Savage's, because the latter does not assume probabilities, or any numbers for that matter, to be part of the data. Anscombe and Aumann, however, viewed the use of objective probabilities as a merit, because they believed that people think in terms of subjective probabilities after they have internalized the concept of objective probability. Be that as it may, there is no doubt that the AA model has become the main testbed for new models of decision under uncertainty.[43]

Axioms We now make a few assumptions on the primitive preference $\succsim$. The first one is a standard weak order axiom.

AA.1 WEAK ORDER: $\succsim$ on $\mathcal{F}$ is complete and transitive.

[43] See Ghirardato, Maccheroni, Marinacci, and Siniscalchi (2003) for a subjective underpinning of the AA setup.

The next axiom is a monotonicity assumption: if state by state an act f delivers a weakly better (risky) consequence than an act g, then f should be weakly preferred to g. It is a basic rationality axiom.

AA.2 MONOTONICITY: for any $f, g \in \mathcal{F}$, if $f(s) \succsim_{\Delta} g(s)$ for each $s \in S$, then $f \succsim g$.

Next we have an independence axiom, which is peculiar to the AA setting since it relies on its vector structure.

AA.3 INDEPENDENCE: for any three acts $f, g, h \in \mathcal{F}$ and any $0 < \alpha < 1$, we have

$$f \succ g \Rightarrow \alpha f + (1-\alpha)h \succ \alpha g + (1-\alpha)h. \tag{10}$$

According to this axiom, the DM's preference over two acts f and g is not affected by mixing them with a common act h. In the special case when all these acts are constant, axiom AA.3 reduces to von Neumann-Morgenstern's original independence axiom on lotteries.

We close with standard Archimedean and nontriviality assumptions.[44]

AA.4 ARCHIMEDEAN: let f, g, and h be any three acts in $\mathcal{F}$ such that $f \succ g \succ h$. Then, there are $\alpha, \beta \in (0,1)$ such that $\alpha f + (1-\alpha)h \succ g \succ \beta f + (1-\beta)h$.

AA.5 NONDEGENERACY: there are $f, g \in \mathcal{F}$ such that $f \succ g$.

We can now state the AA subjective expected utility theorem.

Theorem 1 *Let $\succsim$ be a preference defined on $\mathcal{F}$. The following conditions are equivalent:*

(i) $\succsim$ satisfies axioms AA.1-AA.5;

(ii) there exists a non-constant function $u : X \to \mathbb{R}$ and a probability measure $P : \Sigma \to [0,1]$ such that, for all $f, g \in \mathcal{F}$, $f \succsim g$ if and only if

$$\int_S \left(\sum_{x \in \operatorname{supp} f(s)} u(x) f(s) \right) dP(s) \geq \int_S \left(\sum_{x \in \operatorname{supp} g(s)} u(x) g(s) \right) dP(s). \tag{11}$$

Moreover, P is unique and u is cardinally unique.[45]

The preference functional $V : \mathcal{F} \to \mathbb{R}$ in (11) has the form

$$V(f) = \int_S \left(\sum_{x \in \operatorname{supp} f(s)} u(x) f(s) \right) dP(s) \tag{12}$$

[44] See Gilboa (2009) for some more details on them.

[45] Throughout the paper, cardinally unique means unique up to positive affine transformations.

and consists of two parts. The inner part

$$\sum_{x \in \operatorname{supp} f(s)} u(x) f(s) \tag{13}$$

is the expected utility of the lottery $f(s)$ that act f delivers when state s obtains. It is easy to see that this expected utility represents the DM's risk preference $\succsim_{\Delta}$. The outer part

$$\int_S \left(\sum_{x \in \operatorname{supp} f(s)} u(x) f(s) \right) dP(s)$$

averages all expected utilities (13) according to the probability P, which quantifies the DM's beliefs over the state space.

The classical models of Savage and Anscombe-Aumann were considered the gold standard of decision under uncertainty, despite the challenge posed by Ellsberg's experiments. In the 1980s, however, several alternatives were proposed, most notably models based on probabilities that are not necessarily additive, or on sets of probabilities. We now turn to review these contributions and some of the current research in the area.

3.2 Choquet expected utility

The first general-purpose, axiomatically-based non-Bayesian decision model was the Choquet Expected Utility (CEU) model proposed by David Schmeidler in 1982, which appeared as Schmeidler (1989). Schmeidler's starting point was that the Bayesian model is a straightjacket that does not allow the DM to express her own degree of confidence in her beliefs. Schmeidler gave the example of two coins, one that has been tested extensively and is known to be fair, and the other about which nothing is known. He noted that a Bayesian would probably have 50%-50% beliefs regarding the result of the toss of either coin, but that these beliefs differ: in one case, the DM practically knows that each side of the coin has probability of 50% of coming up. In the other case, the numbers 50%-50% are obtained with a shrug of one's shoulders, relying on symmetry of ignorance rather than symmetry of information.[46] Observe that Schmeidler's two-coin example is very close to Ellsberg's two-urn experiment. However, Schmeidler was not motivated by the desire to explain Ellsberg's results; rather, he considered the standard theory and found it counter-intuitive.

Schmeidler (1989) suggested to model probabilities by set functions that are not necessarily additive. For example, if H (T) designates the event "the unknown coin falls with H (T) up", and ν is the measure of credence, we may have

$$\nu(H) + \nu(T) < \nu(H \cup T)$$

[46] See Fischhoff and Bruine De Bruin (1999) for experimental evidence on how people use 50%-50% statements in this sense.

Thus, the "probability" of events, as measured by our willingness to bet on them, may not satisfy the standard axioms of probability theory. Schmeidler referred to them as *non-additive probabilities*, and required that they be positive and monotone with respect to set inclusion. Such mathematical entities are also known by the term *capacities*. Formally, given an event algebra Σ of state space S, a set function $\nu : \Sigma \to [0,1]$ is a capacity if

(i) $\nu(\emptyset) = 0$ and $\nu(S) = 1$;

(ii) $E \subseteq E'$ implies $\nu(E) \leq \nu(E')$.

Dempster (1967) and Shafer (1976) also suggested a theory of belief in which the degree of belief in an event did not obey additivity. They focused on the representation of uncertainty by *belief functions*. There is a vast literature that followed, often referred to as "imprecise probabilities" (see Walley, 1991). Most of this literature, however, does not address the question of decision making. By contrast, Schmeidler had decision theory in mind, and he sought a notion of integration that would generalize standard expectation when the capacity ν happens to be additive. Such a notion of integration was suggested by Choquet (1953).

Choquet Integral To understand the gist of the Choquet integral,[47] suppose that Σ is a σ-algebra (e.g., the power set 2^S) and consider a positive and bounded Σ-measurable function $\phi : S \to \mathbb{R}$. The *Choquet integral* of ϕ with respect to a capacity ν is given by:

$$\int \phi \, d\nu = \int_0^\infty \nu\left(\{s \in S : \phi(s) \geq t\}\right) dt, \tag{14}$$

where on the right-hand side we have a Riemann integral. To see why the Riemann integral is well defined, let $E_t = \{s \in S : \phi(s) \geq t\}$ be the the upper contour set of ϕ at t. Since ϕ is Σ-measurable, $E_t \in \Sigma$ for all $t \in \mathbb{R}$. Define the *survival function* $G_\nu : \mathbb{R} \to \mathbb{R}$ of ϕ with respect to ν by

$$G_\nu(t) = \nu(E_t), \qquad \forall t \in \mathbb{R}.$$

Using this function, we can write (14) as

$$\int \phi \, d\nu = \int_0^\infty G_\nu(t) \, dt.$$

The upper contour sets $\{E_t\}_{t \in \mathbb{R}}$ are nested, with $E_t \supseteq E_{t'}$ if $t \leq t'$. Since ν is a capacity, $\nu(E_t) \geq \nu(E_{t'})$ if $t \leq t'$, and so G_ν is a decreasing function. Moreover, since ϕ is positive and bounded, the function G_ν is positive with compact support. By standard results on Riemann integration, the Riemann integral $\int_0^\infty G_\nu(t) \, dt$ exists, and so the Choquet integral (14) is well defined. Moreover, it is easy to see that when ν is additive the Choquet integral reduces to a standard Lebesgue integral.

[47] We refer the interested reader to Denneberg (1994) and to Marinacci and Montrucchio (2004) for detailed expositions of Choquet integration.

When ϕ is a simple function,[48] the Choquet integral can be written in a couple of useful equivalent ways. For simplicity, suppose that S is a finite state space $\{1, ..., n\}$ and set $\Sigma = 2^S$. Assume further that the states of the world are ordered so that ϕ is decreasing:

$$\phi(1) \geq \phi(2) \geq ... \geq \phi(n) \geq 0$$

Then, it is easy to see that

$$\begin{aligned} \int_S \phi \, dv &= [\phi(1) - \phi(2)] \, \nu(\{1\}) \\ &\quad + [\phi(2) - \phi(3)] \, \nu(\{1,2\}) \\ &\quad \cdots \\ &\quad + [\phi(n-1) - \phi(n)] \, \nu(\{1, ..., n\}) \, . \end{aligned}$$

Equivalently,

$$\begin{aligned} \int_S \phi \, dv &= \phi(1)\nu(\{1\}) \\ &\quad + \phi(2) [\nu(\{1,2\}) - \nu(\{1\})] \\ &\quad + \phi(3) [\nu(\{1,2,3\}) - \nu(\{1,2\})] \\ &\quad \cdots \\ &\quad + \phi(n) [\nu(S) - \nu(\{1, ..., n-1\})] \, . \end{aligned}$$

This last way of writing the Choquet integral sheds light on its additivity properties. For, it shows that the Choquet integral $\int_S \phi \, dv$ is the standard Lebesgue integral of ϕ with respect to an (additive) probability measure $P : 2^S \to [0, 1]$ defined by

$$\begin{aligned} P(1) &= \nu(\{1\}) \\ P(2) &= \nu(\{1,2\}) - \nu(\{1\}) \\ &\cdots \\ P(n) &= \nu(S) - \nu(\{1, ..., n-1\}) \end{aligned}$$

If we were to consider a different function ψ, we may find that the appropriate additive measure differs from ϕ. More precisely, if it so happens that

$$\psi(1) \geq \psi(2) \geq \cdots \geq \psi(n)$$

that is, that a single permutation of the states render both ϕ and ψ non-increasing, then the Choquet integrals of both ϕ and ψ can be replaced by standard integrals with respect to the *same* measure P.

[48] That is, there is a partition $\{E_i\}_{i=1}^n \subseteq \Sigma$ of S and a finite collection $\{\alpha_i\}_{i=1}^n$ of scalars such that $\phi = \sum_{i=1}^n \alpha_i 1_{E_i}$.

In this case, we say that ϕ and ψ are *comonotonic* (short for "commonly monotonic"), and we find that the integral will also preserve its additivity:

$$\int_S (\phi + \psi) dv = \int_S \phi \, dv + \int_S \psi dv. \tag{15}$$

If, however, we find that

$$\phi(1) > \phi(2) \quad \text{and} \quad \psi(1) < \psi(2)$$

we may conclude that ϕ and ψ are not comonotonic: any permutation of the states that makes ϕ non-increasing has to place state 2 after state 1, and the converse holds for ψ. In this case, each of the Choquet integrals of ϕ and of ψ equals an integral of the respective function with respect to a certain probability measure, but it will typically not be the same measure. As a result, the Choquet integral generally does not satisfy additivity, except under comonotonicity.

In general state spaces S, not necessarily finite, two Σ-measurable functions $\phi : S \to \mathbb{R}$ and $\psi : S \to \mathbb{R}$ are said to be comonotonic if

$$[\phi(s) - \phi(s')] [\psi(s) - \psi(s')] \geq 0, \qquad \forall s, s' \in S, \tag{16}$$

that is, if it is never the case that $\phi(s) > \phi(s')$ and $\psi(s) < \psi(s')$ for some states s and s'. Dellacherie (1971) showed that (15) holds in general state spaces if ϕ and ψ are comonotonic. This important additivity result can be proved by suitably generalizing the argument outlined before for finite state spaces.

Schmeidler (1989) Schmeidler (1989) axiomatized Choquet expected utility in the AA setup. The key innovation relative to the AA axioms AA.1-AA.4 was to restrict the independence axiom AA.3 to *comonotonic* acts, that is, acts $f, g \in \mathcal{F}$ for which it is never the case that both $f(s) \succ f(s')$ and $g(s) \prec g(s')$ for some states of the world s and s'. This is the preference version of comonotonicity.

S.3 COMONOTONIC INDEPENDENCE: for any pairwise comonotonic acts $f, g, h \in \mathcal{F}$ and any $0 < \alpha < 1$,

$$f \succ g \Rightarrow \alpha f + (1-\alpha)h \succ \alpha g + (1-\alpha)h. \tag{17}$$

According to this axiom, the DM's preference between two comonotonic acts f and g is not affected by mixing them with another act h that is comonotonic with both. The intuition behind this axiom can best be explained by observing that the classical independence axiom may not be very compelling in the presence of uncertainty. For example, assume that there are two states of the world, and two vNM lotteries $P \succ Q$. Let $f = (P, Q)$ and $g = (Q, P)$. Suppose that, due to ignorance about the state of the world, the DM is driven to express indifference, $f \sim g$. By AA's independence, for every h we will observe

$$\frac{1}{2}f + \frac{1}{2}h \sim \frac{1}{2}g + \frac{1}{2}h$$

However, for $h = g$ this implies that $\frac{1}{2}f + \frac{1}{2}g \sim g$, despite the fact that the act $\frac{1}{2}f + \frac{1}{2}g$ is risky while g is uncertain.

In this example, g can serve as a hedge against the uncertainty inherent in f, but it clearly cannot hedge against itself. The standard independence axiom is too demanding, because it does not distinguish between mixing operations $\alpha f + (1-\alpha)h$ that reduce uncertainty (via hedging) and mixing operations that do not. Restricting the independence axiom to pairwise comonotonic acts neutralizes this asymmetric effect of hedging.

Similar remarks can be made to explain why the Ellsberg's experiments patterns violate the independence axiom AA.3. For instance, consider Ellsberg's single urn experiment. Too see that its pattern (8) violates AA.3, consider the mixed bets

$$\frac{1}{2}1_R + \frac{1}{2}1_G \quad \text{and} \quad \frac{1}{2}1_Y + \frac{1}{2}1_G.$$

By definition,

$$\left(\frac{1}{2}1_R + \frac{1}{2}1_G\right)(s) = \begin{cases} \frac{1}{2} & s \in R \cup G \\ 0 & \text{else} \end{cases} = \left(\frac{1}{2}1_{R\cup G} + \frac{1}{2}0\right)(s),$$

that is,

$$\frac{1}{2}1_R + \frac{1}{2}1_G = \frac{1}{2}1_{R\cup G} + \frac{1}{2}0.$$

Similarly,

$$\frac{1}{2}1_Y + \frac{1}{2}1_G = \frac{1}{2}1_{Y\cup G} + \frac{1}{2}0.$$

Hence, by AA.3 we have

$$1_R \succ 1_Y \Longrightarrow \frac{1}{2}1_R + \frac{1}{2}1_G \succ \frac{1}{2}1_Y + \frac{1}{2}1_G \Longleftrightarrow \frac{1}{2}1_{R\cup G} + \frac{1}{2}0 \succ \frac{1}{2}1_{Y\cup G} + \frac{1}{2}0 \tag{18}$$

and

$$1_{Y\cup G} \succ 1_{R\cup G} \Longrightarrow \frac{1}{2}1_{Y\cup G} + \frac{1}{2}0 \succ \frac{1}{2}1_{R\cup G} + \frac{1}{2}0,$$

a contradiction. Instead, under comonotonic independence we do not reach this contradiction since bets 1_R, 1_Y and 1_G are not pairwise comonotonic, and so the first implication in (18) may not hold.

Using Axiom S.3, Schmeidler (1989) was able to prove the following representation theorem, which generalizes the subjective expected utility representation established by Theorem 1 by allowing for possibly non-additive probabilities. The proof of the result is based on some results on Choquet integration established in Schmeidler (1986).

Theorem 2 *Let $\succsim$ be a preference defined on $\mathcal{F}$. The following conditions are equivalent:*

(i) $\succsim$ satisfies axioms AA.1, AA.2, S.3, AA.4, and AA.5;

(ii) there exists a non-constant function $u : X \to \mathbb{R}$ and a capacity $\nu : \Sigma \to [0,1]$ such that, for all $f, g \in \mathcal{F}$, $f \succsim g$ if and only if

$$\int_S \left(\sum_{x \in \operatorname{supp} f(s)} u(x) f(s) \right) d\nu(s) \geq \int_S \left(\sum_{x \in \operatorname{supp} g(s)} u(x) g(s) \right) d\nu(s). \tag{19}$$

Moreover, ν is unique and u is cardinally unique.

Gilboa (1987), Wakker (1989a, 1989b), and Nakamura (1990) established purely subjective versions of Schmeidler's representation result.[49] Sarin and Wakker (1992) showed that the existence of a suitable rich collection of unambiguous events substantially streamlines the derivation of Schmeidler's representation through a simple cumulative dominance condition.

3.3 Maxmin expected utility

Schmeidler's model is a generalization of Anscombe-Aumann's in a way that allows us to cope with uncertainty, or ambiguity. The capacity in the model can be interpreted as a lower bound on probabilities. Specifically, let $\Delta(\Sigma)$ be the collection of all finitely additive probability measures $P : \Sigma \to [0,1]$ and define the *core* of ν to be, as in cooperative game theory,

$$core(\nu) = \{P \in \Delta(\Sigma) : P(E) \geq \nu(E) \text{ for all } E \in \Sigma\}.$$

If $core(\nu) \neq \emptyset$, we may think of $\nu(E)$ as the lower bound on $P(E)$, and then ν is a concise way to represent a set of probabilities, presumably those that are considered possible.[50] This interpretation was also suggested by Dempster (1967) and Shafer (1976) for the special case of capacities that are belief functions.

Schmeidler (1986) has shown that if ν is *convex* in the sense that

$$\nu(E) + \nu(E') \leq \nu(E \cup E') + \nu(E \cap E'), \qquad \forall E, E' \in \Sigma,$$

then

$$\int_S \phi \, dv = \min_{P \in core(\nu)} \int_S \phi \, dP \tag{20}$$

[49] Nakamura and Wakker's papers use versions of the so-called *tradeoff method* (see Kobberling and Wakker, 2003, for a detailed study of this method and its use in the establishment of axiomatic foundations for choice models).

[50] As Diaconis and Freedman (1986) observe in the Conclusion of their classic paper on Bayesian consistency, "Often, a statistician has prior information on a problem (say as to the rough order of magnitude of a key parameter), but does not have really a sharply defined prior probability distribution. Many different distributions would have the right qualitative features and a Bayesian typically chooses one of the basis of mathematical convenience."

for every Σ-measurable bounded function $\phi : S \to \mathbb{R}$ (see also Rosenmueller, 1971 and 1972). Thus, when the capacity ν happens to be convex (e.g., a belief function a la Dempster-Shafer), Choquet integration has a simple and intuitive interpretation: a DM who evaluated an act f by the Choquet integral of its utility profile $u \circ f$ can be viewed as if she entertained a *set* of possible probabilities, $core(\nu)$, and evaluated each act by its minimal expected utility, over all probabilities in the set.

There is a simple behavioral condition that characterizes CEU preferences with convex ν. To introduce it, denote by $B_0(\Sigma)$ the vector space of all simple functions $\phi : S \to \mathbb{R}$ and consider the Choquet functional $I : B_0(\Sigma) \to \mathbb{R}$ given by $I(\phi) = \int \phi \, dv$. This functional is easily seen to be concave when (20) holds. Actually, according to a classic result of Choquet (1953), I is concave if and only if its capacity ν is convex.[51] This concavity property suggests the following convexity axiom, due to Schmeidler (1989), which models a negative attitude toward ambiguity.

S.6 UNCERTAINTY AVERSION: for any $f, g \in \mathcal{F}$ and any $0 < \alpha < 1$, we have

$$f \sim g \Rightarrow \alpha f + (1-\alpha) g \succsim f.$$

Thus, uncertainty aversion states that mixing, through randomization, between equivalent acts can only make the DM better off. For example, in Ellsberg's example it is natural to expect that DMs prefer to hedge against ambiguity by mixing acts IIB and IIW, that is,

$$\alpha IIB + (1-\alpha) IIW \succsim IIB \sim IIW, \qquad \forall \alpha \in [0,1].$$

This mixing can be thus viewed as a form of hedging against ambiguity that the DM can choose.[52]

Theorem 3 *In Theorem 2, $\succsim$ satisfies axiom S.6 if and only if the capacity ν in (19) is convex.*

This result of Schmeidler (1989) shows that convex capacities characterize ambiguity averse Choquet expected utility DMs (in the sense of axiom S.6). Since most DMs are arguably ambiguity averse, this is an important result in Choquet expected utility theory. Moreover, relating this theory to maximization of the worst-case expected utility over a set of probabilities has several advantages. First, it obviates the need to understand the unfamiliar concept of Choquet integration. Second, it provides a rather intuitive, if extreme, cognitive account of the decision process: as in classical statistics, the DM entertains several probability measures as potential beliefs. Each such "belief" induces an expected utility index for each act. Thus, each act has many expected utility values. In the absence of second-order beliefs, the cautious DM chooses the worst-case expected utility as summarizing the act's desirability. Wakker (1990, 1991) established several important behavioral properties and characterizations of concave/convex capacities in the CEU model.

[51] See Marinacci and Montrucchio (2004) p. 73. They show on p. 78 that (20) can be derived from this result of Choquet through a suitable application of the Hahn-Banach Theorem.

[52] Klibanoff (2001a, 2001b) studied in detail the relations between randomization and ambiguity aversion.

Gilboa and Schmeidler (1989) This account of Choquet expected utility maximization also relates to the maxmin criterion of Wald (1950; see also Milnor, 1954). However, there are many natural sets of probabilities that are not the core of any capacity. Assume, for example, that there are three states of the world, $S = \{1, 2, 3\}$. Assume that the DM is told that, if state 1 is not the case, then the (conditional) probability of state 2 is at least 2/3. If this is all the information available to her, she knows only that state 2 is at least twice as likely than state 3. Hence the set of probability vectors $P = (p_1, p_2, p_3)$ that reflects the DM's knowledge consists of all vectors such that

$$p_2 \geq 2p_3$$

It is easy to verify that this set is not the core of a capacity. Similarly, one may consider a DM who has a certain probability measure P in mind, but allows for the possibility of error in its specification. Such a DM may consider a set of probabilities

$$C = \{Q \in \Delta(\Sigma) : \|P - Q\| < \varepsilon\}$$

for some norm $\|\cdot\|$ and $\varepsilon > 0$, and this set is not the core of any capacity (such sets were used in Nishimura and Ozaki, 2007).

It therefore makes sense to generalize Choquet expected utility with convex capacities to the maxmin rule, where the minimum is taken over general sets of probabilities. Decision rules of this type have been suggested first by Hurwicz (1951), under the name of Generalized Bayes-minimax principle, and then by Smith (1961), Levi (1974, 1980), and Gärdenfors and Sahlin (1982).[53]

Gilboa and Schmeidler (1989) provided an axiomatic model of maxmin expected utility maximization ("MMEU", also referred to as "MEU"). This model is also formulated in the AA framework and, like the Choquet expected utility model, is based on a suitable weakening of the Independence Axiom AA.3. Schmeidler's comonotonic independence axiom restricted AA.3 to the case that all acts are pairwise comonotonic. This rules out obvious cases of hedging, but it may allow for more subtle ways in which expected utility can be "smoothed out" across states of the world. A more modest requirement restricts the independence condition to the case in which the act h is constant:

GS.3 C-INDEPENDENCE: for all acts $f, g \in \mathcal{F}$ and all lottery acts p,

$$f \succsim g \Rightarrow \alpha f + (1-\alpha)p \succsim \alpha g + (1-\alpha)p, \qquad \forall \alpha \in [0,1].$$

C-Independence is essentially weaker than Comonotonic Independence S.3 because lottery (constant) acts are comonotonic with all other acts.[54] The axiom is arguably easier to accept because

[53] Recently, related ideas appeared in mathematical finance (see Artzner, Delbaen, Eber, and Heath, 1997, 1999).

[54] Schmeidler required that all three acts be pairwise comonotonic, whereas C-Independence does not restrict attention to comonotonic pairs (f, g). Thus, C-Independence is not, strictly speaking, weaker than Comonotonic Independence. However, in the presence of Schmeidler's other axioms, Comonotonic Independence is equivalent to the version in which f and g are not required to be comonotonic.

the mixture with a lottery act can be viewed as a change of the unit of measurement. Indeed, this axiom may be viewed as the preference version of the following property of real-valued functionals: a functional $I : B_0(\Sigma) \to \mathbb{R}$ is said to be *translation invariant* if

$$I(\alpha\phi + k) = \alpha I(\phi) + I(k), \qquad \forall \alpha \in \mathbb{R},$$

given any $\phi \in B_0(\Sigma)$ and any constant function k.[55]

Gilboa and Schmeidler thus used a weaker version of the independence axiom, but they also imposed the uncertainty aversion axiom S.6. Both axioms GS.3 and S.6 follow from the independence axiom AA.3. Thus, the following representation result, due to Gilboa and Schmeidler (1989), generalizes Theorem 1 by allowing for possibly nonsingleton sets of probabilities.

Theorem 4 *Let $\succsim$ be a preference defined on $\mathcal{F}$. The following conditions are equivalent:*

(i) $\succsim$ satisfies axioms AA.1, AA.2, GS.3, AA.4, AA.5, and S.6;

(ii) there exists a non-constant function $u : X \to \mathbb{R}$ and a convex and compact set $C \subseteq \Delta(\Sigma)$ of probability measures such that, for all $f, g \in \mathcal{F}$,

$$f \succsim g \Leftrightarrow \min_{P \in C} \int_S \left(\sum_{x \in \operatorname{supp} f(s)} u(x) f(s) \right) dP(s) \geq \min_{P \in C} \int_S \left(\sum_{x \in \operatorname{supp} f(s)} u(x) g(s) \right) dP(s), \quad (21)$$

Moreover, C is unique and u is cardinally unique.

The set C is a singleton if and only if $\succsim$ satisfies the Independence Axiom AA.3. A slightly more interesting result actually holds, which shows that maxmin expected utility DMs reduce to subjective expected utility ones when their choices do not involve any hedging against ambiguity.[56]

Proposition 5 *In Theorem 4, C is a singleton if and only if, for all $f, g \in \mathcal{F}$,*

$$f \sim g \Rightarrow \frac{1}{2} f + \frac{1}{2} g \sim g.$$

When C is not a singleton, the model can express more complex states of knowledge, reflected by various sets C of probabilities. For applications in economic theory, the richness of the maxmin model seems to be important. In particular, one may consider any model in economic theory and enrich it by adding some uncertainty about several of its parameters. By contrast, in order to formulate Choquet expected utility, one needs to explicitly consider the state space and the capacity defined on it. Often, this exercise may be intractable.

[55] See Ghirardato, Klibanoff, and Marinacci (1998) for details.
[56] See Ghirardato, Maccheroni, and Marinacci (2004) for details.

By contrast, for some practical applications such as in medical decision making, the richness of the maxmin model may prove a hindrance. Wakker (2010) presents the theory of decision making under risk and under ambiguity geared for such applications. He focuses on capacities as a way to capture ambiguity, rather than on sets of probabilities.[57]

The maxmin model allows for more degrees of freedom than the CEU model, but it does not generalize it. In fact, the overlap of the two models is described in Theorem 3 and occurs when the uncertainty averse axiom S.6 holds. But, whereas uncertainty aversion – through axiom S.6 – is built into the decision rule of the maxmin model, Choquet expected utility can express attitudes of uncertainty liking. This observation in part motivated the search by Ghirardato, Maccheroni, and Marinacci (2004) of a class of preferences that may not satisfy S.6 and is able to encompass both CEU and MMEU preferences. We review this contribution below.

Finally, Casadesus-Masanell, Klibanoff, and Ozdenoren (2000), Ghirardato, Maccheroni, Marinacci, and Siniscalchi (2003), and Alon and Schmeidler (2010) established purely subjective versions of Gilboa and Schmeidler's representation result.[58]

Countably additive priors Theorem 4 considers the set $\Delta(\Sigma)$ of all finitely additive probabilities. In applications, however, it is often important to consider countably additive probabilities, which have very convenient analytical properties that many important results in probability theory crucially rely upon.

The behavioral condition that underlies countably additive priors is Monotone Continuity, introduced by Arrow (1970) to characterize countable additivity of the subjective probability P in Savage's model.

MC MONOTONE CONTINUITY: If $f, g \in \mathcal{F}$, $x \in X$, $\{E_n\}_{n\geq 1} \in \Sigma$ with $E_1 \supseteq E_2 \supseteq \ldots$ and $\bigcap_{n\geq 1} E_n = \emptyset$, then $f \succ g$ implies that there exists $n_0 \geq 1$ such that $xE_{n_0}f \succ g$.

Marinacci (2002a) and Chateauneuf, Maccheroni, Marinacci, and Tallon (2005) showed that this condition keeps characterizing countable additivity in the MMEU model. Next we state a version of their results, a countably additive counterpart of Theorem 4. Here $Q \ll P$ means that Q is absolutely continuous with respect to P, i.e., $P(E) = 0$ implies $Q(E) = 0$ for all $E \in \Sigma$.

[57] Wakker (2010) also introduces the gain-loss asymmetry that is one of the hallmarks of Prospect Theory (Kahneman and Tversky, 1979). The combination of gain-loss asymmtry with rank-dependent expected utility (Quiggin, 1982, Yaari, 1987) resulted in Cumulative Prospect Theory (CPT, Tversky and Kahneman, 1992). When CPT is interpreted as dealing with ambiguity, it is equivalent to Choquet expected utility with the additional refinement of distinguishing gains from losses.

[58] For a critical review of the maxmin and other non-Bayesian models, see Al-Najjar and Weinstein (2009) (see Mukerji , 2009, and Siniscalchi, 2009b, for a discussion).

Theorem 6 *In Theorem 4, $\succsim$ satisfies Axiom MC if and only if all probabilities in C are countably additive. In this case, there exists $P \in C$ such that $Q \ll P$ for all $Q \in C$.*

Besides the countable additivity of priors, axiom MC also delivers the existence of a "control" prior $P \in C$ relative to which all other priors $Q \in C$ are absolutely continuous.[59]

In decision theory the use of countably additive priors has been often debated, most forcefully by de Finetti and Savage themselves, who argued that it is a purely technical property that, if anything, actually impairs the analysis (e.g., over countable state spaces it is not possible to define uniform priors that are countably additive). However, Arrow's characterization of countably additive priors in Savage's model through Monotone Continuity and its MMEU version in Theorem 6 show that behaviorally this technically most useful property requires a relatively small extra baggage compared to the basic axioms of the finitely additive case.[60]

Equivalent priors A minimal consistency requirement among priors in C is that they agree on what is possible or impossible. Formally, this is the case if any two priors P and P' in C are *equivalent*, i.e., if they are mutually absolutely continuous ($P(E) = 0$ if and only if $P'(E) = 0$ for all $E \in \Sigma$). Epstein and Marinacci (2007) provide a behavioral condition that ensures this minimal consistency among priors, which is especially important in dynamic problems that involve priors' updating.

Interestingly, this condition turns out to be a translation in a choice under uncertainty setup of a classic axiom introduced by Kreps (1979) in his seminal work on menu choices. Given any two consequences x and y, let

$$x \vee y = \begin{cases} x & \text{if } x \succsim y \\ y & \text{otherwise} \end{cases}$$

and given any two acts f and g, define the act $f \vee g$ by $(f \vee g)(s) = f(s) \vee g(s)$ for each $s \in S$.

GK GENERALIZED KREPS: For all $f, f', g \in \mathcal{F}$, $f \sim f \vee f' \Rightarrow f \vee g \sim (f \vee g) \vee f'$.

In every state, the act $f \vee f'$ gives the better of the two outcomes associated with f and f'. Thus we say that $f \vee f'$ weakly improves f in 'the direction' f'. GK requires that if an improvement of f in direction f' has no value, then the same must be true for an improvement in direction f' of any act (here $f \vee g$) that improves f. The next result of Epstein and Marinacci (2007) shows that for

[59] As Chateauneuf et al (2005) show, this control prior exists because, under Axiom MC, the set C is weakly compact, a stronger compactness condtion than the weak*-compactness that C features in Theorem 4. Their results have been generalized to variational preferences by Maccheroni et al (2006).

[60] In this regard, Arrow (1970) wrote that "the assumption of Monotone Continuity seems, I believe correctly, to be the harmless simplification almost inevitable in the formalization of any real-life problem." See Kopylov (2010) for a recent version of Savage's model under Monotone Continuity.

In many applications, countable additivity of the measure(s) necessitates the restriction of the algebra of events to be a proper subset of 2^S. Ignoring many events as "non-measurable" may appear as sweeping the continuity problem under the measurability rug. However, this approach may be more natural if one does not start with the state space S as primitive, but derives it as the semantic model of a syntactic system, where propositions are primitive.

maxmin preferences this seemingly innocuous axiom is equivalent to the mutual absolute continuity of priors.

Theorem 7 *In Theorem 4, $\succsim$ satisfies Axiom GK if and only if the probabilities in C are equivalent.*

3.4 Unanimity preferences

Another way to deal with ambiguity is to relax the completeness of preferences. Indeed, because of the poor information that underlies ambiguity, the DM may not be able to rank some pairs of acts. If so, one of the most basic assumptions in decision theory, namely, that preferences are complete, may be relaxed because of ambiguity.

This is the approach proposed by Truman Bewley. Incomplete preferences were already studied by Aumann (1962), interpreted as a DM's inability to decide between some pairs of alternatives. Building on Aumann's work, Bewley presented in 1986 a model of incomplete preferences in the context of uncertainty, which appeared as Bewley (2002). In his model the Weak Order Axiom AA.1 is replaced by two weaker assumptions.

B.1a PARTIAL ORDER**:** $\succsim$ on $\mathcal{F}$ is reflexive and transitive.

Hence, $\succsim$ is no longer required to be complete. The DM, however, knows her tastes: the only reason for incompleteness is ignorance about probabilities. For this reason, Bewley assumes the next weak form of completeness, which only applies to lottery acts.

B.1b C-COMPLETENESS**:** for every lottery acts $p, q \in \Delta(X)$, $p \succsim q$ or $q \succsim p$.

In other words, B.1 requires the risk preference $\succsim_\Delta$ to be complete. Using these two axioms, Gilboa, Maccheroni, Marinacci, and Schmeidler (2010) established the following general form of Bewley's representation theorem.[61]

Theorem 8 *Let $\succsim$ be a preference defined on $\mathcal{F}$. The following conditions are equivalent:*

(i) $\succsim$ satisfies axioms B.1, and AA.2-AA.5;

(ii) there exists a non-constant function $u : X \to \mathbb{R}$ and a convex and compact set $C \subseteq \Delta(\Sigma)$ of probability measures such that, for all $f, g \in \mathcal{F}$,

$$f \succsim g \Leftrightarrow \int_S \left(\sum_{x \in \operatorname{supp} f(s)} u(x) f(s) \right) dP(s) \geq \int_S \left(\sum_{x \in \operatorname{supp} f(s)} u(x) g(s) \right) dP(s), \qquad \forall P \in C. \tag{22}$$

[61] A caveat: the unanimity rule (22) is slightly different from Bewley's, who represents strict preference by unanimity of strict inequalities. This is generally not equivalent to representation of weak preference by unanimity of weak inequalities.

Moreover, C is unique and u is cardinally unique.

In this representation a set of probability measures C arises, interpreted as the probabilistic models that are compatible with the DM's information. Two acts f and g are comparable only when their expected utilities with respect to the probabilities in C unanimously rank one act over the other. If this is not the case – that is, if the probabilities in C do not agree in ranking of the two acts– the DM is unable to rank the two acts.

When preferences are incomplete, the model does not always specify what the DM will do. In particular, acts are not evaluated by a numerical index V that represents preferences and that makes it possible to formulate the optimization problems that most economic applications feature. To complete the model, one needs to add some assumptions about choices in case preferences do not have a maximum. One possibility is to assume that there exists a status quo, namely, an alternative that remains the default choice unless it is dethroned by another alternative that is unanimously better. This might be a rather reasonable descriptive model, especially of organizations, but it is considered by many to be less than rational. Recently, Ortoleva (2010) reconsidered Bewley's inertia insight from a different angle by showing, within a full-fledged axiomatic model, how status quo biases may lead to incomplete preferences.

Another approach suggests to complete preferences based on the same set of probabilities C. Gilboa et al. (2010) offer a model involving two preference relations, and show that certain axioms, stated on each relation separately as well as relating the two, are equivalent to a joint representation of the two relations by the same set of probabilities C: one by the unanimity rule, and the other – by the maxmin rule. Their results provide a bridge between the two classic representations (21) and (22), as well as a possible account by which maxmin behavior might emerge from incomplete preferences.

3.4.1 Unanimity, scenarios, and uncertainty aversion

Ghirardato, Maccheroni, and Marinacci (GMM, 2004) used some insights from Bewley's unanimity representation to remove the Uncertainty Aversion axiom S.6 in the derivation of Gilboa and Schmeidler (1989) and, in this way, to propose a class of preferences that encompasses both Choquet and maxmin preferences. To this end, they consider the following definition.

Definition 9 *A preference $\succsim$ on $\mathcal{F}$ is said to be invariant biseparable if it satisfies axioms AA.1, AA.2, GS.3, AA.4, and AA.5.*

Invariant biseparable (IB) preferences thus satisfy all AA axioms, except for the independence axiom AA.3, which is replaced by the C-Independence axiom GS.3 of Gilboa and Schmeidler (1989).[62]

[62] The name biseparable originates in Ghirardato and Marinacci (2001, 2002), which we will discuss later.

Thanks to this key weakening, invariant biseparable preferences include as special cases both CEU and MMEU preferences: the former constitute the special case when the Comonotonic Independence axiom S.3 holds, while the latter – when the Uncertainty Aversion axiom S.6 holds.

The main tool that GMM use to study IB preferences is an auxiliary relation $\succsim^*$ on $\mathcal{F}$. Specifically, given any two acts $f, g \in \mathcal{F}$, act f is said to be *unambiguously (weakly) preferred* to g, written $f \succsim^* g$, if

$$\alpha f + (1-\alpha)\, h \succsim \alpha g + (1-\alpha)\, h$$

for all $\alpha \in [0,1]$ and all $h \in \mathcal{F}$. In words, $f \succsim^* g$ holds when the DM does not find any possibility of hedging against or speculating on the ambiguity that she may perceive in comparing f and g. GMM argue that this DM's choice pattern reveals that ambiguity does not affect her preference between f and g, and this motivates the "unambiguously preferred" terminology.

The unambiguous preference relation is, in general, incomplete. This incompleteness is due to ambiguity

Lemma 10 *The following statements hold:*

(i) If $f \succsim^* g$, then $f \succsim g$.

(ii) $\succsim^*$ satisfies axioms B.1, AA.2, and AA.3

(iii) $\succsim^*$ is the maximal restriction of $\succsim$ satisfying the independence axiom AA.3.[63]

By (i) and (ii), the unambiguous preference $\succsim^*$ is a restriction of the primitive preference relation $\succsim$ that satisfies reflexivity, transitivity, monotonicity, and independence. By (iii), it is the maximal such restriction that satisfies independence.[64]

The next result proves, along the lines of the Bewley-type representation (22), that the unambiguous preference can be represented by a set of priors.

Proposition 11 *Let $\succsim$ be an IB preference on $\mathcal{F}$. Then, there exists a function $u : X \to \mathbb{R}$ and a convex and compact set $C \subseteq \Delta(\Sigma)$ of probability measures such that, for all $f, g \in \mathcal{F}$,*

$$f \succsim^* g \Leftrightarrow \int_S u(f)\; dP(s) \geq \int_S u(g)(s)\; dP(s), \qquad \forall P \in C. \tag{23}$$

In words, f is unambiguously weakly preferred to g if and only if every probability $P \in C$ assigns a weakly higher expected utility to f. It is natural to interpret each prior $P \in C$ as a "possible scenario" that the DM envisions, so that unambiguous preference corresponds to preference in every scenario. GMM thus argue that C represents the (subjective) perception of ambiguity of the DM, and that the DM perceives ambiguity in a decision problem if C is not a singleton.

[63] That is, if $\succsim' \subseteq \succsim$ and $\succsim'$ satisfies independence, then $\succsim' \subseteq \succsim^*$.
[64] This latter feature of $\succsim^*$ relates this notion to an earlier one by Nehring (2001), as GMM discuss.

The relation $\succsim^*$ thus makes it possible to elicit a set of priors C for a general IB preference $\succsim$. When $\succsim$ is a MMEU preference, C is the set of priors of the maxmin representation (21). When $\succsim$ is a CEU preference that satisfies axiom S.6, C is the core of the representing capacity ν.[65]

More generally, GMM prove a representation theorem for IB preferences based on the set C, which generalizes Theorems 2 and 4. To this end, given any act f consider its expected utility profile $\left\{\int_S u(f)\, dP(s) : P \in C\right\}$ under C. Write $f \asymp g$ if two acts f and g feature isotonic profiles, that is,

$$\int_S u(f(s))\, dP'(s) \geq \int_S u(f(s))\, dP''(s) \Leftrightarrow \int_S u(g(s))\, dP'(s) \geq \int_S u(g(s))\, dP''(s), \qquad \forall P', P'' \in C.$$

Intuitively, in this case the DM perceives a similar ambiguity in both acts. For example, $p \asymp q$ for all lottery acts, which are unambiguous.

It is easy to see that $\asymp$ is an equivalence relation. Denote by $[f]$ the relative equivalence class determined by an act f, and by $\mathcal{F}_{|\asymp}$ the quotient space of $\mathcal{F}$ that consists of these equivalence classes.

Theorem 12 *Let $\succsim$ be an IB preference on $\mathcal{F}$. Then, there exists a function $u : X \to \mathbb{R}$, a convex and compact set $C \subseteq \Delta(\Sigma)$ of probability measures, and a function $a : \mathcal{F}_{|\asymp} \to [0,1]$ such that $\succsim$ is represented by the preference functional $V : \mathcal{F} \to \mathbb{R}$ given by*

$$V(f) = a([f]) \min_{P \in C} \int_S u(f(s))\, dP(s) + (1 - a([f])) \max_{P \in C} \int_S u(f(s))\, dP(s), \tag{24}$$

where u and C represent $\succsim^*$ in the sense of (23).

Moreover, C is unique, u is cardinally unique, and a is unique on $\mathcal{F}_{|\asymp}$ (with the exclusion of the equivalence class $[p]$ of lottery acts).

In this representation, due to GMM, the revealed perception of ambiguity, embodied by the set C, is separated from the DM's reaction to it, modelled by the function a. Both C and a are derived endogenously within the model. When a is constant equal to 1, we get back to the maxmin representation. Otherwise, we have a more general choice criterion that may well exhibit ambiguity loving (the polar case is, clearly, when a is a constant equal to 0).

Giraud (2005) and Amarante (2009) studied invariant biseparable preferences, with novel important insights. Amarante established an alternative characterization of IB preferences through the two stage form

$$V(f) = \int_\Delta \left(\int_S u(f(s))\, dP(s) \right) d\nu(P)$$

where ν is a capacity over the set of measures $\Delta = \Delta(\Sigma)$ on S. In a statistical decision theory vein, the capacity ν quantifies DM's beliefs over the possible models P. Giraud thoroughly studies

[65] GMM also show the form that C takes for some CEU preferences that do not satisfy S.6.

a similar representation, motivated by the desire to incorporate probabilistic information in a choice under ambiguity framework.

Finally, Siniscalchi (2006a) investigates an interesting class of invariant biseparable preferences that satisfy a local no-hedging condition that gives preferences a piecewise structure that makes them SEU on each component (see Castagnoli, Maccheroni, and Marinacci, 2003, for a related representation).

a-MEU Preferences In the special case when the function a is constant the representation (24) reduces to

$$V(f) = a \min_{P \in C} \int_S u(f(s))\, dP(s) + (1-a) \max_{P \in C} \int_S u(f(s))\, dP(s). \tag{25}$$

This is the a-MEU criterion that Jaffray (1989) suggested to combine Hurwicz (1951)'s criterion (see also Arrow and Hurwicz, 1972) with a maxmin approach. Intuitively, $a \in [0,1]$ measures the degree of the individual's pessimism, where $a = 1$ yields the maxmin expected utility model, and $a = 0$ – its dual, the maxmax expected utility model. However, this apparently natural idea turned out to be surprisingly tricky to formally pin down. GMM provided a specific axiom that reduces the IB representation to (25), where C represent $\succsim^*$ in the sense of (23). Because of this latter clause, when $a \in (0,1)$ it is not possible to take any pair u and C as a given and assume that the DMs' preferences are represented by the corresponding a-MEU criterion (25). In a nutshell, the issue is the uniqueness properties of C in (25), which are problematic when $a \in (0,1)$. We refer the reader to GMM and to Eichberger, Grant, and Kelsey (2008) and to Eichberger, Grant, Kelsey, and Koshevoy (2011) for more on this issue. (The latter paper shows that for finite state spaces the a-MEU axiomatized as a very special case of (24) by GMM only allows for $\alpha = 0$ or $\alpha = 1$).

3.5 Smooth preferences

The MMEU model discussed above is often viewed as rather extreme: if, indeed, a set of probability measures C is stipulated, and each act f is mapped to a range of expected utility values, $\left\{\int_S u(f)dp \mid p \in C\right\}$, why should such an f be evaluated by the minimal value in this interval? This worst-case scenario approach seems almost paranoid: why should the DM assume that nature[66] will choose a probability as if to spite the DM? Isn't it more plausible to allow for other ways that summarize the interval by a single number?

The extreme nature of the maxmin model is not evident from the axiomatic derivation of the model. Indeed, this model is derived from Anscombe-Aumann's by relaxing their independence axiom in two ways: first, by restricting it to mixing with a constant act (h above) and, second, by

[66] Relations betwenn ambiguity and games against nature are discussed in Hart, Modica, and Schmeidler (1994), Maccheroni, Marinacci, and Rustichini (2006a, 2006b), and Ozdenoren and Peck (2008).

assuming uncertainty aversion. These weaker axioms do not seem to reflect the apparently-paranoid attitude of the maxmin principle. A question then arises, how do these axioms give rise to such extreme uncertainty attitude?

In this context it is important to recall that the axiomatic derivation mentioned above is in the revealed preferences tradition, characterizing behavior that could be represented in a certain mathematical formula. An individual who satisfies the axioms can be thought of *as if* she entertained a set C of priors and maximized the minimal expected utility with respect to this set. Yet, this set of priors need not necessarily reflect the individual's knowledge. Rather, information and personal taste jointly determine the set C. Smaller sets may reflect both better information and a less averse uncertainty attitude. For example, an individual who bets on a flip of a coin and follows the expected utility axioms with respect to a probability $p = 0.5$ of "Head" may actually know that the probability p is 0.5, or she may have no clue about p but chooses the model $p = 0.5$ because she is insensitive to her ignorance about the true data generating process. Thus, information and attitude to uncertainty are inextricably intertwined in the set C. More generally, it is possible that the individual has objective information that the probability is in a set D, but behaves according to the maxmin expected utility rule with respect to a set $C \subset D$, reflecting her uncertainty attitude. This intuition has motivated the model of Gajdos, Hayashi, Tallon, and Vergnaud (2008) that axiomatically established the inclusion $C \subset D$ (some related ideas can be found in Wang, 2003a, and Giraud, 2005).

If, however, the set of priors C is interpreted cognitively a la Wald, that is, as the set of probabilities that are consistent with objectively available information, one may consider alternatives to the maxmin rule that, under this Waldean interpretation, has an extreme nature. One approach to address this issue is to assume that the DM has a prior probability over the possible probability distributions in C. Thus, if $\Delta(\Sigma)$ is the space of all "first order" probability distributions (viewed as data generating processes), and μ is a "second order" prior probability over them, one can use μ to have an averaging of sorts over all expected utility values of an act f.

Clearly, the expectation of expectations is an expectation. Thus, if one uses μ to compute the expectation of the expected utility, there will exist a probability $\hat{p}$ on S, given by

$$\hat{p} = \int_{\Delta(\Sigma)} p d\mu$$

such that for every act f (and every utility function u)

$$\int_{\Delta(\Sigma)} \left(\int_S u(f) dp \right) d\mu = \int_S u(f) d\hat{p}$$

In this case, the new model cannot explain any new phenomena, as it reduces to the standard Bayesian model. However, if the DM uses a non-linear function to evaluate expected utility values, one may explain non-neutral attitudes to uncertainty. Specifically, assume that

$$\varphi : \mathbb{R} \to \mathbb{R}$$

is an increasing function, and an act f is evaluated by

$$V(f) = \int_{\Delta(\Sigma)} \varphi \left(\int_S u(f) dp \right) d\mu.$$

In this representation, μ is read as representing information (about the probability model p), whereas φ reflects attitude towards ambiguity, with a concave φ corresponding to ambiguity aversion, similarly to the way that concave utility represents risk aversion in the classical model of expected utility under risk. In this way we have a separation between ambiguity perception, an information feature modelled by μ and its support, and ambiguity attitude, a taste trait modelled by φ and its shape.

This decision rule has been axiomatized by Klibanoff, Marinacci, and Mukerji (2005). It has become to be known as the *smooth model* of ambiguity because, under mild assumptions, V is a smooth functional, whereas the Choquet expected utility and the maxmin expected utility functionals are typically not everywhere differentiable (over the space of acts).

The notion of second order probabilities is rather old and deserves a separate survey.[67] This idea is at the heart of Bayesian statistics, where Bayes's rule is retained and a probability over probabilities over a state space is equivalent to a probability over the same space. Within decision theory, Segal (1987) already suggested that Ellsberg's paradox can be explained by second-order probabilities, provided that we allow the decision maker to violate the principle of reduction of compound lotteries. Specifically, Segal's model assumed that the second-order probabilities are used to aggregate first-order expectations via Quiggin's (1982) anticipated utility. Other related models have been proposed by Nau (2001, 2006, 2010), Chew and Sagi (2008), Ergin and Gul (2009), and Seo (2009). Halevy and Feltkamp (2005) proposed another approach according to which the decision maker does not err in the computation of probabilities, but uses a mis-specified model, treating a one-shot choice as if it were repeated.

As compared to Choquet expected utility maximization, the smooth preferences model, like the maxmin model, has the advantage of having a simple and intelligible cognitive interpretation. As opposed to both Choquet and maxmin expected utility models, smooth preferences have the disadvantage of imposing non-trivial epistemological demands on the DM: the smooth model requires the specification of a prior over probability models, that is, of a probability μ over a much larger space, $\Delta(\Sigma)$, something that may be informationally and observationally demanding.

That said, beyond the above mentioned separation, the smooth preferences model enjoys an additional advantage of tractability. Especially if one specifies a simple functional form for φ, one

[67] Bayes (1763) himself writes in his Proposition 10 that "the chance that the probability of the event lies somewhere between ... " (at the beginning of his essay, in Definition 6 Bayes says that "By chance I mean the same as probability").

gets a simple model in which uncertainty/ambiguity attitudes can be analyzed in a way that parallels the treatment of risk attitudes in the classical literature. Specifically, assume that

$$\varphi(x) = -\frac{1}{\alpha}e^{-\alpha x}$$

for $\alpha > 0$. In this case, the DM can be said to have a constant ambiguity aversion α; when $\alpha \to 0$, the DM's preferences converge to Bayesian preferences with prior $\hat{p}$, whereas when $\alpha \to \infty$, preferences converge to MMEU preferences relative to the support of μ. (See Klibanoff, Marinacci, and Mukerji, 2005, for details.) Thus, the smooth ambiguity aversion model can be viewed as an extension of the maxmin model, in its Waldean interpretation.

3.6 Variational preferences

Maccheroni, Marinacci, and Rustichini (MMR, 2006a) suggested and axiomatized an extension of the maxmin model in order to better understand the theoretical foundations of the works of Hansen and Sargent on model uncertainty in macroeconomics (see the surveys Hansen, 2007, and Hansen and Sargent, 2008). These works consider agents who take into account the possibility that their (probabilistic) model Q may not be the correct one, but only an approximation. For this reason, they rank acts f according to the following choice criterion

$$V(f) = \min_{P \in \Delta(\Sigma)} \left\{ \int_S u(f(s))\, dP(s) + \theta R(P \| Q) \right\}, \tag{26}$$

where $\theta > 0$, and $R(\cdot \| Q) : \Delta(\Sigma) \to [0, \infty]$ is the relative entropy with respect to Q.

Preferences $\succsim$ on $\mathcal{F}$ represented by criterion (26) are called *multiplier preferences* by Hansen and Sargent. The relative entropy $R(P \| Q)$ measures the relative likelihood of the alternative models P with respect to the reference model Q. The positive parameter θ reflects the weight that agents are giving to the possibility that Q might not be the correct model (as θ becomes larger, agents focus more on Q as the correct model, giving less importance to the alternatives P).

Model uncertainty, which motivated the study of multiplier preferences, is clearly akin to the problem of ambiguity, underlying maxmin preferences. Yet, neither class of preferences is nested in the other. A priori, it was not clear what are the commonalities between these models and how they can be theoretically justified. To address this issue, MMR introduced and axiomatized a novel class of preferences that includes both multiplier and maxmin preferences as special cases.

Specifically, observe that the maxmin criterion (21) can be written as

$$V(f) = \min_{P \in \Delta(\Sigma)} \left\{ \int_S u(f(s))\, dP(s) + \delta_C(P) \right\}, \tag{27}$$

where $\delta_C : \Delta \to [0, \infty]$ is the indicator function of C given by

$$\delta_C(P) = \begin{cases} 0 & \text{if } P \in C, \\ \infty & \text{otherwise.} \end{cases}$$

Like the relative entropy, the indicator function is a convex function defined on the simplex $\Delta(\Sigma)$. This suggests the following general representation

$$V(f) = \min_{P \in \Delta(\Sigma)} \left\{ \int u(f(s))\, dP(s) + c(P) \right\}, \tag{28}$$

where $c : \Delta(\Sigma) \to [0, \infty]$ is a convex function on the simplex. MMR call *variational* the preferences $\succsim$ on $\mathcal{F}$ represented by (28). Multiplier and maxmin preferences are the special cases of variational preferences where c is, respectively, the relative entropy $\theta R(\cdot \| q)$ and the indicator function δ_C.

MMR establish a behavioral foundation for the representation (28), which in turn offers a common behavioral foundation for multiplier and maxmin preferences. Their axiomatization is based on a relaxation of the C-Independence GS.3 of Gilboa and Schmeidler. To understand it, consider the following equivalent form of GS.3.

Lemma 13 *A binary relation $\succsim$ on $\mathcal{F}$ satisfies axiom GS.3 if and only if, for all $f, g \in \mathcal{F}$, $p, q \in \Delta(X)$, and $\alpha, \beta \in (0, 1]$, we have:*

$$\alpha f + (1-\alpha)p \succsim \alpha g + (1-\alpha)p \Rightarrow \beta f + (1-\beta)q \succsim \beta g + (1-\beta)q.$$

Lemma 13 (MMR p. 1454) shows that axiom GS.3 actually involves two types of independence: independence relative to mixing with constants and independence relative to the weights used in such mixing. The next axiom, due to MMR, retains the first form of independence, but not the second one.

MMR.3 WEAK C-INDEPENDENCE: If $f, g \in \mathcal{F}$, $p, q \in \Delta(X)$, and $\alpha \in (0, 1)$,

$$\alpha f + (1-\alpha)p \succsim \alpha g + (1-\alpha)p \Rightarrow \alpha f + (1-\alpha)q \succsim \alpha g + (1-\alpha)q.$$

Axiom MMR.3 is therefore the special case of axiom GS.3 in which the mixing coefficients α and β are required to be equal. In other words, axiom MMR.3 is a very weak independence axiom that requires independence only with respect to mixing with lottery acts, provided the mixing weights are kept constant.

This is a significant weakening of axiom GS.3. One might wonder, why would the DM follow MMR.3 but not GS.3 in its full strength. To see this, consider the re-statement of axiom GS.3 in Lemma 13 in the case that the weights α and β are very different, say α is close to 1 and β is close to 0. Intuitively, acts $\alpha f + (1-\alpha)p$ and $\alpha g + (1-\alpha)p$ can then involve far more uncertainty than acts $\beta f + (1-\beta)q$ and $\beta g + (1-\beta)q$, which are almost constant acts. As a result, we expect that, at least in some situations, the ranking between the genuinely uncertain acts $\alpha f + (1-\alpha)p$ and $\alpha g + (1-\alpha)p$ can well differ from that between the almost constant acts $\beta f + (1-\beta)q$ and $\beta g + (1-\beta)q$. By contrast, Axiom MMR.3 is not susceptible to this critique: since only the same coefficient α is used in both sides of the implication, the axiom does not involve acts that differ in their overall uncertainty, as it were.

The representation result of MMR is especially sharp when the utility function u is unbounded (above or below), that is, when its image $u(X) = \{u(x) : x \in X\}$ is an unbounded set. In an AA setup this follows from the following assumption (see Kopylov, 2001).

AA.7 UNBOUDEDNESS: There exist $x \succ y$ in X such that for all $\alpha \in (0,1)$ there exists $z \in X$ satisfying either $y \succ \alpha z + (1-\alpha) x$ or $\alpha z + (1-\alpha) y \succ x$.

We can now state the representation result of MMR, which generalizes Theorem 4 by allowing for general functions $c : \Delta(\Sigma) \to [0,\infty]$. Here x_f denotes the certainty equivalent of act f; i.e., $f \sim x_f$.

Theorem 14 *Let $\succsim$ be a binary relation on $\mathcal{F}$. The following conditions are equivalent:*

(i) $\succsim$ satisfies conditions AA.1, AA.2, MMR.3, AA.4, AA.5, S.6, and AA.7;

(ii) there exists an affine function $u : X \to \mathbb{R}$, with $u(X)$ unbounded, and a grounded,[68] convex, and lower semicontinuous function $c : \Delta(\Sigma) \to [0,\infty]$ such that, for all $f, g \in \mathcal{F}$

$$f \succsim g \Leftrightarrow \min_{P \in \Delta(\Sigma)} \left(\int_S u(f(s))\, dP(s) + c(P) \right) \geq \min_{p \in \Delta(\Sigma)} \left(\int u(g(s))\, dP(s) + c(P) \right). \quad (29)$$

For each u there is a unique $c : \Delta(\Sigma) \to [0,\infty]$ satisfying (29), given by

$$c(p) = \sup_{f \in \mathcal{F}} \left(u(x_f) - \int_S u(f(s))\, dP(s) \right). \quad (30)$$

MMR show how the function c can be viewed as an index of ambiguity aversion, as we will discuss later in Section 4. Alternatively, they observe that the function c can be interpreted as the cost of an adversarial opponent of selecting the prior P. In any case, formula (30) allows to determine the index c from behavioral (e.g., experimental) data in that it only requires to elicit u and the certainty equivalents x_f.

Behaviorally, maxmin preferences are the special class of variational preferences that satisfy the C-Independence axiom GS.3. For multiplier preferences, however, MMR did not provide the behavioral assumption that characterize them among variational preferences. This question left open by MMR was answered by Strzalecki (2011), who found the sought-after behavioral conditions. They turned out to be closely related to some of Savage's axioms. Strzalecki's findings thus completed the integration of multiplier preferences within the framework of choice under ambiguity.

[68] The function $c : \Delta(\Sigma) \to [0,\infty]$ is grounded if its infimum value is zero.

The weakening of C-Independence in MMR.3 has a natural variation in which independence is restricted to a particular lottery act, but not to a particular weight α. Specifically, one may require that, for the worst possible outcome x_* (if such exists),

$$\alpha f + (1-\alpha)x_* \succsim \alpha g + (1-\alpha)x_* \Leftrightarrow \beta f + (1-\beta)x_* \succsim \beta g + (1-\beta)x_*$$

for every two acts $f, g \in \mathcal{F}$ and every $\alpha, \beta \in (0,1]$,

This condition has been used by Chateauneuf and Faro (2009), alongside other conditions, to derive the following representation: there exists a so-called *confidence* function φ on $\Delta(\Sigma)$, and a confidence threshold α, such that acts are evaluated according to

$$V(f) = \min_{\{P \in \Delta(\Sigma) | \varphi(P) \geq \alpha\}} \left[\frac{1}{\varphi(P)} \int_S u(f(s)) dP(s) \right]$$

This decision rule suggests that the DM has a degree of confidence $\varphi(P)$ in each possible prior P. The expected utility associated with a prior P is multiplied by the inverse of the confidence in P, so that a low confidence level is less likely to determine the minimum confidence-weighted expected utility of f.

The intersection of the classes of variational preferences with confidence preferences is the maxmin model, satisfying C-Independence in its full force.[69] See also Ghirardato, Maccheroni, and Marinacci (2005) for other characterizations of C-Independence.

3.6.1 Beyond independence: uncertainty averse preferences

All the choice models that we reviewed so far feature some violation of the independence axiom AA.3, which is the main behavioral assumption questioned in the literature on choice under ambiguity in a AA setup. In order to better understand this class of models, Cerreia-Vioglio, Maccheroni, Marinacci, and Montrucchio (2008) recently established a common representation that unifies and classifies them. Since a notion of minimal independence among uncertain acts is, at best, elusive both at a theoretical and empirical level, this common representation does not use any independence condition on uncertain acts, however weak it may appear.

Cerreia-Vioglio et al. (2008) thus studied uncertainty averse preferences, that is, complete and transitive preferences that are monotone and convex, without any independence requirement on uncertain acts. This general class of preferences includes as special cases variational preferences, confidence preferences, as well as smooth preferences with a concave φ.

Though no independence assumption is made on uncertain acts, to calibrate risk preferences Cerreia-Vioglio et al. assumed standard independence on lottery acts.

[69] This is so because one axiom relates preferences between mixtures with different coefficients α, β and the other – between mixtures with different constant acts x_*, p.

CMMM.3 RISK INDEPENDENCE: If $p, q, r \in \Delta(X)$ and $\alpha \in (0,1)$, $p \sim q \Rightarrow \alpha p + (1-\alpha) r \sim \alpha q + (1-\alpha) r$.

Along with the other axioms, CMMM.3 implies that the risk preference $\succsim_\Delta$ satisfies the von Neumann-Morgenstern axioms. In the representation result of Cerreia-Vioglio et al. (2008) functions of the form $G : \mathbb{R}\times\Delta(\Sigma) \to (-\infty, \infty]$ play a key role. Denote by $\mathcal{G}(\mathbb{R}\times\Delta(\Sigma))$ the class of these functions such that:

(i) G is quasiconvex on $\mathbb{R}\times\Delta(\Sigma)$,

(ii) $G(\cdot, P)$ is increasing for all $P \in \Delta(\Sigma)$,

(iii) $\inf_{P\in\Delta(\Sigma)} G(t, P) = t$ for all $t \in T$.

We can now state a version of their main representation theorem.

Theorem 15 *Let $\succsim$ be a binary relation on $\mathcal{F}$. The following conditions are equivalent:*

(i) $\succsim$ satisfies axioms AA.1, AA.2, CMMM.3, AA.4, AA.5, S.6, AA.7;

(ii) there exists a non-constant affine $u : X \to \mathbb{R}$, with $u(X) = \mathbb{R}$, and a lower semicontinuous $G : \mathbb{R}\times\Delta(\Sigma) \to (-\infty, \infty]$ that belongs to $\mathcal{G}(\mathbb{R}\times\Delta(\Sigma))$ such that, for all f and g in $\mathcal{F}$,

$$f \succsim g \Leftrightarrow \min_{P\in\Delta(\Sigma)} G\left(\int u(f)\, dP, P\right) \geq \min_{P\in\Delta(\Sigma)} G\left(\int u(g)\, dP, P\right). \tag{31}$$

The function u is cardinally unique and, given u, the function G in (31) is given by

$$G(t, P) = \sup_{f\in\mathcal{F}} \left\{ u(x_f) : \int u(f)\, dP \leq t \right\} \qquad \forall (t, p) \in \mathbb{R}\times\Delta(\Sigma). \tag{32}$$

In this representation DMs can be viewed as if they considered, through the term $G\left(\int u(f)\, dP, P\right)$, all possible probabilities P and the associated expected utilities $\int u(f)\, dP$ of act f. They then behave as if they summarized all these evaluations by taking their minimum. The quasiconvexity of G and the cautious attitude reflected by the minimum in (31) derive from the convexity of preferences. Their monotonicity, instead, is reflected by the monotonicity of G in its first argument.

The representation (31) features both probabilities and expected utilities, even though no independence assumption whatsoever is made on uncertain acts. In other words, this representation establishes a general connection between the language of preferences and the language of probabilities and utilities, in keeping with the tradition of the representation theorems in choice under uncertainty.

Cerreia-Vioglio et al. (2008) show that G can be interpreted as index of uncertainty aversion, in the sense of Section 4 below. Moreover, (32) shows that this index can be elicited from choice behavior.

Variational preferences correspond to additively separable functions G, i.e., these preferences are characterized by

$$G(t, P) = t + c(P)$$

where $c : \Delta(\Sigma) \to [0, \infty]$ is a convex function. In this case (31) reduces to the variational representation (29)

Smooth preferences with concave ϕ correspond to the uncertainty aversion index given by

$$G(t, P) = t + \min_{\nu \in \Gamma(P)} I_t(\nu \parallel \mu) \tag{33}$$

where $I_t(\cdot \parallel \mu)$ is a suitable statistical distance function that generalizes the classic relative entropy, and $\Gamma(P)$ is the set of all second-order probabilities ν that are absolutely continuous with respect to μ and that have P as their reduced, first-order, probability measure on S.

3.7 Other classes of preferences

The scope of this paper does not allow us to do justice to the variety of decision models that have been suggested in the literature to deal with uncertainty in a non-probabilistic way, let alone the otherwise growing literature in decision theory.[70] Here we only mention a few additional approaches to the problem of ambiguity.

As mentioned above, Segal (1987, 1990) suggested a risk-based approach to uncertainty, founded on the idea that people do not reduce compound lotteries. Recently, Halevy (2007) provided some experimental evidence on the link between lack of reduction of compound lotteries and ambiguity, and Seo (2009) carried out an in depth theoretical analysis of this issue. Since failure to reduce compound lotteries is often regarded as a mistake, this source of ambiguity has a stronger positive flavor than the absence of information, which is our main focus.

Stinchcombe (2003), Olszewski (2007), and Ahn (2008) model ambiguity through sets of lotteries, capturing exogenous or objective ambiguity. (See also Jaffray, 1988, who suggested related ideas). Preferences are defined over these sets, with singleton and nonsingleton ones modelling risky and ambiguous alternatives, respectively. For example, these sets can be ranked either according to the criterion $V(A) = \left(\int_A \phi \circ u d\mu\right) / \mu(A)$ where ϕ and μ model ambiguity attitudes (Ahn, 2008) or the criterion $V(A) = \alpha \min_{l \in A} U(l) + (1 - \alpha) \max_{l \in A} U(l)$ where α models ambiguity attitudes (Olszewski, 2007). Viero (2009) combines this approach with the Anscombe-Aumann model.

Chateauneuf, Eichberger, and Grant (2007) axiomatize *neo-additive Choquet expected utility*, a tractable CEU criterion of the "Hurwicz" form $V(f) = \alpha \int u(f(s))\, dP(s) + \beta \max_s u(f(s)) + (1 - \alpha - \beta) \min_s u(f(s))$. Through the values of the weights α and β, the preference functional

[70] Other sub-fields include choices from menus, decision under risk, minmax regret approaches, and others. On the first of these, see Limpan and Pesendorfer (2011).

V captures in a simple way different degrees of optimism and pessimism, whose extreme forms are given by the min and max of $u(f(s))$.

Gajdos, Hayashi, Tallon, and Vergnaud (2008) axiomatize, as discussed before, a model with objective information. Preferences are defined over pairs (f, C) of acts and sets of probabilities (that represent objective information). Such pairs are ranked through the functional $V(f, C) = \min_{p \in \varphi(C)} \int u(f(s))\, dP(s)$, where $\varphi(C) \subseteq C$ is the subset of C that we denoted in the earlier discussion as D. When $\varphi(C) = C$, we get back to the MMEU model.

Gul and Pesendorfer (2008) suggested *subjective expected uncertain utility theory*, according to which acts can be reduced to bilotteries, each specifying probabilities for ranges of outcome values, where these probabilities need not be allocated to sub-ranges. Arlo-Costa and Helzner (2010a) propose to deal with the comparative ignorance hypothesis of Tversky and Fox (1995), and present experimental findings that challenge the explanation provided by the latter. (See also Arlo-Costa and Helzner, 2010b).

Siniscalchi (2009a) axiomatizes *vector expected utility*, in which Savage's acts are assessed according to $V(f) = \int u(f(s))\, dP(s) + A\left(\left(\int \xi_i \cdot u(f(s))\, dP(s)\right)_{i=1,\ldots,n}\right)$ where the first term on the right hand side is a baseline expected-utility evaluation and the second term is an adjustment that reflects DMs' perception of ambiguity and their attitudes toward it. In particular, ξ_i are random variables with zero mean that model different sources of ambiguity (see Siniscalchi, 2009a, p. 803).

Given the variety of the models of decision making that allow for non-neutral approaches to ambiguity, one is led to ask, how should we select a model to work with? There are at least three possible approaches to this problem. First, one may follow the classical empirical tradition and compare the different models by a "horse-race". The model that best explains observed phenomena should be used for prediction, with the usual trade-offs between the model's goodness of fit and its simplicity and generality. The degree to which models fit the data should be measured both for their assumptions and for their conclusions. (Indeed, the assumptions are also, in a trivial sense, conclusions.) Thus, this approach calls both for experimental tests of particular axioms and of entire models, as well as for empirical tests of theories based on these models. Importantly, when engaging in such an endeavor, one should be prepared to find that a model may be the most appropriate for analyzing certain phenomena but not for others. Thus, for example, it is possible that smooth preferences are the best model for the behavior of organizations, whereas variational preferences are a better description for the behavior of individuals. Or that labor search models are best explained by the maxmin model, while financial investments call for the Hurwicz-Jaffray model, and so forth.

For qualitative analysis, one may adopt a second approach, which does not commit to a particular model of decision under uncertainty, but uses representatives of these models in order to gain robust insights. Adopting this approach, a researcher may start with a benchmark Bayesian model, and add the uncertainty ingredient using any of the models mentioned above, as a sensitivity analysis of the Bayesian model. In this approach, theoretical convenience may be an important guideline. However, it will be advisable to trust only the qualitative conclusions that emerge from more than one model. That is, sensitivity analysis itself should not be too sensitive.

Finally, in light of the variety of models and the theoretical difficulties in selecting a single one, one may choose a third approach, which attempts to obtain general conclusions within a formal model, without committing to a particular theory of decision making. This approach has been suggested in the context of risk by Machina (1982). In this celebrated paper, facing a variety of decision models under risk, Machina attempted to show that much of economic analysis of choice under risk can be carried through without specifying a particular model. More concretely, Machina stipulated a functional on lotteries (with given probabilities) that was smooth enough to allow local approximations by linear functions. The gradient of the functional was considered to be a *local* utility function. Machina has shown that some results in economic theory could be derived by allowing the local utility function to vary, as long as it satisfied the relevant assumptions. Machina's approach was therefore not about decision theory per se. It was about the degree to which decision theory mattered: it showed that, for some applications, economists need not worry about how people really make decisions, since a wide range of models were compatible with particular qualitative conclusions.

A similar approach has been suggested for decisions under uncertainty. An early example of this approach is the notion of biseparable preferences, suggested by Ghirardato and Marinacci (2001), and mentioned above. *Biseparable preferences* are any monotone and continuous preferences over general acts that, when restricted to acts f with only two outcomes, say, x and y, can be described by the maximization of

$$J(f) = u(x)\nu(A) + (u(x) - u(y))(1 - \nu(A))$$

where ν is a capacity and

$$f(s) = \begin{cases} x & s \in A \\ y & s \notin A \end{cases}$$

with $x \succ y$. Biseparable preferences include both CEU and MMEU. Ghirardato and Marinacci (2001) provide a definition of uncertainty aversion that does not depend on the specific model of decision making and applies to all biseparable preferences.

More recently, Machina (2005) suggested a general approach to preferences under uncertainty which, similarly to Machina (1982), assumes mostly smoothness and monotonicity of preferences, but remains silent regarding the actual structure of preferences, thereby offering a highly flexible model.

4 Ambiguity aversion

Schmeidler's axiom S.6 provided a first important characterization of ambiguity aversion, modelled through a preference for hedging/randomization. Epstein (1999) and Ghirardato and Marinacci (2002) studied this issue from a different perspective, inspired by Yaari (1969)'s analysis of comparative risk attitudes.

Here we present the approach of Ghirardato and Marinacci because of its sharper model implications. This approach relies on two key ingredients:

(i) A comparative notion of ambiguity aversion that, given any two preferences $\succsim_1$ and $\succsim_2$ on $\mathcal{F}$, says when $\succsim_1$ is more ambiguity averse than $\succsim_2$.

(ii) A benchmark for neutrality to ambiguity; that is, a class of preferences $\succsim$ on $\mathcal{F}$ that are viewed as neutral to ambiguity.

The choice of these ingredients in turn determines the absolute notion of ambiguity aversion, because a preference $\succsim$ on $\mathcal{F}$ is classified as ambiguity averse provided it is more ambiguity averse than an ambiguity neutral one.

The comparative notion (i) is based on comparisons of acts with lottery acts that deliver a lottery p at all states. We consider them here because they are the most obvious example of unambiguous acts, that is, acts whose outcomes are not affected by the unknown probabilities.

Consider DM_1 and DM_2, whose preferences on $\mathcal{F}$ are $\succsim_1$ and $\succsim_2$, respectively. Suppose that

$$f \succsim_1 p,$$

that is, DM_1 prefers the possibly ambiguous act f to the unambiguous lottery act p. If DM_1 is more ambiguity averse than DM_2 it is natural to expect that DM_2 will also exhibit such preferences:

$$f \succsim_2 p.$$

For, if DM_1 is bold enough to have $f \succsim_1 p$, then DM_2 – who dislikes ambiguity no more than DM_1 – must be at least equally bold.

We take this as the behavioral characterization of the comparative notion of ambiguity aversion.

Definition 16 *Given two preferences* $\succsim_1$ *and* $\succsim_2$ *on* $\mathcal{F}$*,* $\succsim_1$ *is* more ambiguity averse than $\succsim_2$ *if, for all* $f \in \mathcal{F}$ *and* $p \in \Delta(X)$*,*

$$f \succsim_1 p \Rightarrow f \succsim_2 p. \tag{34}$$

As benchmark for neutrality to ambiguity we consider subjective expected utility (SEU) preferences on $\mathcal{F}$. These preferences intuitively embody ambiguity neutrality. They might not be the only preference embodying ambiguity neutrality, but they seem to be the most obvious ones.[71]

Methodologically, like the choice of lottery acts as the unambiguous acts in the comparison (34), also the neutrality benchmark is chosen by making the weakest prejudgment on which preferences qualify for this role. Sharp model implications will follow, nevertheless, as we will see momentarily.

Having thus prepared the ground, we can define ambiguity aversion

Definition 17 *A preference relation $\succsim$ on $\mathcal{F}$ is* ambiguity averse *if it is more ambiguity averse than some SEU preference on $\mathcal{F}$.*

The next result, due to Ghirardato and Marinacci (2002), applies these notions to the maxmin expected utility (MEU) model. Here $u_1 \approx u_2$ means that there exist $\alpha > 0$ and $\beta \in \mathbb{R}$ such that $u_1 = \alpha u_2 + \beta$.

Theorem 18 *Given any two MMEU preferences $\succsim_1$ and $\succsim_2$ on $\mathcal{F}$, the following conditions are equivalent:*

(i) $\succsim_1$ is more ambiguity averse than $\succsim_2$,

(ii) $u_1 \approx u_2$ and $C_1 \subseteq C_2$ (provided $u_1 = u_2$).

Given that $u_1 \approx u_2$, the assumption $u_1 = u_2$ is just a common normalization of the two utility indices. Therefore, Theorem 18 says that more ambiguity averse MMEU preferences are characterized, up to a normalization, by smaller sets of priors C. Therefore, the set C can be interpreted as an *index of ambiguity aversion*.

This result thus provides a behavioral foundation for the comparative statics exercises in ambiguity through the size of the sets of priors C that play a key role in the economic applications of the MMEU model. In fact, a central question in these applications is how changes in ambiguity attitudes affect the relevant economic variables.

An immediate consequence of Theorem 18 is that, not surprisingly, MMEU preferences are always ambiguity averse. That is, they automatically embody a negative attitude toward ambiguity, an attitude inherited from axiom S.6.

The condition $u_1 \approx u_2$ ensures that risk attitudes are factored out in comparing the MMEU preferences $\succsim_1$ and $\succsim_2$. This is a dividend of the risk calibration provided by the AA setup via the

[71] Epstein (1999) takes the standard for ambiguity neutrality to be preferences that are probabilistically sophisticated in the sense of Machina and Schmeidler (1992). In his approach Theorem 18 below does not hold.

risk preference $\succsim_\Delta$ discussed in Section 3.1. In a Savage setup, where this risk calibration is no longer available, Definition 16 has to be enriched in order to properly factor out risk attitudes, so that they do not interfere with the comparison of ambiguity attitudes (see Ghirardato and Marinacci, 2002, for details on this delicate conceptual issue).

Maccheroni, Marinacci, and Rustichini (2006a) generalize Theorem 18 to variational preferences by showing that the condition $C_1 \subseteq C_2$ takes in this case the more general form $c_1 \leq c_2$. The function c can thus be viewed as an index of ambiguity aversion that generalizes the sets of priors C. Variational preferences are always ambiguity averse, a fact that comes as no surprise since they satisfy axiom S.6.

For CEU preferences, Ghirardato and Marinacci (2002) show that more ambiguity averse CEU preferences are characterized, up to a common normalization of utility indexes, by smaller capacities ν. More interestingly, they show that CEU preferences are ambiguity averse when the cores of the associated capacities are nonempty. Since convex capacities have nonempty cores, CEU preferences that satisfy axiom S.6 are thus ambiguity averse. The converse, however, is not true since there are capacities with nonempty cores that are not convex. Hence, there exist ambiguity averse CEU preferences that do not satisfy S.6, which is thus a sufficient but not necessary condition for the ambiguity aversion of CEU preferences. Ghirardato and Marinacci (2002) discuss at length this feature of CEU preferences, and we refer the interested reader to that paper for details (see also Chateauneuf and Tallon, 2002, who present a notion of weak ambiguity aversion for CEU preferences, as well as Montesano and Giovannone, 1996, who investigate how CEU preferences may reflect aversion to increasing ambiguity).

Unambiguous events Unambiguous events should be events over which decision makers do not perceive any ambiguity. Intuitively, in terms of functional forms an event E is unambiguous for a preference $\succsim$ if:

(i) $\nu(E) + \nu(E^c) = 1$ when $\succsim$ is CEU;

(ii) $P(E) = P'(E)$ for all $P, P' \in C$ when $\succsim$ is MMEU and, more generally, for all $P, P' \in \operatorname{dom} c$ when $\succsim$ is variational;[72]

(iii) $p(E) = k$ μ-a.e. for some $k \in [0, 1]$ when $\succsim$ is smooth.

A few behavioral underpinnings of these notions of unambiguous event have been proposed by Nehring (1999), Epstein and Zhang (2001), Ghirardato and Marinacci (2002), Zhang (2002), Ghirardato, Maccheroni, and Marinacci (2004), Klibanoff, Marinacci, and Mukerji (2005), and Amarante

[72] $\operatorname{dom} c$ is the effective domain of the function c; i.e., $\operatorname{dom} c = \{P \in \Delta(S) : c(p) < +\infty\}$.

and Feliz (2007) (who also provide a discussion of some of the earlier notions which we refer the interested to).

5 Updating beliefs

How should one update one's beliefs when new information is obtained? In the case of probabilistic beliefs there is an almost complete unanimity that Bayes's rule is the only sensible way to update beliefs. Does it have an equivalent rule for the alternative models discussed above? The answer naturally depends on the particular non-Bayesian model one adopts. At the risk of over-generalizing from a small sample, we suggest that Bayes's rule can typically be extended to non-Bayesian beliefs in more than one way. Since the focus of this survey is on static preferences, we mention only a few examples, which by no means exhaust the richness of dynamic models.

For instance, if one's beliefs are given by a capacity ν, and one learns that an event B has obtained, one may assign to an event A the weight corresponding to the straightforward adaptations of Bayes's formula:

$$\nu\left(A|B\right) = \frac{\nu\left(A \cap B\right)}{\nu(B)}$$

However, another formula has been suggested by Dempster (1967, see also Shafer, 1976) as a special case of his notion of merging of belief functions:

$$\nu\left(A|B\right) = \frac{\nu\left(\left(A \cap B\right) \cup B^c\right) - \nu\left(B^c\right)}{1 - \nu\left(B^c\right)}$$

Clearly, this formula also boils down to standard Bayesian updating in case ν is additive. Yet, the two formulae are typically not equivalent if the capacity ν fails to be additive. Each of these formulae extends some, but not all, of the interpretations of Bayesian updating from the additive to the non-additive case.

If beliefs are given by a set of priors C, and and event B is known to have occurred, a natural candidate for the set of priors on B is simply the same set C, where each probability is updated according to Bayes rule. This results in *full Bayesian updating (FBU)*, defining the set of priors (on B)

$$C_B = \{\, p\left(\cdot|B\right) \mid p \in C\,\}$$

FBU allows standard learning given each possible prior, but does not reflect any learning about the set of priors that should indeed be taken into consideration. It captures Bayesian learning (conditional on a prior) but not the statistical inference typical of classical statistics, namely, the selection of subsets of distributions from an a priori given set of distributions. If we were to think of each prior p in C as an expert, who expresses her probabilistic beliefs, FBU can be interpreted as if each expert were learning from the evidence B, while the DM does not use the evidence to decide which experts' advice to heed.[73]

Following this line of reasoning, and in accordance with statistical principles, one may wish to select probabilities from the set C based on the given event B. One, admittedly extreme way of doing so is to adopt the maximum likelihood principle. This suggests that only the priors that a priori used to assign the highest probability to the event B should be retained among the relevant ones. Thus, *maximum likelihood updating (MLU)* is given by

$$C_B^M = \left\{ p\left(\cdot|B\right) \middle| p \in \arg\max_{q\in C} q(B) \right\}$$

If one's beliefs are given by a convex capacity, or, equivalently, by a set C which is the core of a convex capacity, MLU is equivalent to Dempster-Shafer's updating. This rule has been axiomatized by Gilboa and Schmeidler (1993), whereas FBU, suggested by Jean-Yves Jaffray, has been axiomatized by Pires (2002).

FBU and MLU are both extreme. Using the experts metaphor, FBU retains all experts, and gives as much weight to those who were right as to those who were practically proven wrong in their past assessments. By contrast, MLU completely ignores any expert who was not among the maximizers of the likelihood function. It therefore makes sense to consider intermediate methods, though, to the best of our knowledge, none has been axiomatized to date.

The tension between FBU and MLU disappears if the set of priors C is *rectangular* (Epstein and Schneider, 2003) in the sense that it can be decomposed into a set of current-period beliefs, coupled with next-period conditional beliefs, in such a way that any combination of the former and the latter is in C. Intuitively, rectangularity can be viewed as independence of sorts: it suggests that whatever happens in the present period does not teach us which prior (or expert) is to be trusted more in the next period. Formally, the set of conditional probabilities on the given event B using all priors and the set obtained using only the maximum likelihood ones coincide. Related arguments, in particular how rectangular sets of priors would lead to consistent dynamic MMEU behavior, were made by Sarin and Wakker (1998) (see in particular their Theorem 2.1). See also Epstein and Schneider (2007), who consider updating in a more explicit model, distinguishing between the set of parameters and the likelihood functions they induce.

Epstein and Schneider (2003) consider preferences over consumption processes, and axiomatize a decision rule that extends MMEU to the dynamic set-up recursively. Their axioms also guarantee that the set of priors C is rectangular. The recursive structure means that the maxmin expected utility at a given period for the entire future can also be written as maxmin expected utility over the present period and the discounted continuation (MMEU) value starting in the following period.

[73] See Seidenfeld and Wasserman (1993) who study counter-intuitive updating phenomena in this context.

Wang (2003b) proposed a related recursive approach.

This recursive approach extends beyond the MMEU model. It has similarly been applied to extend smooth preferences (see Klibanoff, Marinacci, and Mukerji, 2009, and Hayashi and Miao, 2010) and variational preferences to dynamic set-ups (see Maccheroni, Marinacci, and Rustichini, 2006b). Equipped with a variety of models of behavior with ambiguous beliefs, which are adapted to deal with dynamic problems recursively, the stage is set to analyze economic problems in not-necessarily Bayesian ways.

Another approach to updating was proposed by Hanany and Klibanoff (2007, 2009). They retain dynamic consistency by allowing the update rule to depend not only on original beliefs and new information, but also on the choice problem. In the case of the MMEU model, their approach consists of selecting a subset of priors, and updating them according to Bayes rule, while the relevant subset of priors generally depends on the act chosen before the arrival of new information.

A different route was pursued by Siniscalchi (2006b), who investigated choices over decision trees rather than over temporal acts. This modification allows him to consider sophisticated choices, characterized through a natural notion of consistent planning, under ambiguity.

An important problem relating to updating is the long-run behavior of beliefs. Suppose that a non-Bayesian decision maker faces a process that is, in a well-defined sense, repeated under the same conditions. Will she learn the true process? Will the set of probabilities converge in the limit to the true one? A partial answer was given in the context of capacities, where laws of large numbers have been proved by Marinacci (1999, 2002) and Maccheroni and Marinacci (2005). The behavior of the set of probabilities in the context of the maxmin model was analyzed in Epstein and Schneider (2007).

6 Applications

There are many economic models that lead to different qualitative conclusions when analyzed in a Bayesian way as compared to the alternative, non-Bayesian theories. The past two decades have witnessed a variety of studies that re-visited classical results and showed that they need to be qualified when one takes ambiguity into account. The scope of this paper allows us to mention but a fraction of them. The following is a very sketchy description of a few studies, designed only to give a general idea of the scope of theoretical results that need to be re-examined in light of the limitations of the Bayesian approach.[74]

Dow and Werlang (1992) analyzed a simple asset pricing model. They showed that, if an economic agent is ambiguity averse as in the CEU or MMEU model, then there will be a *range* of prices at

[74] Mukerji and Tallon (2004) survey early works in this area.

which she will wish neither to buy nor to sell a financial asset. This range will be of non-zero length even if one ignores transaction costs. To see the basic logic of this result, consider two states of the world, where the probability of the first state, p, is only known to lie in the interval $[0.4, 0.6]$. (This will also be the core of a convex capacity.) Assume that a financial asset X yields 1 in the first state and -1 in the second. The MMEU model values both X and $-X$ at -0.2. In a Bayesian model, p would be known, and the agent would switch, at a certain price π, from demanding X to offering it. This is no longer the case when p is not known. In this case, assuming ambiguity aversion, there will be an interval of prices π at which neither X nor $-X$ will seem attractive to the agent. This may explain why people refrain from trading in certain markets. It can also explain why at times of greater volatility one may find lower volumes of trade: with a larger set of probabilities that are considered possible, there will be more DMs who prefer neither to buy nor to sell.[75] The question of trade among uncertainty averse agents has been also studied in Billot, Chateauneuf, Gilboa and Tallon (2000), Kajii and Ui (2006, 2009), and Rigotti, Shannon, and Strzalecki (2008).

Epstein and Miao (2003) use uncertainty aversion to explain the home bias phenomenon in international finance, namely, the observation that people prefer to trade stocks of their own country rather than foreign ones. The intuition is that agents know the firms and the stock market in their own country better than in foreign ones. Thus, there is more ambiguity about foreign equities than about domestic ones. A Bayesian analysis makes it more difficult to explain this phenomenon: when a Bayesian DM does not know the distribution of the value of a foreign equity, she should have beliefs over it, reducing uncertainty to risk. Thus, a Bayesian would behave in qualitatively similar ways when confronting known and unknown distributions. By contrast, the notion that agents are ambiguity averse may more readily explain why they prefer to trade when the value distribution is closer to being known than when there is a great deal of ambiguity about it.

There are many other applications of ambiguity aversion to models of asset pricing. For example, Epstein and Schneider (2008) show that models involving ambiguity can better capture market reaction to the quality of information than can Bayesian models (see also Epstein and Schneider, 2010), while Gollier (2011) shows that ambiguity aversion may not reinforce risk aversion and investigates how this may affect asset prices. Other recent asset pricing applications include Garlappi, Uppal and Wang (2007), Caskey (2009), Miao (2009), Ju and Miao (2010), Miao and Wang (2011) (see also Guidolin and Rinaldi, 2011).

The MMEU model has also been employed in a job search model by Nishimura and Ozaki (2004). They ask how an unemployed agent will react to increasing uncertainty in the labor market. In a Bayesian model, greater uncertainty might be captured by higher variance of the job offers that the

[75] This argument assumes that the decision maker starts with a risk-free portfolio. A trader who already holds an uncertain position may be satisfied with it with a small set of probabilities, but wish to trade in order to reduce uncertainty if the set of probabilities is larger.

agent receives. Other things being equal, an increase in variance should make the agent less willing to accept a given offer, knowing that he has a chance to get better ones later on. This conclusion is a result of the assumption that all uncertainty is quantifiable by a probability measure. Nishimura and Ozaki (2004) show that for an ambiguity averse agent, using the MMEU model, the conclusion might be reversed: in the presence of greater uncertainty, modeled as a larger set of possible priors, the agent will be more willing to take a given job offer rather than bet on waiting for better ones in the future.

Hansen, Sargent, and Tallarini (1999) and Hansen, Sargent, and Wang (2002) compare savings behavior under expected utility maximization with savings behavior of a *robust DM* who behaves in accordance with the multiple prior model. They show that the behavior of a robust DM puts the market price of risk much closer to empirical estimates than does the behavior of the classical expected utility maximizer, and, in particular, can help account for the equity premium. Hansen and Sargent (2001, 2008) apply multiplier preferences to macroeconomic questions starting from the viewpoint that, whatever the probability model a policy maker might have, it cannot be known with certainty. They ask how robust economic policy would be to variations in the underlying probability, and find conclusions that differ qualitatively from classical results. See also Miao (2004), who studies the consumption-savings decision in a different set-up.

Other (published) applications of ambiguity averse preferences include Epstein and Wang (1994, 1995), who explain financial crashes and booms, Mukerji (1998), who explains incompleteness of contracts, Chateauneuf, Dana, and Tallon (2000), who study optimal risk-sharing rules with ambiguity averse agents, Greenberg (2000), who finds that in a strategic set-up a player may find it beneficial to generate ambiguity about her strategy choice, Mukerji and Tallon (2001), who show how incompleteness of financial markets my arise because of ambiguity aversion, Rigotti and Shannon (2005), who characterize equilibria and optima and study how they depend on the degree of ambiguity, Bose, Ozdenoren and Pape (2006), who study auctions under ambiguity, Nishimura and Ozaki (2007), who show that an increase in ambiguity changes the value of an investment opportunity differently than does an increase in risk, Easley and O'Hara (2009, 2010), who study how ambiguity affects market participation, and Treich (2010), who studies when the value of a statistical life increases under ambiguity aversion.

As mentioned above, this list is but a sample of applications and has no claim even to be a representative sample.

7 Conclusion

Uncertainty is present in practically every field of economic enquiry. Problems in growth and finance, labor and development, political economy and industrial organization lead to questions of uncertainty and require its modeling.

For the most part, economic theory has strived to have a unifying approach to decision making in general, and to decision under uncertainty in particular. It is always desirable to have simple, unifying principles, especially if, as is the case with expected utility theory, these principles are elegant and tractable.

At the same time, expected utility theory appears to be too simple for some applications. Despite its considerable generality, there are phenomena that are hard to accommodate with the classical theory. Worse still, using the classical theory alone may lead to wrong qualitative conclusions, and may make it hard to perceive certain patterns of economic behavior that may be readily perceived given the right language.

At this point it is not clear whether a single paradigm of decision making under uncertainty will ever be able to replace the Bayesian one. It is possible that different models will prove useful to varying degrees in different types of problems. But even if a single paradigm will eventually emerge, it is probably too soon to tell which one it will be.

For the time being, it appears that economic theory may benefit from having more than a single theory of decision under uncertainty in its toolbox. The Bayesian model is surely a great candidate to remain the benchmark. Moreover, often it is quite obvious that the insights learned from the Bayesian analysis suffice. For example, Akerlof's (1970) lemons model need not be generalized to incorporate ambiguity. Its insight is simple and clear, and it will survive in any reasonable model. But there are other models in which the Bayesian analysis might be misleadingly simple. In some cases, adding a touch of ambiguity to the model, often in whatever model of ambiguity one fancies, suffices to change the qualitative conclusions. Hence it seems advisable to have models of ambiguous beliefs in our toolbox, and to test each result, obtained under the Bayesian assumptions, for robustness relative to the presence of ambiguity.

8 References

Ahn, D. (2008), "Ambiguity without a State Space", *Review of Economic Studies*, 75: 3-28.

Akerlof, G. A. (1970), "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism", *The Quarterly Journal of Economics*, 84: 488-500.

Al-Najjar, N. and J. L. Weinstein (2009) "The Ambiguity Aversion Literature: A Critical Assessment", *Economics and Philosophy*, 25: 249-284.

Alon, S., and D. Schmeidler (2010), "Purely Subjective Maximin Expected Utility", Tel Aviv, mimeo.

Amarante, M. (2009), "Foundations of Neo-Bayesian Statistics", *Journal of Economic Theory*, 144: 2146-2173.

Amarante, M., and E. Feliz (2007), "Ambiguous Events and Maxmin Expected Utility", *Journal of Economic Theory*, 134: 1-33.

Anscombe, F. J. and R. J. Aumann (1963), "A Definition of Subjective Probability", *Annals of Mathematics and Statistics*, 34: 199-205.

Arlo-Costa, H. and J. Helzner (2010a), "Ambiguity Aversion: The Explanatory Power of Indeterminate Probabilities", *Synthese*, 172: 37-55.

——— (2010b), "Ellsberg Choices: Behavioral Anomalies or New Normative Insights?" *Philosophy of Science,* forthcoming.

Arrow, K. J. (1951), "Alternative Approaches to the Theory of Choice in Risk-Taking Situations", *Econometrica,* 19: 404-437.

——— (1970), *Essays in the Theory of Risk-Bearing.* Amsterdam: North-Holland.

——— (1980), "Review of 'Foundations: Essays in Philosophy, Logic, Mathematics and Economics',", *Journal of Political Economy*, 88: 636-638.

Arrow, K.J., and L. Hurwicz. (1972), "An Optimality Criterion for Decision Making under Ignorance", In C.F. Carter and J.L. Ford, *Uncertainty and Expectations in Economics.* Basil Blackwell: Oxford.

Artzner, P., F. Delbaen, J.M. Eber, and D. Heath (1997), "Thinking Coherently", *Risk,* 10: 68-71

——— (1999), "Coherent Measures of Risk", *Mathematical Finance,* 9: 203-228

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

——— (1974), “Subjectivity and Correlation in Randomized Strategies”, *Journal of Mathematical Economics,* 1: 67-96.

——— (1976), “Agreeing to Disagree”, *Annals of Statistics,* 4: 1236-1239.

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

Bather, J. (1996), “A Conversation with Herman Chernoff”, *Statistical Science*, 11: 335-350.

Bayarri, M. J., and J. O. Berger (2004), “The Interplay of Bayesian and Frequentist Analysis”, *Statistical Science*, 19: 58-80.

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

Berger, J. (2004), “The Case for Objective Bayesian Analysis”, *Bayesian Analysis*, 1: 1-17.

Bernoulli, D. (1738), “Specimen theoriae novae de mensura sortis”, *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, (translated as “Exposition of a new theory on the measurement of risk”, *Econometrica*, 22, 23-36, 1954).

Bernoulli, J. (1713), *Ars Conjectandi.* Thurneysen Brothers: Basel (translated by E. D. Sylla in “The Art of Conjecturing”, Johns Hopkins University Press, 2005).

Bertrand J. (1907), *Calcul de probabilite.* Gauthiers Villars: Paris (2^{nd} edition).

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

Billot, A. Chateauneuf, I. Gilboa and J.-M. Tallon (2000), “Sharing Beliefs: Between Agreeing and Disagreeing”, *Econometrica*, 68: 685-694.

Bose, S., Ozdenoren, E., and A. Pape (2006), “Optimal Auctions with Ambiguity”, *Theoretical Economics*, 1: 411-438.

Bridgman, P. W. (1927), *The Logic of Modern Physics.* New York: Macmillan.

Carnap, R. (1923), “Uber die Aufgabe der Physik und die Andwednung des Grundsatze der Einfachstheit”, *Kant-Studien*, 28: 90-107.

Casadesus-Masanell, R., P. Klibanoff, and E. Ozdenoren (2000), “Maxmin Expected Utility over Savage Acts with a Set of Priors”, *Journal of Economic Theory*, 92: 35-65.

Caskey, J. (2009), “Information in equity markets with ambiguity-averse investors”, *Review of Financial Studies*, 22: 3595-3627.

Castagnoli, E., F. Maccheroni, and M. Marinacci (2003), "Expected utility with multiple priors", in *Proceedings of ISIPTA 2003*.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2008), "Uncertainty Averse Preferences", *Journal of Economic Theory,* forthcoming.

Chateauneuf, A., R.-A. Dana, and J.-M. Tallon (2000), "Optimal risk-sharing rules and equilibria with Choquet expected utility", *Journal of Mathematical Economics*, 34: 191-214.

Chateauneuf, A., J. Eichberger, and S. Grant (2007), "Choice under uncertainty with the best and worst in mind: Neo-additive capacities", *Journal of Economic Theory*, 137: 538-567.

Chateauneuf, A., F. Maccheroni, M. Marinacci, and J.-M. Tallon (2005), "Monotone continuous multiple priors," *Economic Theory*, 26: 973-982.

Chateauneuf, A. and J. H. Faro (2009), "Ambiguity Through Confidence Functions", *Journal of Mathematical Economics*, 45: 535-558.

Chateauneuf, A., and J.-M. Tallon (2002), "Diversification, convex preferences, and non-empty core in the Choquet expected utility model", *Economic Theory*, 19: 509-523.

Chew, H. S., and J. Sagi (2008), Small Worlds: Modeling Attitudes toward Sources of Uncertainty, *Journal of Economic Theory*, 139: 1-24.

Choquet, G. (1953), "Theory of Capacities", *Annales de l'Institut Fourier,* 5: 131-295.

Cifarelli, D. M. and E. Regazzini (1996), "de Finetti's Contribution to Probability and Statistics", *Statistical Science*, 11: 253-282.

Cyert, R. M., and M. H. DeGroot (1974), "Rational Expectations and Bayesian Analysis", *Journal of Political Economy*, 82: 521-536.

Daston, L. (1995), *Classical Probability in the Enlightenment.* Princeton: Princeton University Press.

de Finetti, B. (1931), "Sul Significato Soggettivo della Probabilità", *Fundamenta Mathematicae*, 17: 298-329.

——— (1937), "La Prevision: ses Lois Logiques, ses Sources Subjectives", *Annales de l'Institut Henri Poincare*, 7: 1-68 (translated in *Studies in Subjective Probability*, edited by H.E. Kyburg and H.E. Smokler, Wiley, 1963).

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

Dekel, E., B. L. Lipman, and A. Rustichini (2001), "Representing Preferences with a Unique Subjective State Space", *Econometrica*, 69: 891-934.

Dellacherie, C. (1971), “Quelques commentaires sur le prolongements de capacités”, *Lecture Notes in Math.* 191. New York: Springer-Verlag.

Dempster, A. P. (1967). “Upper and Lower Probabilities Induced by a Multivalued Mapping”, *Annals of Mathematical Statistics,* 38: 325-339.

Denneberg D. (1994), *Non-Additive Measure and Integral.* Dordrecht: Kluwer.

Diaconis, P., and D. Freedman (1986), On the Consistency of Bayes Estimates, *Annals of Statistics*, 14: 1-26

Dow, J., and S. R. C. Werlang (1992), “Uncertainty aversion, risk aversion, and the optimal choice of portfolio”, *Econometrica*, 60: 197-204.

Dreze, J. H. (1961), “Les Fondements Logiques de l’Utilite Cardinale et de la Probabilite Subjective”, *La Decision.* Colloques Internationaux du CNRS.

——— (1987), *Essays on Economic Decision under Uncertainty.* Cambridge, UK: Cambridge University Press.

Easley, D., and M. O’Hara (2009), “Ambiguity and Nonparticipation: The Role of Regulation”, *Review of Financial Studies* 22: 1817-1843.

——— (2010), “Microstructure and ambiguity”, *Journal of Finance*, 65: 1817-1846.

Eichberger, J., S. Grant, and D. Kelsey (2008), “Differentiating Ambiguity: an Expository Note”, *Economic Theory*, 36: 327-336.

Eichberger, J., S. Grant, D. Kelsey and G. A. Koshevoy (2011), “The α-MEU Model: A Comment”, *Journal of Economic Theory*, forthcoming.

Efron, B. (1986), “Why Isn’t Everyone a Bayesian?”, *The American Statistician*, 40: 1-11 (with discussion).

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

Epstein, L. (1999), “A Definition of Uncertainty Aversion”, *Review of Economic Studies*, **66**: 579-608.

Epstein, L. G. and M. Marinacci (2007), “Mutual Absolute Continuity of Multiple Priors”, *Journal of Economic Theory*, 137: 716-720.

Epstein, L. G., M. Marinacci, and K. Seo (2007), “Coarse Contingencies and Ambiguity”, *Theoretical Economics*, 2: 355-394.

Epstein, L. G. and J. Miao (2003), “A Two-Person Dynamic Equilibrium under Ambiguity”, *Journal of Economic Dynamics and Control*, 27: 1253-1288.

Epstein, L. G. and M. Schneider (2003), “Recursive Multiple Priors”, *Journal of Economic Theory*, 113: 32-50.

——— (2007), “Learning under Ambiguity”, *Review of Economic Studies*, 74: 1275-1303.

——— (2008),“Ambiguity, Information Quality and Asset Pricing”, *Journal of Finance,* 63: 197-228.

——— (2010), “Ambiguity and Asset Markets”, *Annual Review of Financial Economics*, 2: 315-346.

Epstein, L. G. and T. Wang (1994), “Intertemporal Asset Pricing under Knightian Uncertainty”, *Econometrica*, 62: 283-322.

——— (1995), “Uncertainty, Risk-Neutral Measures and Security Price Booms and Crashes”, *Journal of Economic Theory*, 67: 40-82.

Epstein, L. G. and J. Zhang, (2001), “Subjective Probabilities on Subjectively Unambiguous Events”, *Econometrica*, 69: 265-306.

Ergin, H. and F. Gul (2009), “A Theory of Subjective Compound Lotteries”, *Journal of Economic Theory*, 144: 899-929.

Ergin, H. and T. Sarver (2009), “A Subjective Model of Temporal Preferences”, Northwestern and WUSTL, mimeo.

Fischhoff, B., and W. Bruine De Bruin (1999), “Fifty–Fifty=50%?”, *Journal of Behavioral Decision Making*, 12: 149-163.

Fishburn, P.C. (1970) *Utility Theory for Decision Making.* John Wiley and Sons.

Frisch, R. (1926), “Sur un problème d'économie pure”, *Norsk Matematisk Forenings Skrifter*, 1: 1-40.

Gajdos, T., T. Hayashi, J.-M. Tallon, and J.-C. Vergnaud (2008), “Attitude toward Imprecise Information”, *Journal of Economic Theory*, 140: 27-65.

Gärdenfors, P. and N.-E. Sahlin (1982), “Unreliable probabilities, risk taking, and decision making”, *Synthese*, 53:361-386.

Garlappi, L., R. Uppal, and T. Wang, (2007), “Portfolio selection with parameter and model uncertainty: a multi-prior approach”, *Review of Financial Studies*, 20: 41-81.

Ghirardato, P. (2002), “Revisiting Savage in a Conditional World”, *Economic Theory*, 20: 83-92.

Ghirardato, P., P. Klibanoff, and M. Marinacci (1998), “Additivity with Multiple Priors”, *Journal of Mathematical Economics*, 30: 405-420.

Ghirardato, P. and M. Marinacci (2001), "Risk, Ambiguity, and the Separation of Utility and Beliefs", *Mathematics of Operations Research*, 26: 864–890.

——— (2002), "Ambiguity Made Precise: A Comparative Foundation", *Journal of Economic Theory*, 102: 251–289.

Ghirardato, P., F. Maccheroni, and M. Marinacci (2004), "Differentiating Ambiguity and Ambiguity Attitude", *Journal of Economic Theory*, 118: 133-173.

——— (2005), "Certainty Independence and the Separation of Utility and Beliefs", *Journal of Economic Theory*, 120: 129-136.

Ghirardato, P., F. Maccheroni, M. Marinacci, and M. Siniscalchi (2003), "Subjective Foundations for Objective Randomization: A New Spin on Roulette Wheels", *Econometrica* 71: 1897-1908.

Gilboa, I. (1987), "Expected Utility with Purely Subjective Non-Additive Probabilities", *Journal of Mathematical Economics*, 16: 65-88.

——— (2009), *Theory of Decision under Uncertainty*, 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 (2008), "Probabilities in Economic Modeling", *Journal of Economic Perspectives*, 22: 173-188.

——— (2009), "Is It Always Rational to Satisfy Savage's Axioms?", *Economics and Philosophy*, forthcoming.

——— (2010), "Rationality of Belief Or: Why Savage's axioms are neither necessary nor sufficient", *Synthese*, forthcoming.

Gilboa, I. and D. Schmeidler (1989), "Maxmin Expected Utility with a Non-Unique Prior", *Journal of Mathematical Economics*, 18: 141-153. (Working paper, 1986.)

——— (1993), "Updating Ambiguous Beliefs", *Journal of Economic Theory*, 59: 33-49.

Giraud, R. (2005) "Objective Imprecise Probabilistic Information, Second Order Beliefs and Ambiguity Aversion : An Axiomatization" in *Proceedings of ISIPTA 2005*.

Gollier, C., (2011), "Does Ambiguity Aversion Reinforce Risk Aversion? Applications to Portfolio Choices and Asset Pricing", *Review of Economic Studies*, forthcoming.

Greenberg, J. (2000), "The Right to Remain Silent", *Theory and Decisions*, 48: 193-204.

Guidolin, M., and F. Rinaldi (2011), "Ambiguity in Asset Pricing and Portfolio Choice: a Review of the Literature", IGIER-Bocconi, mimeo.

Gul, F. and W. Pesendorfer (2008), "Measurable Ambiguity", Princeton, mimeo.

Hacking, I. (1975), *The Emergence of Probability.* Cambridge: Cambridge University Press.

Halevy, Y. (2007), "Ellsberg Revisited: An Experimental Study", *Econometrica*, 75: 503-536.

Halevy, Y. and V. Feltkamp (2005), "A Bayesian Approach to Uncertainty Aversion", *Review of Economic Studies*, 72: 449-466.

Hanany, E. and P. Klibanoff (2007), "Updating Preferences with Multiple Priors", *Theoretical Economics*, 2: 261-298.

——— (2009), "Updating Ambiguity Averse Preferences", *The B.E. Journal of Theoretical Economics,* 9 (Advances): Article 37.

Hansen, L. P. (2007), "Beliefs, Doubts, and Learning: Valuing Macroeconomic Risk", *American Economic Review*, 97: 1-30.

Hansen, L. P., and T. J. Sargent (2001), Robust control and model uncertainty, *American Economic Review*, 91: 60-66.

——— (2008), *Robustness.* Princeton: Princeton University Press.

Hart, S. Modica, S. and D. Schmeidler (1994), "A Neo^2 Bayesian Foundation of the Maxmin Value for Two-Person Zero-SUM Games, *International Journal of Game Theory,* 23: 347-358.

Hayashi, T., and J. Miao (2010), "Intertemporal Substitution and Recursive Smooth Ambiguity Preferences", *Theoretical Economics*, forthcoming.

Hurwicz, L. (1951), "Some Specification Problems and Application to Econometric Models", *Econometrica,* 19: 343-344.

Huygens, C. (1657), *De Ratiociniis in Ludo Aleae.* Amsterdam: van Schooten (translated by E. D. Sylla in "The Art of Conjecturing", Johns Hopkins University Press, 2005).

Jaffray, J.-Y. (1988), "Application of Linear Utility Theory to Belief Functions", in *Uncertainty and Intelligent Systems*: 1-8. Berlin: Springer.

Jeffrey, R. C. (1965). *The Logic of Decision.* New-York: McGraw-Hill.

Ju, N., and J. Miao (2010), "Ambiguity, Learning, and Asset Returns", Boston University, mimeo.

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

Kajii, A. and T. Ui (2006), "Agreeable Bets with Multiple Priors", *Journal of Economic Theory* 128: 299-305.

——— (2009), "Interim Efficient Allocations under Uncertainty", *Journal of Economic Theory*, 144: 337-353.

Kass, R. E. and L. Wasserman (1996), "The Selection of Prior Distributions by Formal Rules," *Journal of the American Statistical Association*, 91:1343-1370.

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*, XIV: 109–123.

Klibanoff, P. (2001a) "Stochastically independent randomization and uncertainty aversion", *Economic Theory*, 18: 605-620.

——— (2001b), "Characterizing uncertainty aversion through preference for mixtures", *Social Choice and Welfare*, 18: 289-301.

Klibanoff, P., M. Marinacci, and S. Mukerji (2005), "A Smooth Model of Decision Making under Ambiguity," *Econometrica*, 73: 1849-1892.

——— (2009) "Recursive Smooth Ambiguity Preferences", *Journal of Economic Theory*, 144: 930-976.

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

Kobberling, V. and P. P. Wakker (2003), "Preference Foundations for Nonexpected Utility: A Generalized and Simplified Technique", *Mathematics of Operations Research,* 28: 395-423.

Kocherlakota, N. R. (2007), "Model Fit and Model Selection", *Federal Reserve Bank of St. Louis Review*, 89: 349-360.

Koopman , B. O. (1940a), "The Axioms and Algebra of Intuitive Probability", *Annals of Mathematics*, 41: 269-292.

——— (1940b), "The Bases of Probability", *Bulletin of the American Mathematical Society*, 46: 763-774.

——— (1941), "Intuitive Probabilities and Sequences", *Annals of Mathematics*, 42: 169-187.

Koopman, B. O., and J. von Neumann (1932), "Dynamical Systems of Continuous Spectra", *Proceedings of the National Academy of Science,* 18: 255-263.

Kopylov, I. (2001), "Procedural Rationality in the Multiple Prior Model", Rochester, mimeo.

——— (2007), “Subjective Probabilities on ‘Small’ Domains”, *Journal of Economic Theory*, 133: 236-265.

——— (2010), “Simple Axioms for Countably Additive Subjective Probability”, UC Irvine, mimeo.

Kraft, C. H., J. W. Pratt, and A. Seidenberg (1959), “Intuitive Probability on Finite Sets”, *Annals of Mathematical Statistics*, 30: 408-419.

Kreps, D. (1988), *Notes on the Theory of Choice.* Westview Press: Underground Classics in Economics.

Kyburg, H. E. and H. E. Smokler (1964), *Studies in Subjective Probability.* New York: John Wiley and Sons.

Laplace, P. S., (1814), *Essai Philophique sur les Probabilites*, Gauthier-Villars: Paris (English edition, 1951, A Philosophical Essay on Probabilities, New York: Dover).

Levi, I. (1974), “On Indeterminate Probabilities”, *Journal of Philosophy* 71: 391-418.

——— (1980), *The Enterprise of Knowledge.* Cambridge, MA: MIT Press.

Lewis, D. (1980), “A Subjectivist’s Guide to Objective Chance”, in *Studies in Inductive Logic and Probability* (R. C. Jeffrey, ed.), Berkeley and Los Angeles: University of California Press.

Lipman, B., and W. Pesendorfer (2011), “Temptation”, this volume.

Maccheroni, F., and M. Marinacci (2005), “A Strong Law of Large Numbers for Capacities”, *The Annals of Probability*, 33: 1171–1178.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006a), “Ambiguity Aversion, Robustness, and the Variational Representation of Preferences”, *Econometrica*, 74: 1447-1498.

——— (2006b), “Dynamic Variational Preference”, *Journal of Economic Theory* 128: 4-44.

Machina, M. J. (1982), “‘Expected Utility’ Analysis without the Independence Axiom”, *Econometrica*, 50: 277-323.

——— (1987), “Choice Under Uncertainty: Problems Solved and Unsolved”, *Economic Perspectives*, 1: 121-154.

——— (2004), “Almost-Objective Uncertainty”, *Economic Theory* 24: 1-54.

——— (2005), “ ‘Expected Utility/Subjective Probability’ Analysis without the Sure-Thing Principle or Probabilistic Sophistication", *Economic Theory*, 26: 1-62.

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

Marinacci, M. (1999), "Limit Laws for Non-Additive Probabilities and Their Frequentist Interpretation", *Journal of Economic Theory*, 84: 145–195.

——— (2002a), "Probabilistic Sophistication and Multiple Priors", *Econometrica* 70: 755-764.

——— (2002b), "Learning from Ambiguous Urns", *Statistical Papers* 43: 143-151.

Marinacci, M. and L. Montrucchio (2004), Introduction to the mathematics of ambiguity, in *Uncertainty in Economic Theory*, (I. Gilboa, ed.), New York: Routledge.

Miao, J. (2004), "A Note on Consumption and Savings Under Knightian Uncertainty", *Annals of Economics and Finance*, 5: 299-311.

——— (2009), "Ambiguity, Risk and Portfolio Choice under Incomplete Information", *Annals of Economics and Finance*, 10: 257-279.

Miao, J. and N. Wang (2011), "Risk, Uncertainty, and Option Exercise", *Journal of Economic Dynamics and Control*, forthcoming.

Milgrom, P. and N. Stokey (1982), "Information, Trade, and Common Knowledge", *Journal of Economic Theory*, 26: 17-27.

Milnor, J. (1954), "Games Against Nature", in: *Decision Processes*, ed. R.M. Thrall, C.H. Coombs & R.L. Davis. New York, Wiley.

Montesano, A. and F. Giovannone (1996), "Uncertainty aversion and aversion to increasing uncertainty", *Theory and Decision*, 41: 133-148.

Mukerji, S. (1998), "Ambiguity Aversion and the Incompleteness of Contractual Form", *American Economic Review*, 88: 1207-1232.

——— (2009), "Foundations of ambiguity and economic modelling", *Economics and Philosophy*, 25: 297-302.

Mukerji, S., and J.-M. Tallon (2001), "Ambiguity aversion and incompleteness of financial markets", *Review of Economic Studies*, 68: 883-904.

——— (2004), "An Overview of Economic Applications of David Schmeidler's Models of Decision Making under Uncertainty, in *Uncertainty in Economic Theory*, (I. Gilboa, ed.), New York: Routledge.

Nakamura, Y. (1990), "Subjective expected utility with non-additive probabilities on finite state spaces". *Journal of Economic Theory,* 51: 346–366.

Nau, R.F. (2001, 2006), "Uncertainty Aversion with Second-Order Utilities and Probabilities", *Management Science* 52: 136-145 (see also *Proceedings of ISIPTA 2001*).

——— (2010). "Risk, Ambiguity, and State-Preference Theory", *Economic Theory*, forthcoming.

Nehring, K. (1999), "Capacities and Probabilistic Beliefs: a Precarious Coexistence", *Mathematical Social Sciences*, 38: 197-213.

——— (2009), "Imprecise Probabilistic Beliefs as a Context for Decision Making under Ambiguity", forthcoming in *Journal of Economic Theory* 144*:* 1054-1091.

Nishimura, K., and H. Ozaki (2004), "Search and Knightian Uncertainty", *Journal of Economic Theory*, 119: 299-333.

——— (2007), "Irreversible Investment and Knightian Uncertainty", *Journal of Economic Theory*, 136: 668-694.

Olszewski, W. B. (2007), "Preferences over Sets of Lotteries", *Review of Economic Studies*, 74: 567-595

Ore, O. (1960), "Pascal and the Invention of Probability Theory," *American Mathematical Monthly*, 67: 409-419.

Ortoleva, P. (2010), "Status Quo Bias, Multiple Priors and Uncertainty Aversion", *Games and Economic Behavior,* 69: 411-424.

Ozdenoren, E., and J. Peck (2008), "Ambiguity Aversion, Games against Nature, and Dynamic Consistency", *Games and Economic Behavior*, 62: 106-115.

Pascal, B. (1670), *Pensées sur la Religion et sur Quelques Autres Sujets*.

Pearl, J. (1986), "Fusion, Propagation, and Structuring in Belief Networks", *Artificial Intelligence*, 29: 241–288.

Pires, C. P. (2002), "A Rule for Updating Ambiguous Beliefs", *Theory and Decision*, 33: 137-152.

Quiggin, J. (1982), "A Theory of Anticipated Utility", *Journal of Economic Behaviorand Organization*, 3: 225-243.

Ramsey, F. P. (1926a), "Truth and Probability", in R. Braithwaite (ed.), (1931), *The Foundation of Mathematics and Other Logical Essays*. London: Routledge and Kegan.

——— (1926b), "Mathematical Logic", *Mathematical Gazette*, 13: 185-194.

Rigotti, L., and C. Shannon (2005), "Uncertainty and Risk in Financial Markets", *Econometrica*, 73: 203-243.

Rigotti, L., C. Shannon, and T. Strzalecki (2008), "Subjective Beliefs and ex ante Trade", *Econometrica*, 76: 1167-1190.

Rosenmueller, J. (1971), “On Core and Value,” *Methods of Operations Research,* 9: 84-104.

——— (1972), “Some Properties of Convex Set Functions, Part II” *Methods of Operations Research*, 17: 287-307.

Sarin, R., and P. P. Wakker, (1992), “A Simple Axiomatization of Nonadditive Expected Utility”, *Econometrica*, 60: 1255-1272.

——— (1998), “Dynamic Choice and Nonexpected Utility”, *Journal of Risk and Uncertainty*, 17: 87-119.

Savage, L. J. (1951), “The Theory of Statistical Decision,” *Journal of the American Statistical Association*, 46: 55-67.

——— (1954), *The Foundations of Statistics.* New York: John Wiley and Sons. (Second edition in 1972, Dover)

Schervish, M. J. (1995), *Theory of Statistics*, New York: Springer-Verlag.

Schmeidler, D. (1986), “Integral Representation without Additivity”, *Proceedings of the American Mathematical Society*, 97: 255-261.

——— (1989), “Subjective Probability and Expected Utility without Additivity”, *Econometrica*, 57: 571-587. (Working paper, 1982).

Seo, K. (2009), “Ambiguity and Second-Order Belief”, *Econometrica*, **77**: 1575-1605.

Segal, U. (1987), “The Ellsberg Paradox and Risk Aversion: An Anticipated Utility Approach”, *International Economic Review*, 28: 175–202.

——— (1990).“Two-Stage Lotteries without the Reduction Axiom”, *Econometrica*, 58: 349–377.

Seidenfeld, T. and L. Wasserman (1993), “Dilation for Sets of Probabilities”, *The Annals of Statistics*, 21: 1139-1154.

Shafer, G. (1976), *A Mathematical Theory of Evidence.* Princeton: Princeton University Press.

——— (1986), “Savage Revisited”, *Statistical Science*, 1: 463-486.

Shapley, L. S. (1972), “Cores of Convex Games”, *International Journal of Game Theory*, 1: 11-26. (Working paper, 1965).

Siniscalchi, M. (2006a), “A behavioral characterization of plausible priors”, *Journal of Economic Theory*, 128: 91-135.

——— (2006b), “Dynamic choice under ambiguity”, *Theoretical Economics*, forthcoming.

——— (2009a), "Vector Expected Utility and Attitudes toward Variation", *Econometrica*, 77: 801-855.

——— (2009b), "Two out of three ain't bad: a comment on 'The ambiguity aversion literature: A critical assessment'" *Economics and Philosophy*, 25: 335-356.

Smith, C. A. B. (1961), "Consistency in Statistical Inference and Decision", *Journal of the Royal Statistical Society, Series B*, 23: 1-25.

Stinchcombe, M. (2003), "Choice and games with ambiguity as sets of probabilities", UT Austin, mimeo.

Strzalecki, T. (2011), "Axiomatic Foundations of Multiplier Preferences", *Econometrica*, 79: 47-73.

Suppe, F. (1977), *The structure of scientific theories.* Champaign: University of Illinois Press.

Treich, N. (2010), "The value of a statistical life under ambiguity aversion", *Journal of Environmental Economics and Management*, 59: 15-26.

Tversky, A., and C. Fox (1995), "Weighing Risk and Uncertainty", *Psychological Review*, 102: 269–283.

Tversky, A., and D. Kahneman (1992), "Advances in Prospect Theory: Cumulative Representation of Uncertainty", *Journal of Risk and Uncertainty*, 5: 297-323.

van Fraassen, B. C. (1989), *Laws and Symmetry.* Oxford: Oxford University Press.

Viero, M.-L. (2009) "Exactly What Happens After the Anscombe-Aumann Race? Representing Preferences in Vague Environments", *Economic Theory,* 41: 175-212.

von Neumann, J. and O. Morgenstern (1947), *Theory of Games and Economic Behavior.* Princeton: Princeton University Press (2^{nd} edition).

Wakker, P. P. (1989a), Continuous Subjective Expected Utility with Nonadditive Probabilities, *Journal of Mathematical Economics*, 18: 1-27.

——— (1989b), *Additive Representations of Preferences: a New Foundation of Decision Analysis.* Dordrecht: Kluwer.

——— (1990), "Characterizing Optimism and Pessimism Directly through Comonotonicity", *Journal of Economic Theory*, 52: 453-463.

——— (1991), "Testing and Characterizing Properties of Nonadditive Measures through Violations of the Sure-Thing Principle", *Econometrica*, 69: 1039-1059.

——— (2010), *Prospect Theory.* Cambridge: Cambridge University Press.

Wald, A. (1950), *Statistical Decision Functions.* New York: John Wiley and Sons.

Walley, P. (1991), *Statistical Reasoning with Imprecise Probabilities.* London: Chapman and Hall.

Wang, T. (2003a), "A Class of Multi-Prior Preferences", UBC, mimeo.

——— (2003b), "Conditional Preferences and Updating" *Journal of Economic Theory*, 108: 286-321.

Welch, B. L. (1939), "On Confidence Limits and Sufficiency, and Particular Reference to Parameters of Location", *Annals of Mathematical Statistics*, 10: 58-69.

Weyl, H. (1944), "David Hilbert", *Obituary Notices of Fellows of the Royal Society*, 4: 547-553.

Yaari, M. E. (1969), "Some Remarks on Measures of Risk Aversion and on their Uses", *Journal of Economic Theory*, 1: 315-329.

——— (1987), "The Dual Theory of Choice under Risk", *Econometrica*, 55: 95-115.

Zhang, J. (2002), "Subjective Ambiguity, Expected Utility, and Choquet Expected Utility, *Economic Theory*, 20: 159-181.