Rationality, Decisions *and* Large Worlds

Mareile Drechsler

A thesis submitted to the Department of Philosophy, Logic and Scientific Method
of the London School of Economics and Political Science
for the degree of Doctor of Philosophy, October 2012.

To my parents.

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 57 782 words.

Mareile Drechsler

Statement of conjoint work

I confirm that Chapter 5 was jointly co-authored with Konstantinos Katsikopoulos and Gerd Gigerenzer at the Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin. I contributed 60% of this work.

Abstract

Taking Savage's (1954) subjective expected utility theory as a starting point, this thesis distinguishes three types of uncertainty which are incompatible with Savage's theory for small worlds: ambiguity, option uncertainty and state space uncertainty.

Under ambiguity agents cannot form a unique and additive probability function over the state space. Option uncertainty exists when agents cannot assign unique consequences to every state. Finally, state space uncertainty arises when the state space the agent constructs is not exhaustive, such that unforeseen contingencies can occur.

Chapter 2 explains Savage's notions of small and large worlds, and shows that ambiguity, option and state space uncertainty are incompatible with the small world representation. The chapter examines whether it is possible to reduce these types of uncertainty to one another.

Chapter 3 suggests a definition of objective ambiguity by extending Savage's framework to include an exogenous likelihood ranking over events. The definition allows for a precise distinction between ambiguity and ambiguity attitude. The chapter argues that under objective ambiguity, ambiguity aversion is normatively permissible.

Chapter 4 gives a model of option uncertainty. Using the two weak assumptions that the status quo is not uncertain, and that agents are option uncertainty averse, we derive status quo bias, the empirical tendency for agents to choose the status quo over other available alternatives. The model can be seen as rationalising status quo bias.

Chapter 5 gives an axiomatic characterisation and corresponding representation theorem for the priority heuristic, a heuristic which predicts binary decisions between lotteries particularly well. The chapter analyses the normative implications of this descriptive model.

Chapter 6 defends the pluralist view of decision theory this thesis assumes. The chapter discusses possible applications of the types of uncertainty defined in the thesis, and concludes.

Contents

List of Tables

Acknowledgements

First and foremost, I would like to thank Richard Bradley. Richard's feedback has been dependable, timely and astute, and in four years I cannot recall a single occasion where Richard has not had time for a discussion and advice. It truly is an extraordinary privilege to have had the opportunity to learn from, and work with Richard: his advice has shaped and improved this project greatly. Above all, however, Richard has always had my best interest at heart.

I am also very indebted to Katie Steele, who has done a remarkable job as a second supervisor. I have learned much throughout the past years from Katie, and have benefitted greatly from discussions with her. Katie is brilliantly precise, a quality which has helped improve my work considerably. Also, she has mastered the dual role of supervisor and friend whilst compromising on neither of these two roles. Katie is a person whose support I have been able to count on.

I would also like to thank Konstantinos Katsikopoulos and Gerd Gigerenzer at the Center for Adaptive Behavior and Cognition of the Max Planck Institute for Human Development, Berlin. Konstantinos and Gerd have from the outset placed great trust in my academic abilities, and have given me many opportunities. I have benefitted greatly from detailed discussions with Konstantinos, and I have always felt that his understanding and skills are complementary to my own. Gerd has invested much in my work, and has on many occasions made time for me in his extraordinarily busy schedule; this support and trust has gone a long way in building confidence in my own abilities.

At the LSE, I have, particularly at the beginning of my PhD, had many interesting and instructive discussions with Franz Dietrich, who has had the patience to explain many bits of theory to me, and improved my grasp of the concepts of decision theory. I would also like to thank Christian List for interesting discussions and encouragement. Erik Curiel has had the kindness to teach some of the basics of measurement theory to me, a skill which has proved useful within this PhD. Jason Alexander has been extremely supportive of my work in spite of the fact that he has not officially been my supervisor; he has read and given helpful feedback on my work many times. I would also like to thank Alex Voorhoeve for having written many a reference letter in support of my applications for scholarships, and for having been available for discussions at any time. It has been a great pleasure to co-teach Philosophy of Economics in 2010 – 11 with Nancy Cartwright and Armin Schulz, both of whom have been very kind and encouraging throughout my PhD. I would also particularly like to thank Luc Bovens, who has always been available for a chat and advice when I needed it.

Furthermore, I would like to thank Christian Bach, to whom it is due that I applied to the LSE in the first place: back in the year 2005 in Freiburg, Germany, he drew my attention to the interesting research done at the Philosophy Department at the LSE, and I was immediately sold on the idea that the LSE was the place to be. He has also put me in touch with Conrad Heilmann, who has already during my Masters at the LSE in 2006 – 07 proved to be an extremely generous and selfless source of support. Throughout the years, Conrad has given me much useful advice, and has been a reliable friend.

I would like to thank Tibor Neugebauer at the University of Luxembourg for having invited me to the conference *Economic Science Association European Conference, Luxembourg* in 2011, and for the encouragement he has given me. Andrea Isoni, who attended my talk in Luxembourg, has given me the great opportunity to give a talk at the DR@W forum at the University of Warwick in 2012. Philippe Mongin has kindly drawn my attention to the conference *Congrès Société de Philosophie Analytique* in Paris in 2012; I am very grateful and indebted to him for instructive discussions.

I am particularly thankful to the *Fonds National de la Recherche, Luxembourg*, without whom this project would not have been possible. At the FNR, I would like to thank Ulrike Kohl, Angelina Frank and Tom Jakobs for continued support throughout my PhD.

Last but not least, I would like to thank my support team: My parents, Astrid and Manfred Drechsler, who have done anything in their means to help me through this PhD, and my brother Florian, who, as a mathematician, has answered many a mathematical question over the years – he has always been someone I can look up to. Finally, I would like to thank Henry for being a source of strength and peace.

Chapter 1

Introduction

Savage's *Foundations of Statistics*, published in 1954, remains the classic work on normative decision theory, and the basis for many economic models and standard texts on decision theory. In the opening paragraph, Savage remarks

> *It is often argued academically that no science can be more secure than its foundations, and that, if there is controversy about the foundations, there must be even greater controversy about the higher parts of the science. As a matter of fact, the foundations are the most controversial parts of many, if not all, sciences.*

It is indeed very important that a theory is based on the "right" basic principles or axioms. However, these may be difficult to pin down. Although many academics are agreed that Savage's own postulates are very compelling, there has been much controversy over the extent to which they are applicable, and over the question whether they constitute requirements of rationality. This thesis is concerned with the foundations of normative decision theory, and more specifically, the correct principles to use in decision making under uncertainty. The overarching question this thesis grapples with is how a rational actor should respond to various sources, degrees and types of uncertainty, and which axioms and decision rules seem appropriate in those cases where Savage's theory is limited.

Before considering the role of the foundations of decision theory further, let us step back to consider the role of decision theory more generally. Clearly, in our everyday lives we are confronted with innumerable decisions, and nowadays making decisions is, perhaps, more difficult then ever: it is well known that ordering a coffee at Starbucks requires a total of eight decisions[1]. We make choices between partners, political parties, haircuts and holiday plans, mortgages and masters degrees. Decisions vary in their importance within the course of our lives: many are irrelevant (such as our choice of coffee at Starbucks), but many others require detailed reflection.

The role of *normative decision theory* is to answer the question how a rational agent *ought to* make decisions. The answer to this question is usually taken to be that a rational agent ought to choose that act amongst all available acts which maximises their expected utility, where the expected utility of an act is calculated relative to the agent's utility valuations on the consequences of the act, and the agent's (personal) beliefs regarding the likelihood of these consequences.

Savage's (1954) subjective expected utility theory, which can be seen as the foundation of Bayesian decision theory, shows that an agent whose preferences among acts satisfy a set of simple and intuitively compelling axioms will act as if they maximised their subjective expected utility, relative to a utility function on consequences and a probability measure on the state space. Thereby, the axioms of Savage's theory are usually interpreted as requirements of rationality.

Savage's theory is designed to apply to *small worlds*, decision settings in which the agent's problem can be represented using a decision matrix consisting of an exogenously given state space, a set of consequences, and a set of acts. Using a small world decision matrix, the agent then chooses that act in the set of acts which yields the highest subjective expected utility. A small world decision matrix is interpreted as containing all information which is relevant to the agent's decision problem.

However, as this thesis will argue, not all decision problems can be represented within a small world decision matrix. In particular, in order for a decision problem to be representable using a small world decision matrix, the agent must be able

[1] If you don't believe this, here is the list of basic decisions to be made in ordering a coffee. *Drink type*: Mocha, Latte, or Caramel Macchiato. *Drink size*: Short, Tall, Grande, Venti. *Drink style*: iced or warm. *Caffeination of espresso*: regular, decaf., or half-caf.. *Amount of espresso*: number of espresso shots. *Milk type*: non-fat, 2%, whole, half and half, and soy. *Syrup type*: any of 15 flavours. *Whipped cream*: with or without. Also, there are a number of more obscure dimensions, such as cup types and the exact temperature of the coffee; these are not listed within the eight decisions mentioned.

to reduce their uncertainty over what to do to uncertainty over what the true state in a given set of states is. This will be feasible, as we shall argue, only in special cases. Other types of uncertainty which are incompatible with a small world representation may affect the agent's decision problem. In particular, this thesis distinguishes three types of uncertainty: *ambiguity*, where the agent is uncertain with respect to the true probability distribution over the state space, *option uncertainty*, where the agent is uncertain what consequences follow from the exercise of any given action, and *state space uncertainty*, where the state space the agent entertains may not be exhaustive. We will call decision problems which feature ambiguity, option uncertainty or state space uncertainty *large worlds*, cases where a small world representation of the decision problem is not feasible.

This thesis asks the question how a rational agent should make decisions in large worlds. We will argue that under uncertainty, Savage's axiomatic characterisation is not as compelling as it is in the small world case. However, when one confines ones attention to a particular kind of uncertainty, it is possible to identify requirements of rationality suitable to the large world decision setting the agent is faced with. The axiomatic frameworks thus obtained are very similar to Savage's framework for small worlds, as most elements of Savage's framework can be retained under uncertainty. However, extending Savage's theory to allow for a variety of sources of uncertainty may yield interesting new implications, as this thesis hopes to show.

This chapter is structured as follows: section 1.1 introduces Savage's (1954) framework and explains the axioms of subjective expected utility theory. Section 1.2 introduces Anscombe and Aumann's (1963) reformulation of Savage's theory, and clarifies in what respects it differs from Savage's original framework. Section 1.3 introduces the notion of rationality normative decision theory is based on, and shows the importance of probability theory within decision theory. Section 1.4 presents the well-known paradoxes of rationality, the Allais and Ellsberg paradoxes. Section 1.5 considers the notions of small and large worlds, and the significance of uncertainty in decision theory. Section 1.6 gives an outline of the chapters of this thesis.

1.1 Savage's framework

Savage's achievement consisted in combining the subjective view of probability predominant at his time with von Neumann and Morgenstern's decision theory. The axioms of the theory are intuitively convincing, and are therefore often interpreted as standards of rationality. Savage's theorem shows that given a set of axioms on the agent's preferences over acts, the agent will behave as if he attached utilities to consequences and probabilities to states of the world. An agent whose preferences can be so characterised will then make decisions as if he maximised expected utility, relative to the corresponding subjective probability distribution over the state space.

The agent's decision problem in Savage's framework[2] consists in choosing between *acts*, which are functions from *states of the world* to *consequences*. In particular, in Savage's theory the set of acts includes all functions from states to consequences. The state space is assumed to consist of mutually exclusive and collectively exhaustive states, which detail all the relevant exogenous contingencies an agent's decision may depend on. Events are then collections of states, and are therefore subsets of the state space. The set of consequences details the outcomes of acts at all states. Also, Savage's theory assumes states and consequences to be primitives of the theory; acts are defined in terms of states and consequences. Consider the following definitions:

States of the world: $S = \{..., s, ...\}$.

Events: $\mathcal{E} := 2^S = \{..., A, B, E, F, ...\}$.

Consequences: $X = \{..., x, ...\}$.

Acts: $\mathcal{A} := X^S = \{..., f(\cdot), g(\cdot), ...\}$.

The agent is assumed to have preferences over acts, expressed as a relation $\succeq$ on $\mathcal{A}$, where $f \succeq g$ is to be read as "the agent weakly prefers act f to act g". The relation $\succeq$ is assumed to have a corresponding symmetric equivalence relation $\sim$, denoting "indifference", as well as an asymmetric part, $\succ$, denoting "strict preference". An event E is said to be *null* if $f \sim g$ for every f, g in $\mathcal{A}$ which differ on E. Preferences over $\mathcal{A}$ induce preferences over consequences,

[2] We follow Machina and Schmeidler (1992) in the exposition of Savage's framework.

since consequences can be understood as constant acts, which lead to the same consequence in every state. Then, Savage's axioms are:

[P1] (Weak Order): $\succeq$ *is a weak order on* $\mathcal{A}$*:*
(i) (Completeness): Either $f \succeq g$*, or* $g \succeq f$*.*
(ii) (Transitivity): If $f \succeq g$ *and* $g \succeq h$*, then* $f \succeq h$*.*

[P2] (Sure-Thing Principle): *For all events* E *and all acts* $f(\cdot), f^*(\cdot), g(\cdot)$ *and* $h(\cdot)$*:*

$$\begin{bmatrix} f^*(s) & \text{if} & s \in E \\ g(s) & \text{if} & s \notin E \end{bmatrix} \succeq \begin{bmatrix} f(s) & \text{if} & s \in E \\ g(s) & \text{if} & s \notin E \end{bmatrix}$$
$$\Rightarrow \begin{bmatrix} f^*(s) & \text{if} & s \in E \\ h(s) & \text{if} & s \notin E \end{bmatrix} \succeq \begin{bmatrix} f(s) & \text{if} & s \in E \\ h(s) & \text{if} & s \notin E \end{bmatrix}.$$

[P3] (Eventwise Monotonicity): *For all consequences* x, y*, non-null events* E *and acts* $g(\cdot)$*:*

$$\begin{bmatrix} x & \text{if} & s \in E \\ g(s) & \text{if} & s \notin E \end{bmatrix} \succeq \begin{bmatrix} y & \text{if} & s \in E \\ g(s) & \text{if} & s \notin E \end{bmatrix} \Leftrightarrow x \succeq y.$$

[P4] (Comparative Probability): *For all events* A, B *and outcomes* $x^* \succeq x$ *and* $y^* \succeq y$*:*

$$\begin{bmatrix} x^* & \text{if} & A \\ x & \text{if} & \neg A \end{bmatrix} \succeq \begin{bmatrix} x^* & \text{if} & B \\ x & \text{if} & \neg B \end{bmatrix}$$
$$\Rightarrow \begin{bmatrix} y^* & \text{if} & A \\ y & \text{if} & \neg A \end{bmatrix} \succeq \begin{bmatrix} y^* & \text{if} & B \\ y & \text{if} & \neg B \end{bmatrix}.$$

[P5] (Nondegeneracy): *There exist outcomes* x *and* y *such that* $x \succ y$*.*

[P6] (Small Event Continuity):*For all acts* $f(\cdot) \succeq g(\cdot)$ *and outcome* x *there exists a finite set of events* $\{A_1, A_2, ..., A_n\}$ *forming a partition of* S *such that:*

$$f(\cdot) \succ \begin{bmatrix} x & \text{if} & s \in A_i \\ g(s) & \text{if} & s \notin A_i \end{bmatrix} \text{and} \begin{bmatrix} x & \text{if} & s \in A_j \\ f(s) & \text{if} & s \notin A_j \end{bmatrix} \succ g(\cdot)$$

for all $i, j \in \{1, ..., n\}$.

[P7] (Uniform Monotonicity): *For all events* E *and all acts* $f(\cdot)$ *and* $f^*(\cdot)$, *if*

$$\begin{bmatrix} f^*(s) & \text{if} & s \in E \\ g(s) & \text{if} & s \notin E \end{bmatrix} \succeq (\preceq) \begin{bmatrix} x & \text{if} & s \in E \\ g(s) & \text{if} & s \notin E \end{bmatrix}$$

for all $g(\cdot)$ *and each* $x \in f(E)$, *then:*

$$\Rightarrow \begin{bmatrix} f^*(s) & \text{if} & s \in E \\ h(s) & \text{if} & s \notin E \end{bmatrix} \succeq (\preceq) \begin{bmatrix} f(s) & \text{if} & s \in E \\ h(s) & \text{if} & s \notin E \end{bmatrix}$$

for all $h(\cdot)$.

The completeness part of P1 requires that the agent be able to rank all acts in the order of preference: Either the agent strictly prefers act g to act h or vice versa, or he is indifferent between the two. This axiom precludes indecisiveness on the part of the agent. The transitivity component of P1 holds that if an agent prefers act f to act g, and also act g to act h, then he should also prefer act f to act h.

P2 holds that if two acts have different subacts ($f^*(s)$ and $f(s)$ respectively) on some event E, but agree on the event $\neg E$, then the ranking between the act should not depend on the common subact on $\neg E$. This axiom implies the separability of preferences across mutually exclusive events. The sure thing principle is a crucial element of the framework, since it implies that the expected utility function is linear in probabilities. We will discuss the axiom in greater detail in the context of Allais' and Ellsberg's experiments at the end of this chapter, as well as in Chapter 3.

The Eventwise Monotonicity condition, P3, requires that replacing a consequence on a non-null event with another consequence which the agent prefers should make that act preferable; this reading brings out the "monotonicity" aspect – more is better – of the axiom. Technically, the axiom holds that the preference for consequence x over y conditionally on event E should be independent of act $g(s)$ obtaining on the complement of E. The axiom can thus also be read as a "state-independence" condition: The evaluation of consequences should not hinge on the state they obtain in.

Axiom P4, Comparative Probability, maintains that the subjective beliefs the agent reveals through preferences over acts must be consistent: Given that the agent has the preference $x^* \succeq x$, then a preference of an act which yields x^* if event A occurs and x if A does not occur over an act which pays out x^* if B occurs and x if B does not occur reveals that the agent believes A to be more likely than B. Replacing x^* with y^* and x with y should not make a difference to subjective beliefs regarding the likelihood of A and B, given that $y^* \succeq y$. The axiom imposes that the agent's personal beliefs should be independent of the consequences used to elicit them. Axiom P4 is pivotal in the construction of the subjective probability ranking of events.

The Non-degeneracy axiom, P5, is a non-triviality condition which holds that the agent should not be indifferent between all consequences; this axiom is not very restrictive.

P6, the Small Event Continuity condition, requires that for every consequence x the state space can be partitioned sufficiently finely such that, if the agent has preference $f \succeq g$, replacing a consequence x for the act f on some element of the partition leaves his preference between f and g unchanged. This makes the state space infinitely fine-grained; S is then countable.

Finally, axiom P7, the uniform monotonicity condition, holds that if a consequence x is conditionally worse than any of the consequences of an act $f^*(s)$, then the subact which pays out x should not preferred. This axiom allows for infinite-outcome acts, and ensures the boundedness of the utility function on the set of consequences.

Savage shows that, if the above seven axioms are satisfied by the agent's preferences, then the agent will choose as if he maximised his expected utility relative to his subjective probability and utility functions:

Theorem *(Savage)*: *If $\succeq$ satisfies Axioms 1 – 7, then there exists a unique, finitely additive and nonatomic probability measure $\mu(\cdot)$ on $\mathcal{E}$ and a state-independent and bounded utility function $u : X \to \mathbb{R}$ such that*

$$\begin{gathered} f \succeq g \ \text{ iff} \\ \int_S u[f(s)]d\mu(s) \geq \int_S u[g(s)]d\mu(s) \end{gathered} \tag{1.1}$$

Moreover, u is unique up to a positive linear transformations, and $\mu(E) = 0$ if and only if E is null.

Savage's theorem will form the basis of the discussion of different types of uncertainty contained in Chapter 2, and of the model of option uncertainty given in Chapter 4.

1.2 The Anscombe-Aumann framework

Savage's theory has been reformulated by Anscombe and Aumann (1963), whose framework differs from Savage's in that it allows for the existence of lotteries with objectively known probabilities; in Savage's framework, by contrast, probabilities are only subjectively known. The Anscombe-Aumann framework distinguishes between *roulette lotteries* (called 'lotteries' henceforward), the results of which obtain with known chances, and *horse race lotteries* (called 'acts' henceforward), the outcomes of which occur with subjectively known probabilities.

In the Anscombe-Aumann framework, the state space S is finite. The set of lotteries $\mathcal{L}$ is modelled as finite support probability distributions over the set of outcomes $\mathcal{X}$. A typical lottery is denoted p, and is defined as $p : \mathcal{X} \to [0, 1]$. Also, unlike Savage's model, the Anscombe-Aumann framework permits for mixtures of lotteries. The mixture operation is denoted α, with $\alpha \in [0, 1]$, and for two lotteries $p, q \in \mathcal{L}$, $\alpha p + (1 - \alpha)q$ is defined pointwise over $\mathcal{X}$. The set of acts is denoted $\mathcal{F}$ with typical elements f, g. In contrast to Savage's framework, where acts are functions from states of the world into consequences, in the Anscombe-Aumann framework acts are defined as functions from states of the world S into lotteries $\mathcal{L}$, so that an act pays out a gamble with known chances at every state. The subset $\mathcal{F}_c$ of $\mathcal{F}$ denotes the set of constant acts (i.e. those that yield the same lottery in every state). The agent then holds preferences over acts, with $\succeq$ denoting weak preference. The asymmetric and symmetric components of $\succeq$ are, respectively, denoted $\succ$ and $\sim$. In summary:

States of the world: $S = \{..., s, ...\}$.

Algebra of Events: $\Sigma = \{..., A, B, E, F, ...\}$.

Outcomes: $\mathcal{X} = \{..., x, ...\}$.

Lotteries: $\mathcal{L}$: $[0,1]^{\mathcal{X}} = \{..., p(\cdot), q(\cdot), ...\}$

Acts: $\mathcal{F} := \mathcal{L}^S = \{..., f(\cdot), g(\cdot), ...\}$.

Anscombe and Aumann then impose the following axioms on preferences:

[AA1] (Weak Order): *$\succeq$ is a weak order on $\mathcal{F}$.*

[AA2] (Continuity): *For all p, q, r in $\mathcal{L}$ such that $p \succ q \succ r$, there exist $\alpha, \beta \in]0,1[$ such that*

$$\alpha p + (1-\alpha)r \succ q \succ \beta p + (1-\beta)r.$$

[AA3] (Independence): *For all f, g, h in $\mathcal{F}$ and for every $\alpha \in]0,1[$*

$$f \succeq g \Leftrightarrow \alpha f + (1-\alpha)h \succeq \alpha g + (1-\alpha)h.$$

[AA4] (Monotonicity): *For all f, g in $\mathcal{F}$, if $f(s) \succeq g(s)$ for all $s \in S$, then $f \succeq g$.*

[AA5] (Nontriviality): *There exists at least one pair of acts f, g such that $f \succ g$.*

The interpretation of the weak order and nontrivility conditions matches that of Savage's axioms P1 and P5. The continuity condition performs, in the Anscombe-Aumann framework the same function as Savage's Archimedean axiom, as it results in the continuity of the utility function. The independence axiom is the equivalent, in the Anscombe-Aumann framework, of Savage's axiom P2, the sure-thing principle; it implies the separability of preference across mutually exclusive events. The monotonicity condition holds that if at every state the lottery paid out by act f is preferred by the agent to that paid out by g, then the agent should prefer act f to act g.

Theorem *(Anscombe-Aumann)*: *If $\succeq$ satisfies Axioms AA 1– 5, then there exists a function $\mu \in \mathcal{L}$ and a function $u : \mathcal{X} \to \mathbb{R}$ such that for any $f, g \in \mathcal{F}$*

$$\begin{gathered} f \succeq g \text{ iff} \\ \sum_{s \in S} \mu_s \sum_{x \in \mathcal{X}} f_s(x) u(x) \geq \sum_{s \in S} \mu_s \sum_{x \in \mathcal{X}} g_s(x) u(x) \end{gathered} \tag{1.2}$$

Furthermore, μ is unique and u is unique up to positive affine transformation.

In the Anscombe-Aumann framework, uncertainty is resolved in two steps: in a first step the outcome of acts is determined by the state of the world, and in a second step the outcome of the lottery the act yields is resolved. In the theorem above, μ_s reflects the subjective probability of state s. $f_s(x)$ denotes the probability of outcome x given that state s is true when act f is chosen, and $u(x)$ denotes the utility the agent attributes to the final outcome x.

We will use the Anscombe-Aumann framework in the discussion of ambiguity contained in Chapter 3, since much of the literature on ambiguity and ambiguity aversion is conducted within this framework.

1.3 Rationality

In the introduction, we claimed that rationality is often identified with the maximisation of expected utility, and in particular with the satisfaction of Savage's axioms of subjective expected utility theory. Savage's framework can be seen as the foundation of Bayesian decision theory, which continues to be the paradigm in much of economics. In this section, we will investigate the main claims of Bayesian decision theory.

One can identify at least three tenets of Bayesianism: first, the idea that all uncertainty can be quantified in a single probability distribution satisfying the axioms of probability theory. Second, the stance that agents should update their personal beliefs using Bayes' law. Finally, Bayesianism in decision theory holds that agents must maximise their expected utility relative to their subjective beliefs. Let us investigate each of these tenets in greater detail.

The first tenet of Bayesianism requires agents to form a subjective likelihood ordering over events which can be represented using a unique and additive prior probability distribution $P(\cdot)$ on the state space $(S, 2^S)$. Such a probability distribution will satisfy the axioms of probability theory:

[**Axiom 1**] $P(A) \geq 0$ *for all events* A.

[**Axiom 2**] $P(S) = 1$

[Axiom 3] $P(A \cup B) = P(A) + P(B)$ *for all disjoint events* A *and* B.

Axiom 1 holds that the probability of all events is larger or equal to zero. Axiom 2 holds that the probability of the state space is equal to one. Axiom 3 states that the probability of the union of two disjoint events A and B must be equal to the sum of their individual probabilities. As we shall discuss later, axiom 3 above is violated systematically in experiments: people do not generally hold preferences which are consistent with the existence of an additive probability distribution over the state space, i.e. a distribution for which the sum of the individual probabilities is equal to the probability of their union. Therefore, it has been suggested in the literature that in order to accommodate the empirical evidence suggesting that agents do not always hold beliefs which are representable using a probability distribution satisfying these axioms, the axioms have to be weakened; this topic will be pursued in greater detail in Chapters 2 and 3.

In Savage's theory (and the Anscombe-Aumann reformulation of it), agents will hold a unique probability distribution over the state space satisfying the axioms above. To see how this is generated, consider Savage's postulate P4, the comparative probability axiom:

[P4] (Comparative Probability): *For all events* A, B *and outcomes* $x^* \succeq x$ *and* $y^* \succeq y$:

$$\begin{bmatrix} x^* & \text{if} & A \\ x & \text{if} & \neg A \end{bmatrix} \succeq \begin{bmatrix} x^* & \text{if} & B \\ x & \text{if} & \neg B \end{bmatrix}$$
$$\Rightarrow \begin{bmatrix} y^* & \text{if} & A \\ y & \text{if} & \neg A \end{bmatrix} \succeq \begin{bmatrix} y^* & \text{if} & B \\ y & \text{if} & \neg B \end{bmatrix}.$$

The axiom shows that we can use constant acts x^* and x with $x^* \succeq x$ to construct a likelihood ordering, which we shall denote $\succeq^*$ on S. In particular,

$$A \succeq^* B \Leftrightarrow \begin{bmatrix} x^* & \text{if} & A \\ x & \text{if} & \neg A \end{bmatrix} \succeq \begin{bmatrix} x^* & \text{if} & B \\ x & \text{if} & \neg B \end{bmatrix}$$

The idea is that whenever the agent prefers an act which yields outcome x^* on A and x on $\neg A$ to an act which yields x^* on B and x on $\neg B$, and the agent prefers x^* to x, then they must think that the event A is at least as likely as the event

B. Axiom P4 then additionally imposes that this likelihood ranking of events is independent of the constant acts used to elicit it. Jointly, axioms P1 – P6 imply the existence of a unique and finitely additive probability measure on S:

Theorem *(Savage)*: *If $\succeq$ satisfies Axioms P1 – P6, then $\succeq^*$ can be represented by a unique probability measure on S. That is, there is a unique and finitely additive probability P defined on S, such that for every $A, B \subseteq S$*

$$A \succeq^* B \Longleftrightarrow P(A) \geq P(B)$$

and if $A \subseteq S$ and $0 \leq \rho \leq 1$ there is a $B \subseteq A$ such that

$$P(B) = \rho P(A).$$

Let us now turn to the second tenet of Bayesianism, namely that an agent must update their prior probabilities via Bayes' rule. We have seen that Savage's decision theory implies the existence of a likelihood ordering on events which can be represented by a probability measure; agents thus hold prior probabilities for all events. Then, Bayes' rule requires that the agent updates the prior probabilities to posterior probabilities as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1.3}$$

In the above equation, A and B are events. $P(A|B)$ denotes the posterior probability of A given B, $P(A)$ and $P(B)$ are the prior probabilities of events A and B, and $P(B|A)$ the conditional probability of B given A.

Savage's decision theory, as expounded above, is a static framework, hence Bayes' rule does not come into play directly. However, it is relatively simple to show that Savage's framework implies consistency with Bayes' rule on a dynamic reading of Savage's framework. To this end, consider a *conditional preference relation* $\succeq_A$ which is interpreted as the agent's preference relation upon observing the event A. That is to say, the agent now knows more than that any state in S is true, since he knows that the true state is in A. Now define the preference relation conditional on A as follows:

$$f \succeq_A g \Leftrightarrow \begin{bmatrix} f & \text{if} & A \\ h & \text{if} & \neg A \end{bmatrix} \succeq \begin{bmatrix} g & \text{if} & A \\ h & \text{if} & \neg A \end{bmatrix}$$

Then by a straightforward extension of Savage's theorem, it is possible to show that the conditional preference relation $\succeq_A$ can be represented as follows (Ghirardato, 2002):

Theorem *(Representation of conditional probability)*: *If $\succeq_A$ satisfies Axioms 1 – 7, then there exists a unique and finitely additive probability measure $\mu(\cdot)$ on $\mathcal{E}$ and a state-independent utility function $U : X \to \mathbb{R}$ such that*

$$\begin{gathered} f \succeq_A g \ \text{ iff} \\ \int_S u[f(s)]P_A(ds) \geq \int_S u[g(s)]P_A(ds). \end{gathered} \tag{1.4}$$

In the above equation, P_A is the Bayesian update of the probability measure P conditional on A. So if the agent's conditional preferences $\succeq_A$ satisfy Savage's axioms, then the agent will behave as if they maximised expected utility relative to the posterior probability P_A. However, since this thesis is concerned only with Savage's static framework, we will not pursue the link between Savage's theorem and Bayesian updating any further.

The final tenet of Bayesianism holds that agents must maximise their expected utility. To understand this requirement, note first that decisions were first studied by Pascal and Fermat, who studied people's gambling behaviour from a theoretical point of view. Pascal and Fermat than held that rational choice consisted in choosing the gamble with the highest expected value:

$$EV = \sum_{i=1}^{n} p_i x_i$$

Thereby, 'EV' denotes expected value, and letting $i = \{1, ..., n\}$ denote the event, p_i gives the probability of event i and x_i denotes the payoff of event i. However, the view that rational choice consists in choosing the gamble with the highest expected value quickly came under attack, as it is contradicted by evidence on the St. Petersburg paradox. The setup of the problem is as follows: a fair coin is tossed until "heads" comes up for the first time, at which point the game ends. The payoff of the gamble is dependent on the number of times "tails" has come up in consecutive tosses, and the payoff doubles with every toss of the coin. One therefore wins $\$2^{k-1}$ for k tosses of the coin. The expected value of this gamble is infinite: $EV = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 4 + \frac{1}{16} \cdot 8 + ... = \sum_{k=1}^{\infty} \frac{1}{2} = \infty$. Yet most rational agents would, at best, place a very small sum of money on this bet; this problem therefore became known as a paradox.

In response to the St. Petersburg paradox, Bernoulli suggested that the expected value does not reflect the subjective value a specific amount of money has for a person. Hence, he concluded that rational choice consists in choosing that gamble which has the highest expected utility:

$$EU = \sum p_i u(x_i)$$

where $u(x_i)$ is the utility transformation of payoffs, generally assumed to be monotonically increasing. Of course, Savage's theory satisfies this requirement, as equation (1.1) shows.

One may ask, then, why the tenets of Bayesianism, as characterised here, form a canon of rationality. The answer to this question is commonly given by appeal to the Kantian notion of *practical reason*, the basic human capacity to resolve the question of what to do. Practical reason is a normative approach, since it concerns what the agent rationally *ought* to do. The answer to the normative question of what is best to do is then that it is rational to act in one's own best interest, by maximising (subjective) expected utility. The axioms of (subjective) expected utility theory then embody basic consistency requirements of the agent's *deliberations*; we can judge the agent as rational or irrational depending on whether their deliberations are consistent or not. As we can see, practical reason is a framework, or calculus, of rationality. Notice that practical reason constitutes an *a priori* notion of rationality, which proceeds from universally applicable first principles. The axioms of Savage's theory are usually understood as such a priori consistency, or rationality requirements, the failure of which is attributed to a flaw in the agent's deliberations.

There is, however, much dispute over the question whether Savage's axioms are indeed universally applicable, or whether they constitute requirements of rationality at all. We will answer the questions when Bayesian decision theory should be seen as rational, and what requirements of rationality hold when Bayesianism doesn't, shortly. First, however, we will investigate the two most prominent counterexamples against Savage's theory.

1.4 Paradoxes of Rationality

Savage's axiom P2, the Sure-Thing Principle, requires that preferences are separable across events. This axiom has attracted particularly severe criticism, since it has been shown to be violated systematically in empirical tests. Two different empirical results are particularly noteworthy: the experiments of Allais (1953) and Ellsberg (1961). The next two sections will introduce these two experiments.

The Allais paradox

Maurice Allais tested Savage's sure-thing principle using the example of the gambles given in Table 1.1. Gamble a_1 pays out \$1 million with certainty, whereas gamble a_2 pays out \$5 million with a probability of 10%, \$1 million with a probability of 89% and nothing with a probability of 1%. When asked to choose between gambles a_1 and a_2, most people prefer gamble a_1. Furthermore, gamble a_3 pays out \$1 million with a probability of 11% and \$0 with a probability of 89%, whereas gamble a_4 pays out \$5 million with a 10% probability, and nothing with a 90% probability. When asked to choose between gambles a_3 and a_4, most people prefer gamble a_4.

As we can easily verify, the preference pattern $a_1 \succ a_2$ and $a_4 \succ a_3$ is inconsistent with Savage's *Sure Thing Principle*: gambles a_1 and a_2 have the same outcomes for lottery tickets #12-100, therefore the consequences of these events should be irrelevant to the agent's preference between a_1 and a_2. Also, gambles a_3 and a_4 have the same payoffs for lottery tickets #12-100, therefore, this aspect should be irrelevant to the agent's preference between a_3 and a_4. But, crossing out the column for lottery tickets #12-100, we can see that gamble a_1 is identical to gamble a_3, and gamble a_2 is identical to gamble a_4. Savage's sure thing principle demands that an agent who prefers a_1 to a_2 (a_2 to a_1) should also prefer a_3 to a_4

	#1	#2-#11	#12-100
a_1	\$1M	\$1M	\$1M
a_2	\$0	\$5M	\$1M
a_3	\$1M	\$1M	\$0
a_4	\$0	\$5M	\$0

Table 1.1: The Allais paradox

(a_4 to a_3). The typical preference of a_1 over a_2 and a_4 over a_3 is thus inconsistent with the principle. Allais' problem is usually referred to as a "paradox" since most people find the sure-thing principle intuitively compelling as a requirement of rationality, and simultaneously have the intuition that they would like to choose a_1 and a_4.

We will discuss the Allais paradox in greater detail in Chapter 3, where we will attempt to reconcile the paradox with Savage's theory. However, for now let us consider the various responses that have been made to Allais's paradox. Consider Savage's own response first (1954, p.103):

> *It seems to me that in reversing my preference between Gambles 3 [here: a_3] and 4 [here: a_4] I have corrected an error. There is, of course, an important sense in which preferences, being entirely subjective, cannot be in error; but in a different, more subtle sense they can be. Let me illustrate by a simple example containing no reference to uncertainty. A man buying a car for \$2,134.56 is tempted to order it with a radio installed, which will bring the total price to \$2,228.41, feeling that the difference is trifling. But, when he reflects that, if he already had the car, he certainly would not spend \$93.85 for a radio for it, he realizes that he has made an error.*

As this quote suggests, Savage thinks that upon reflection, people would see that their preference of a_4 over a_3 were in error, and would therefore reverse their preference if given the opportunity. This stance is denied by Shafer (1986), who argues that preferences can not be in error. In particular, Shafer (1986) argues that the example given by Savage is just evidence to the effect that preference is not invariant under different measurements: in the context of buying the car, it seems to the man that it's best to buy it with the radio installed, whereas in case the man were not buying a car, it would seem to them that the radio is too expensive. However, as Shafer argues, Savage assumes that only one of the two ways of asking himself whether or not to buy the radio is correct, namely that where the agent assesses his preference between the radio and the money it costs outside the context of buying a car. But, Shafer points out, it is ultimately up to the man to decide which representation of the decision problem is best suited to determining whether or not he would like to buy the radio; perhaps it is in the context of buying the car that he can best place a value on his desire for a radio. Shafer argues that preferences can not be in error in this way. In Chapter 3, we will argue that the preference pattern commonly revealed in Allais' paradox are not in error, but they may be interpreted as reflecting the difficulty an agent has with constructing a suitable model which allows the agent to form preferences amongst Allais' gambles.

Whilst as we have seen, Savage believes that the preference patterns stated by the agents in Allais' paradox are in error, Savage also holds that if agents truly wish to violate the sure-thing principle even on reflection, then the sure-thing principle is to be abandoned:

> *If, after thorough deliberation, anyone maintains a pair of distinct preferences that are in conflict with the sure-thing principle, he must abandon, or modify, the principle; for that kind of discrepancy seems intolerable in a normative theory. [...] In general, a person who has tentatively accepted a normative theory must conscientiously study situations in which the theory seems to lead him astray; he must decide for each by reflection – deduction will typically be of little relevance – whether to retain his initial impression of the situation or to accept the implications of the theory for it.*

Whilst Savage argues that the sure-thing principle has no normative force if agents wish to violate it even on reflection, Savage also thinks that no rational agent would wish to maintain the preference pattern violating independence if given the chance to revise his decision. So Savage denies that people "truly" prefer a_4 to a_3, they are just taken in by the strong appeal of irrational decisions. As we shall argue in Chapter 3, cases of decision making under ambiguity present just such a case where even on reflection agents wish to violate the sure-thing principle; in contrast, we agree with Savage that in cases of risk, such as in Allais' paradox, the sure-thing principle is compelling as a normative requirement.

Allais himself denies the view that the sure-thing principle is valid even under risk. He believes that adherence to Savage's sure-thing principle should not be considered a question of rationality at all, and that, indeed, choosing according to psychological factors should be permissible. This is expressed in the following quote:

> *Il convient de noter en passant que ces élements ne sont pas qualifiés d'"irrationnels". Il est admis qu'un individu "rationnel" peut avoir une échelle des valeurs psychologiques différentes de l'échelle des valeurs monétaires et qu'il peut avoir une propension plus ou moins grande pour la sécurité ou pour le risque. Il paraît admis que c'est là une question de psychologie et non de "rationnalité"* [3].

We have, then, identified two distinct responses to Allais' paradox: first, denying its relevance on the grounds that it reveals a common flaw of reasoning, and secondly, denying the sure thing principle. The first response is endorsed by Savage, and the second by Allais. Those who believe that the empirical failure of the sure thing principle indicates that the principle be abandoned have suggested variants of expected utility theory without independence. For instance, Machina (1982) provides an expected utility model without independence (the expected utility equivalent of the sure-thing principle).

However, a third response to the Allais paradox is feasible (Steele, 2006), which attempts to reconcile the Allais paradox with the sure-thing principle. This position is defended, for instance, by Broome (1991). According to Broome's argument, the consequences of the gambles in the Allais paradox are not "sure experiences of the deciding person" of the kind Savage had in mind, since they fail to incorporate the agent's attitude to risk. Then, the outcome of the first gamble is more valuable to the agent since the payoffs are obtained with certainty; the "certainty" aspect of lottery a_1 should thus be factored into the outcome of the lottery, such that the payoff of the lottery becomes "\$1 Million + δ", where δ reflects the additional value of the outcome due to certainty of the payoff. Under this new version, call it *Allais**, the sure-thing principle is not longer violated.

[3] It is useful to note in passing that these elements are not labeled "irrational". It is admissible that an individual may have a separate scale of psychological values from the scale of monetary values, and that he may have a larger or smaller propensity toward risk. It seems that this is a question of psychology rather than "rationality".

There are at least two reasons why re-describing outcomes in the way proposed by Broome (1991) is illegitimate. The first makes references to what Broome calls the *rectangular field assumption*, which Broome himself recognises as a counter-argument to his position. In Savage's theory, acts are defined as function from states of the world to consequences. Then the set of acts comprises all possible functions from states to consequences – including the original Allais gambles. The agent is required to have preferences over all acts thus defined. The *rectangular field assumption* therefore implies that the Allais paradox cuts against the sure-thing principle. Re-describing outcomes in the way suggested by Broome is thus incompatible with the rectangular field assumption contained in Savage's theory, and is hence illegitimate.

The second reason why Broome's re-description strategy is unsuccessful is that on Broome's position, probabilities are interpreted *both* as beliefs (which are used as decision weights) *and* as carriers of utility. A Bayesian would reject this position, since a rational decision maker should not attach utility values to beliefs, as Broome's argument would suggest. One basic premise of Bayesian decision theory is that values and beliefs can be separated, a credo Broome's argument breaches.

For these two reasons, Broome's re-description strategy of Allais' paradox seems unsuccessful. However, we agree with Broome's view that the sure-thing principle is compelling as a normative requirement in situations of risk; Chapter 3 will therefore give its own attempt for reconciling Allais' paradox with Savage's sure-thing principle.

The Ellsberg paradox

A second objection to Savage's sure-thing principle was made by Ellsberg (1961). The Ellsberg gambles are given in table 1.2. The setup is as follows: an urn contains 90 balls, 30 of which are red, and the remaining 60 are black or yellow in an unknown proportion. The probability of drawing a red ball is then $1/3$, and the probability of drawing a black (respectively yellow) ball is within the closed interval $[0; 2/3]$. Then, gamble e_1 pays out \$100 if a red ball is drawn, and gamble e_2 pays out \$100 if a black ball is drawn. When choosing between e_1 and e_2, most people opt for e_1. Furthermore, gamble e_3 pays out \$100 if a red or yellow ball

	red	black	yellow
e_1	\$100	\$0	\$0
e_2	\$0	\$100	\$0
e_3	\$100	\$0	\$100
e_4	\$0	\$100	\$100

Table 1.2: The Ellsberg paradox

is drawn from the urn, and e_4 pays out \$100 if a black or yellow ball is drawn. When given the choice between e_3 and e_4, most people choose e_4.

These choices are inconsistent with Savage's sure-thing principle. We can see this if we focus on the last column of the table: gambles e_1 and e_2 have the same payoff if a yellow ball is drawn. Then, the agent's choice between gambles e_1 and e_2 should be dependent only on the outcomes in the events red and black. Also, gambles e_3 and e_4 have the same payoff in the event yellow, so the agent's choice between e_3 and e_4 should also be based only on the events red or black. But then, the restricted gamble e_1 becomes the same as the restricted gamble e_3, and the restricted gamble e_2 becomes the same as the restricted gamble e_4. Therefore, Savage's sure-thing principle requires that an agent who prefers e_1 to e_2 (e_2 to e_1) should also prefer e_3 to e_4 (e_4 to e_3).

The choice of e_1 and e_4 is inconsistent with the existence of a unique probability distribution over the states. This can be easily verified if we focus on the event 'black'. A choice (interpreted as strict preference) of e_1 over e_2 would suggest that the person believes the event 'red' to be more likely than 'black': both gambles pay out the same amount, and the only reason why the agent would prefer e_1 to e_2 is that he believes 'red' to be more likely than 'black'. This indicates that the agent believes the probability of 'black' to be less than $1/3$. But then, a choice of e_4 over e_3 indicates a belief that 'black' is more likely than 'red', since both gambles pay out the same amount in the event 'yellow', and a strict preference of e_4 over e_3 can arise only if the agent holds 'black' to be more likely than 'red'. This yields a probability of 'black' greater than $1/3$.

The results of Ellsberg's experiment are usually interpreted as arising from aversion to the uncertainty over the probabilities of the states 'black' and 'yellow': we have seen that the agent can entertain any probability assignment for the states 'black' and 'yellow' in the range of $[0; 2/3]$. Whenever the agent does not hold a "sharp" prior probability for every state, this type of uncertainty is referred to as "ambiguity". Then, agents who are averse to ambiguity behave as if the uncertainty regarding the correct probability distribution over the states could be turned against them: in each decision situation, they act as if one of the less favourable distributions were the true distribution. In the present case, this implies acting as if the probability of 'black' is less than $1/3$ for gamble e_2, and as if the probability of 'black' was greater than $1/3$ in gamble e_4. We will discuss the concept of ambiguity in greater detail in Chapter 2, and give definitions of the concepts of ambiguity and ambiguity aversion in Chapter 3.

Ellsberg suggested the above gambles in order to show that it is not the case that all uncertainty can be quantified within a single probability distribution, as the first tenet of Bayesianism would suggest. This is expressed in the following quote:

> *A number of sets of constraints on choice-behavior under uncertainty have now been proposed, all more or less equivalent or closely similar in spirit, having the implication that – for a "rational" man –* ***all*** *uncertainties can be reduced to* ***risks.***

However, since there is systematic evidence to the effect that people's preferences are not compatible with beliefs which are representable using a unique and additive subjective prior probability distribution, the first tenet of Bayesianism does not cohere with evidence. I will argue, in the course of this thesis, that ambiguity is a type of uncertainty which is incompatible with Savage's subjective expected utility theory for small worlds, and that it should therefore be understood as a type of uncertainty pertinent to large world settings, which require separate theoretical treatment.

One may wonder, at this point, what the difference between the Allais and Ellsberg paradoxes is. Both constitute a violation of the sure-thing principle and are very similar in structure. The main difference lies in the fact that whilst Allais' paradox may be seen as a violation of the third tenet of Bayesianism, namely that agents maximise expected utility, Ellsberg's paradox is commonly understood as a violation of the first tenet of Bayesianism, that all uncertainty is quantified in a unique and additive probability distribution. More specifically, Allais' paradox is compatible with *probabilistic sophistication*[4], the requirement that the agent's

[4] The concept of probabilistic sophistication is explained in greater detail in section 3.3.

beliefs can be modelled as a probability distribution satisfying the axioms of probability theory given in section 1.3, whereas Ellsberg's experiment is not (Machina and Schmeidler, 1992). This thesis adheres to the view that the third tenet of Bayesianism is a requirement of rationality, whereas the first is not.

The view that agents should not be required to quantify all uncertainty in a unique and additive probability distribution over the state space has famously been argued by Schmeidler (1989) and Gilboa and Schmeidler (1989). The authors give the following example to motivate this claim: suppose you are asked to bet on a coin. You have a coin in your pocket which you have flipped frequently and therefore you know that the relative frequency of heads is approximately 50%. I also have a coin in my pocket, but you know nothing about my coin. Now Bayesian reasoning requires that you should assign probabilities to the events of each of the coins landing heads. Of course, the probability for the first coin landing heads should be 50%. Also, due to the symmetry of one's ignorance with respect to the unknown coin, one should assign a probability of 50% to the second coin landing heads. Now both coins have been assigned a probability of 50% for heads. But this, Schmeidler argues, seems dubious: one would presumably prefer to bet on the first coin, since the probability assignment is based on facts rather than by default.

Guided by this intuition, Gilboa and Schmeidler have developed two different models for choice under ambiguity: First, Schmeidler's (1989) nonadditive probability, or Choquet expected utility (CEU) model relaxes the additivity of probability and thereby permits for the modal behaviour observed in Ellsberg's experiment, and secondly, Gilboa and Schmeidler's (1989) maxmin expected utility (MEU) model, which relaxes the uniqueness of the probability distribution. Both models aim to reconcile Ellsberg's paradox with Savage's framework (or rather, Anscombe and Aumann's reformulation of Savage's framework). In particular, in the CEU model agents maximise their expected utility with respect to nonadditive beliefs called capacities. In contrast, in the MEU model agents will choose that act amongst the set of acts which maximises subjective expected utility under the assumption that the least favourable of all possible probability distributions is the true distribution.

It has been argued in the literature, however, that the MEU model is too extreme in the sense that agents should not be required to choose as if the least favourable of all possible probability distributions were the true distribution; such a decision rule seems too conservative. In particular, Jaffray (1989) has suggested to make use of Hurwicz's (1951) α-criterion in cases of ambiguity. Suppose that c and C are the worst and best payoffs the agent may receive from choosing a particular act. Then Hurwicz's criterion demands that the agent chooses that action which maximises $(1-\alpha)c + \alpha C$, where α reflects how optimistic or pessimistic the agent is with respect to ambiguity (Binmore, 2009). In a similar vein, Ghirardato, Maccheroni and Marinacci (2004) have proffered the so-called α-MEU model, according to which an agent will choose that act which maximises a convex combination of the least and highest expected utilities that could result from the choice of a particular action. We will pursue the topic of ambiguity further in Chapter 3 of this thesis.

1.5 Uncertainty, small worlds and large worlds

Savage's subjective expected utility theory is designed to be suitable to *small world* decision situations. What, however, is a small world? Savage himself distinguishes between small worlds and grand worlds using the following two proverbs: you are in a small world if you can *look before you leap*, and you are in a grand world if you must *cross the bridge when you come to it.* That is to say, the agent is in a small world decision problem if it is feasible to optimise by maximising subjective expected utility, whereas the agent's decision problem is one of a grand world whenever the uncertainty contained in the decision problem is so severe that the agent cannot rationally respond to it. Savage then gives a number of everyday examples of typical small world reasoning, in particular (Savage, 1954, p.8):

1. Whether a particular egg is rotten.
2. Which, if any, in a particular dozen of eggs are rotten.
3. The temperature at noon in Chicago yesterday. [...]
4. The infinite sequence of heads or tails that will result from repeated tosses of a particular (everlasting) coin.

	State	
Options	Good	Rotten
break into bowl	six-egg omelet	no omelet, and five good eggs destroyed
break into saucer	six-egg omelet, and a saucer to wash	five-egg omelet, and a saucer to wash
throw away	five-egg omelet, and one good egg destroyed	five-egg omelet

Table 1.3: A small world decision matrix

Savage then argues that these examples have certain features in common. Most importantly, in each case the agent is uncertain about a some feature, such as the goodness of an egg, the temperature, or the number of heads and tails in coin tosses. This uncertainty can then be expressed, or quantified, within a state space which captures all uncertainty pertinent to the decision problem.

Formally, a *small world model* is associated with the existence of a *decision matrix* such as the one contained in Table 1.3. The table illustrates a man's decision problem when cooking an omelet. In the example, the man has already broken five eggs into a bowl, and now considers breaking the sixth egg into the same bowl. This decision depends on whether he thinks that the sixth egg is good or rotten, and what likelihood these cases have. He then evaluates the available acts of 'breaking the egg into a bowl', 'breaking the egg into a separate saucer', and 'throwing it away' in light of the consequences each of these acts would have in each of the states. We will now discuss the elements of the table, namely a state space, consequences and acts, in greater detail.

Savage characterises a state of the world as *a description of the world leaving no relevant aspect undescribed*. So the state space resolves all uncertainty contained in the decision problem, by enumerating all relevant contingencies the agent's decision problem may depend on. This is contrasted, in Savage's terminology, with the world itself, which is the object the agent is uncertain about, and the true state of the world, which is that state in the state space which obtains as the world unfolds and the uncertainty is resolved.

Savage then introduces the comparative notion of *larger and smaller worlds.* While we will explain this distinction more formally in Chapter 2, the idea is that a larger world contains more details regarding the decision problem than a smaller world. So, for instance, when deliberating whether to take an umbrella to go out for a walk, one might either consider the smaller world states 'rainy' and 'sunny', or the more detailed, larger world states 'rainy and windy', 'rainy and not windy', 'sunny and windy' and 'sunny and not windy'. We can see, then, that the states 'rainy' and 'sunny' form a partition of the larger world state space: the state 'rainy' can be understood as the disjunction of the two states 'rainy and windy' and 'rainy and not windy', and similarly for the state 'sunny'. A smaller world then neglects the distinction between the case where it is windy or not, but does not elide any large world state entirely. However, a small world model *may* elide a large world state entirely, Savage argues, when the state is considered "virtually impossible" by the agent. Savage characterises precisely under what conditions a small world is a satisfactory representation of the "grand world" – an ultimately refined model of the world – this shall be discussed in Chapter 2.

A *consequence*, according to Savage, is then "anything that may happen to a person" (Savage, 1954, p.13), it is construed as an experience of the deciding person, or, as Savage puts it, a "state of the person" as opposed to a "state of the world". A typical consequence will, under this conception of it, detail every aspect of the person's experience which might be relevant to them, such as money, health, the well-being of others, and so forth.

The notion of an *act* is defined as a function from the state space into the set of consequences; states and consequences are primitive notions, and acts are defined derivatively. The set of acts contains all possible functions from states of the world to consequences. More intuitively, Savage argues that an act just consists of a combination of consequences for every state. This notion of acts may, of course, yield acts which cannot be verbalised easily: for instance, what *is* the acts which yields a global temperature rise of 1 degree celsius by 2025, and 3 degrees by 2100? In very simple decision settings such as that of the example of cooking an omelet, it is, however, fairly straightforward to identify an act as that conduct of the agent which brings about, for instance, the consequence of obtaining a 'six-egg omelet' when the sixth egg is good, and 'no omelet and five eggs destroyed' when the sixth egg is rotten.

Savage then considers the case where the agent does not know what consequences follows at a particular state; for instance, where the agent does not know whether one rotten egg will, in fact, spoil the entire omelet or not. This scenario, where agents cannot assign a unique consequence to every state, will be integral to this thesis; we will call this case *option uncertainty*, and discuss its implications in Chapter 2. Savage contends, however, that if the agent is in this situation of uncertainty regarding consequences, then the correct response would be to refine the state space accordingly. For instance, if the agent is unsure of the result of breaking a rotten egg into the bowl containing five good eggs, then the “right” state space to use would not be ‘good’ and ‘rotten’, but rather ‘good and a rotten egg does not spoil the omelet’, ‘good and a rotten egg spoils the omelet’, ‘rotten and a rotten egg does not spoil the omelet’, and ‘rotten and a rotten egg spoils the omelet’. This more refined state space then resolves the uncertainty regarding the consequences of breaking the egg into the bowl fully, such that a unique consequence obtains at every state. In Chapter 2, the argument Savage made to the effect that uncertainty over consequences should be addressed by refinement of the state space will be called the *reduction argument*, and we will argue in Chapter 2 that it cannot be employed in all cases.

A further case Savage considers is that where the decision the agent makes leads to a further decision, such that the formulation of the act ‘break into bowl’ does not fully reflect the options the agent has. For instance, in the case of cooking an omelet the agent might care about what to do when the omelet is indeed spoiled, such as taking the family to the restaurant or eating toast for breakfast instead. In this case, Savage argues, the description of the act contained in the list of possible actions the agent constructed is not sufficiently detailed, such that the agent should replace the act ‘break into bowl’ with a set of acts such as ‘break into bowl, and in case of disaster have toast’, and ‘break into bowl, and in case of disaster take family to the restaurant’. One might call this kind of uncertainty *act uncertainty*. However, we will not consider this case in this thesis, since the set of acts is, in Savage’s theory, derived from states and consequences. Hence, a sufficiently fine-grained set of states and consequences will imply exhaustiveness of the act space.

Let us now turn to the concept of a large world, and let us begin by considering Savage’s definition of these, before proffering the view of large worlds advocated in this thesis. As pointed out above, Savage conceives of large world decision as those cases where the agent must "cross the bridge when he comes to it". Binmore (2007) gives the example of financial economics as a typical case of a large world in Savage's sense: this would be a case where it is not straightforward to see how an agent could rationally, or optimally, respond to the uncertainty he is faced with. More specifically, for typical large world cases it is impossible to construct a state space which satisfies Savage's definition thereof, namely such that it enumerates all aspects of the world leaving no relevant detail undescribed.

With respect to large worlds Savage argues, however, that the most sensible way of resolving such decision situations is to break them down into smaller decision problems which lend themselves to small world representations, thus resolving, step by step, the complicated large world problem. For instance, when we are faced with the complex problem of how much public money should be invested in mitigating climate change within the next twenty years, the most reasonable response would, according to Savage, be to construct a small world model constrained to the forecast horizon for which we have sufficient information, thus "confining attention to so small a world" that it is possible to find an optimal response to the problem thus obtained. So not all decision problems which may appear to require crossing the bridge when one comes to it really are so complex that they prevent rational responses.

According to Savage, then, a large world is a refinement of a small world. In the extreme case of an ultimately refined model, the large world becomes the *grand world*, which includes *all* aspects of a decision problem. The role of the grand world, in Savage's theory, is to peg the concept of optimality: a decision is optimal if it is optimal in a grand world, and a small world representation is suitable as a representation of the decision problem to the extent that the decision in the small world will cohere with the optimal decision made in the grand world. We will elaborate on these notions from a more technical point of view in Chapter 2.

The view of the large world this thesis uses is slightly different from the one Savage has in mind. In particular, in this thesis a large world is understood not as a refinement of a small world model, but as a model in which a small world representation is not feasible. The notion of the large world as used here refers to decision situations where a model *very similar* to the small world decision matrix can be constructed, but where the model explicitly admits uncertainty over some aspects of the small world model. The advantage of this conception of large worlds is that it permits modelling the attitudes an agent has to particular kinds, or sources, of uncertainty with precision – this is the aim of this thesis. We can then model decision problems where the agent can neither fully "look before he leaps" nor on the other hand must "cross the bridge when they come to it"; rather, the agent modelled here rationally responds to uncertainty.

An assumption which is crucial to this endeavour is that the agent is *consciously unaware* of the fact that the representation of the decision problem they construct may be underspecified in some aspects. Once we assume that the agent is consciously unaware, the agent can respond rationally to the uncertainty they are faced with. Indeed, without this assumption, it is not clear what, from a normative point of view, can be said about the agent's reasoning within large worlds.

Finally, it is important to now respond to the two questions to what extent we advocate the Bayesian view of rationality, and what our view of rationality is when the Bayesian framework is too restrictive. The answer to the first question is that Bayesianism embodies the right principles of rationality within the constraints of small world models, and should be employed when a small world decision problem can be constructed. In large world problems where the agent faces significant uncertainty, however, the Bayesian view may be too restrictive, necessitating separate requirements of rationality. Making some progress in specifying the rationality requirements in large worlds is the objective of this thesis.

1.6 Chapter conclusions

This chapter has presented the basic frameworks and concepts this thesis makes use of. In particular, we presented Savage's axiomatic framework and corresponding subjective expected utility theorem. Furthermore, we have explained the Anscombe-Aumann framework and its differences from Savage's theory. The baseline concept of rationality, namely Bayesianism, was explained. We then considered the two most prominent counterexamples against Savage's theorem, the Allais and Ellsberg paradoxes. Finally, we gave an introduction to small and large world models, facilitating our further discussion of uncertainty contained in

Chapter 2. This thesis proceeds as follows:

Chapter 2 first explains Savage's notions of the small and grand world and shows why ambiguity, option and state space uncertainty are incompatible with the small world representation; they may be seen as features of a "large world", an extension of Savage's small world model to cases of uncertainty. Furthermore, the chapter investigates the question to what extent it is possible to reduce these types of uncertainty to one another.

Chapter 3 turns to the topic of ambiguity, arguing in particular that the concept of ambiguity can not be captured accurately if ambiguity is defined subjectively, i.e. in terms of preferences. Subjective definitions of ambiguity may either under- or overestimate the presence of ambiguity, as ambiguity may not be revealed through preferences when it is present, or may be attributed to preference patterns which do not arise out of ambiguity. To solve this issue, we suggest an objective notion of ambiguity, by stipulating the existence of an exogenously given objective likelihood ranking over events. On our definition of ambiguity, careful distinctions between ambiguity and ambiguity attitude are feasible. Moreover, Chapter 3 argues that in situations of ambiguity, Savage's framework is too restrictive; ambiguity aversion should be permissible in objectively ambiguous decision problems.

Chapter 4 gives a formal model of option uncertainty, following Ghirardato (2001). In particular, the model generalises Savage's notion of acts, so that these are no longer functions from states of the world into consequences, but correspondences from states into consequences. We use Ghirardato's framework to show that option uncertainty aversion can be used as an explanation of status quo bias, the tendency that people prefer the status quo over other available alternatives. The two weak conditions that the status quo is not itself uncertain, and that agents are uncertainty averse, suffice to derive status quo bias. The model can be seen as rationalising status quo bias.

Chapter 5 contains a paper co-authored with Konstantinos Katsikopoulos and Gerd Gigerenzer at the Center for Adaptive Behavior and Cognition of the Max Planck Institute for Human Development, Berlin. Whilst the remainder of this thesis can be seen as addressing the question which axioms are reasonable under uncertainty, Chapter 5 asks which axioms are implied by a descriptively accurate model of choice, namely the priority heuristic. The paper gives an axiomatisation, and corresponding representation theorem, of a class of lexicographic models which includes the priority heuristic as a special case.

Chapter 6 explains and defends the pluralistic view of decision theory advocated in this thesis, discusses the role of heuristics under uncertainty and concludes with a discussion of the various applications of the notions of uncertainty developed in this thesis.

Chapter 2

Types of Uncertainty

2.1 Introduction

In the last chapter, we have introduced Savage's framework for small worlds, and identified the main characteristics a small world model must satisfy: the agent's decision problem can be cast in terms of a choice over actions, the outcomes of which hinge on an exogenously given state space. Savage's theorem shows that an agent whose preferences satisfy a number of basic postulates will act as if he were maximising the subjective expectation of utility, relative to a utility function on the set of consequences and a subjective probability function on the set of states of the world.

This chapter argues that many decisions cannot be cast in terms of a small world decision matrix; in particular, those where there is "too much" uncertainty to permit a small-world representation. Let us start with a concrete example: suppose you are the Head of State of Israel, and you must make a decision on the question whether your country should launch a military attack on Iran, on the grounds that you suspect Iran to be building nuclear weapons. There are many complex factors which would influence such a decision, for instance, how likely you think it really is that Iran is building nuclear weapons, and at what stage their development currently is. Also, one would want to predict as precisely as possible the ramifications of the decision to go to war: whether it is possible to find and destroy any potential nuclear missiles, how many lives would be lost, and how likely it is that the conflict might spill over to other states. One may also find it hard to evaluate how desirable each of these consequences are: how do we weigh up lives lost against the threat of a nuclear armed Iran? Finally, one may wonder whether there are any alternatives to going to war, and if so, how good these would be.

The only source of uncertainty consistent with the small-world representation of decisions is uncertainty regarding which state of the world prevails – this type of uncertainty will be called "state uncertainty" – but arguably this is not the only relevant type of uncertainty. In the real world, an agent is not faced with a decision problem, but must rather *construct* it (see, e.g., Ghirardato, 2001): states, consequences and acts are not usually given to the agent. Constructing a decision problem can, however, be highly complex. For instance, according to Savage (28, p.9), a state of the world is "a description of the world, leaving no relevant aspect undescribed". In some cases it may indeed be eminently simple to identify a state space which can be so-described, but in many others, a decision has to be made by the agent as to what counts as a "relevant" aspect and what does not. Similarly, the agent must *decide* on what to include in the set of acts.

Not all decision problems lend themselves to a straightforward small-world representation. Such decision situations are what Savage calls large worlds; cases where the uncertainty is too severe to admit subjective expected utility maximisation. Even in a large world, however, not all uncertainties are alike. The agent may face qualitatively different *kinds*, not just different severities, of uncertainty, and these may require different responses. For instance, the example of Israel's decision whether to launch an attack on Iran above demonstrates that there may be uncertainty over the correct probability distribution over states, as well as uncertainty over the consequences of launching an attack. These kinds of uncertainty may be perceived, by the decision-making agent, very differently: one could imagine that when the possible consequences of one's decision are that many people may lose their lives, then one would be particularly averse to any uncertainty over the consequences of one's actions. This feature of uncertainty over consequences would also suggest that the agent's attitude to this kind of uncertainty would be different to their attitude to uncertainty over the likelihood of states. In the following, we will flesh out this argument, characterising the different types of uncertainty which necessitate distinct treatment.

The chapter is structured as follows: Section 2.2 begins with Savage's own understanding of the small, large and grand worlds, and discuss the limitations of this account. This is followed, in section 2.3, by a classification of the different types of uncertainty characterised here. Section 2.4 then investigates the concept of ambiguity in light of our classifications of types of uncertainty, and considers the question to what extent ambiguity can be reduced to risk. Section 2.5 turns to the concept of option uncertainty, and discusses four different ways of understanding option uncertainty. The concept of state space uncertainty is investigated in section 2.6. Section 2.7 concludes.

2.2 Savage's notions of small, large and grand worlds

Savage distinguishes between small and large worlds using the following two proverbs: you are in a small world when it is possible to "look before you leap", and you are in a large world when you must "cross the bridge when you come to it" (Savage, 1954, p.16). Intuitively, planning ahead by maximisation of subjective expected utility, i.e. looking before you leap, is feasible only in situations where there are grounds to think that certain results would follow from one's actions; in contrast, if the decision situation is, for one reason or another, too complex, then one must make decisions as events unfold, i.e. by crossing the bridge when one comes to it.

Savage's also introduces the notion of the *grand world*, where this is to be understood as an infinitely refined version of the small world, a model in which no detail is elided. For instance, in the grand world consequences are to constitute "sure experiences of the deciding person", that is, they are no longer descriptions of things that happen to an agent in a particular state, but rather levels of satisfaction the agent experiences. On this strong notion of grand world consequences, it almost appears as though what Savage had in mind is that grand world consequences *are* in fact utility levels rather than descriptions of the state of the agent. It is not easy to make sense of this conception of the grand world. A grand world where consequences are mental states of a person will have acts and states which no longer lend themselves to a natural interpretation – how are we to describe a state which brings about a certain level of pleasure or pain? Moreover, in order to obtain consequences which are experiences of the person, we would be forced to subscribe to the idea that our experiences are *fully* determined by exogenously given states, which would imply a deterministic world view, a point to which we shall return shortly. The details of the grand world model will become clear through our discussion of the technical elements thereof.

One may also wonder about the relation between the grand world and the real world. It seems like Savage intended the grand world to be very close to the real world, since it is supposed to take account of all uncertainty which exists in the real world. However, if the grand world is to be a *model* of the real world, then of necessity some abstraction is required. But, as Box (1979) famously pointed out, "all models are wrong" since all models are abstractions – this must then hold too for the grand world model. So if the grand world abstracts from the real world, it cannot be *infinitely* refined as Savage intended, since only the world itself is infinitely refined. Perhaps we are to think of the grand world model as a model of the real world which abstracts only from strictly irrelevant aspects – but, prima facie, any aspect could become relevant to a decision problem, and hence, there is no principled reason to exclude any aspect of the real world from the grand world model. It is then unclear how we are to think of the grand world model; this leads Shafer (1986) to call Savage's grand world "an outrageous fiction". Since the grand world in its interpretation as an infinitely refined small world is so hard to pin down, we will follow Shafer (1986) in giving an example of a small world (Table 2.2) and a refinement thereof (Table 2.3) in order to illustrate Savage's formal account of small and grand worlds.

Let us now turn to Savage's formal framework of the small and grand world. Table 2.1 relates the concepts of the small and grand world model[1]. We will, in the following, explain each element of Table 2.1 in detail. In order to explain the relation between these concepts as clearly as possible, Tables 2.2 and 2.3 give an example of a typical small world and its refinement. In particular, Table 2.2 is the example we gave in Chapter 1 of a person cooking an omelet, and Table 2.3 is a refinement of it (Table 2.3 is taken from Shafer, 1986). In Table 2.3, the decision maker realises that his guest can distinguish between eggs that are less than 36 hours old – these eggs are called 'Fresh' – and eggs that are less fresh – these latter ones are called 'Stale'. Assuming that the first five eggs are all

[1]The notation we use here differs from that which we introduced in Chapter 1. This is in order to give a precise explanation of Savage's formal framework; however, the remainder of this thesis will work with the notation of Chapter 1.

Concept	*Grand world*	*Small world*	*Relation*
States	$S = \{..., s, s', ...\}$	$\bar{S} = \{..., \bar{s}, \bar{s}', ...\}$	$\bar{s} \subseteq S$
Events	$E = \{..., A, B, ...\}$	$\bar{E} = \{..., \bar{A}, \bar{B}, ...\}$	$\left[\bar{B}\right] = \cup_{\bar{s} \in \bar{B}} \bar{s} \subseteq S$
Consequences	$F = \{..., f, g, ...\}$	$\bar{F} = \{..., \bar{f}, \bar{g}, ...\}$	$\bar{f} = \mathbf{f} \subseteq \mathbf{F}$
Acts	$\mathbf{F} = \{..., \mathbf{f}, \mathbf{g}, ...\}$	$\bar{\mathbf{F}} = \{..., \bar{\mathbf{f}}, \bar{\mathbf{g}}, ...\}$	$\bar{\mathbf{f}} \subseteq \bar{\mathbf{F}} = \hat{\mathbf{f}} \subseteq \mathbf{F}$
Probability	$P : S \to [0, 1]$	$\bar{P} : \bar{S} \to [0, 1]$	$\bar{P}(\bar{B}) = P([\bar{B}])$
Utility	$U : F \to \mathbb{R}$	$\bar{U} : \bar{F} \to \mathbb{R}$	$\bar{U}(\bar{f}) = E(\bar{f})$

Table 2.1: The definitions of small and grand world concepts according to Savage (1954, p.84).

equally fresh, an omelet made with exclusively fresh eggs will be called a 'Nero Wolfe omelet'. Refinement of the small world matrix taking the freshness of eggs into account yields the decision matrix in Table 2.3. Notice that, since Table 2.3 is not an infinitely refined model, some of the concepts of Table 2.1 do not translate with exactitude into the examples of Tables 2.2 and 2.3; however, it is nevertheless instructive to study the relation between a small world and its refinement.

Let us begin with the *state space*. As Table 2.1 shows, the grand world state space is denoted S and has typical elements s, s' etc. The small world state space is denoted $\bar{S}$ with elements $\bar{s}, \bar{s'}$. A small world state $\bar{s}$ is, then, both an element of the small world state space, and a subset of the grand world state space S. We can see this easily by reference to Tables 2.2 and 2.3. Take, for instance, the small world state 'Good'. In the refinement, this comprises both of the more refined states 'Good and Fresh' and 'Good and Stale'. Therefore, the small world state 'Good' in Table 2.2 is a subset of the state space in Table 2.3.

Now turn to the concept of *events*. Events are subsets of the state space. The set of grand world events is denoted E, with typical elements A, B. In the small world, events are denoted $\bar{E}$, with typical elements $\bar{A}, \bar{B}$, and so forth. Now, a small world event is, of course, a grand world event. The small world event $\bar{B}$ is, in the grand world, an event $\left[\bar{B}\right]$ which collects all small world states $\bar{s} \in \bar{B}$, and is therefore a subset of the grand world state space S. For instance, the small world event 'Good' in Table 2.2, is in the grand world Table 2.3, a set of states {'Good and Fresh', 'Good and Stale'}.

Options	State	
	Good	Rotten
break into bowl	six-egg omelet	no omelet, and five good eggs destroyed
break into saucer	six-egg omelet, and a saucer to wash	five-egg omelet, and a saucer to wash
throw away	five-egg omelet, and one good egg destroyed	five-egg omelet

Table 2.2: A SMALL WORLD DECISION PROBLEM.

Options	State			
	Good		Rotten	
	Fresh	Stale	Fresh	Stale
break into bowl	six-egg Nero Wolfe omelet	six-egg ordinary omelet	No omelet and five good eggs destroyed	No omelet and five good eggs destroyed
break into saucer	six-egg Nero Wolfe omelet and a saucer to wash	six-egg ordinary omelet and a saucer to wash	Five-egg Nero Wolfe omelet and a saucer to wash	Five-egg ordinary omelet and a saucer to wash
throw away	Five-egg Nero Wolfe omelet and one good egg destroyed	Five-egg ordinary omelet and one good egg destroyed	Five-egg Nero Wolfe omelet	Five egg ordinary omelet

Table 2.3: A REFINEMENT OF THE SMALL WORLD DECISION PROBLEM.

The set of grand world *consequences* is denoted F with typical elements f, g etc. The set of small world consequences is denoted $\bar{F}$, with elements $\bar{f}, \bar{g}$. Small world consequences can then be understood as grand world acts. This is slightly counterintuitive, as one might think that grand world consequences are just more refined versions of small world consequences, just like grand world states are more refined versions of small world states. But Savage thought of grand world consequences as sure experiences of the deciding person, such as 'pleasure' or 'pain', rather than descriptions of the result of an act. Then, a small world consequence, such as 'six-egg omelet' can be understood as a grand world act as follows: it is a function from the very fine-grained, grand world state space to levels of pleasure or enjoyment experienced by the agent when consuming the six egg omelet. We can see, then, that the grand world consequence of experiencing a certain level of satisfaction is not just a more detailed description of the outcome 'six-egg omelet', but rather a consequence determined by the act of consuming a 'six-egg omelet' in a particular (mental) state. Although Table 2.3 is not a grand world, we can use the tables to see how a small world consequence can be seen as a grand world act. Take, for instance, the consequence 'six-egg omelet' in the small world matrix of Table 2.2. This can be seen as an act the consequences of which depend on the states 'Fresh' and 'Stale', then yielding the final consequences 'six-egg Nero Wolfe omelet' and 'six-egg ordinary omelet'.

Formally, Savage writes $\bar{f}(\bar{s})$ for the small world consequence of the small world act $\bar{\mathbf{f}}$ at the state $\bar{s}$. Then the notation Savage uses to convey the intuition that small world consequences can be understood as grand world acts becomes slightly counterintuitive. First, note that the small world consequence $\bar{f}(\bar{s})$ is both an element of $\bar{F}$ and an element of $\mathbf{F}$. Let us focus on the latter interpretation, so we have $\bar{f}(\bar{s}) \subseteq \mathbf{F}$. Now that we understand the small world consequence $\bar{f}(\bar{s})$ as an element of the set of grand world acts, it is clear that the consequences of this act must depend on grand world states. So then, Savage writes $f(s; \bar{s})$ for the grand world consequence of the small world consequence $\bar{f}(\bar{s})$ (understood as a grand world act) at the grand world state s. The counterintuitive aspect of this notation is the following: the notation $f(s; \bar{s})$ would suggest that the function $\mathbf{f}$ in fact depends on two separate variable variables, namely s and $\bar{s}$. But of course, s is an element of $\bar{s}$ as we explained above, so that $\mathbf{f}$ does not depend on two separate variables.

Finally, the set of grand world *acts* is denoted $\mathbf{F}$, with elements $\mathbf{f}, \mathbf{g}$. Grand world acts are, of course, functions from the set of grand world states S to the set of grand world consequences F. The set of small world acts is denoted $\bar{\mathbf{F}}$, with elements $\bar{\mathbf{f}}, \bar{\mathbf{g}}$. Similarly, small world acts are mappings from the set of small world states $\bar{S}$ to the set of small world consequences $\bar{F}$. Formally, Savage then argues that each small world act $\bar{\mathbf{f}}$ uniquely gives rise to a grand world act which he calls $\hat{\mathbf{f}}$ as follows: $\hat{f}(s) \equiv f(s, \bar{s}(s))$. We can then see that the grand world act $\hat{\mathbf{f}}$ depends on the small world state $\bar{s}$, e.g. 'Good', and the particular grand world information 'Fresh' contained in s. Whilst Savage's notation is, as he acknowledges, counterintuitive, the intuition that each small world act uniquely gives rise to a grand world act is easily conveyed by Tables 2.2 and 2.3.

With this formal framework at hand, Savage turns to the question under what conditions a small world model is an adequate representation of the more complex grand world. When a small world model is adequate, it is called a *real microcosm*, and if it is not, a *pseudo microcosm*. There are two conditions, listed in the last two rows of Table 2.1, which must be satisfied for a small world model to be a real microcosm: intuitively, the small world model must yield a probability function over the state space and a utility function over consequences, both of which agree with those obtained from the grand, i.e. infinitely refined, world. In particular, the probability distribution obtained from the grand world, denoted P, assigns a probability $p \in [0, 1]$ to each state in S, and the probability distribution obtained from the small world, denoted $\bar{P}$, assigns a probability $\bar{p} \in [0, 1]$ to each state in $\bar{S}$. As we have seen, small world states are subsets of the grand world state space, and each small world event gives rise to a corresponding grand world event. The probability distributions obtained from the small and grand world should assign the same probability to any event. Secondly, the utility function obtained from the grand world is denoted U and assigns a real number to every grand world consequence, and the utility function obtained from the small world is denoted $\bar{U}$, and it assigns a real number to every small world consequence. As we have seen, every small world consequence can be seen as a grand world act. Then, the utility function obtained from the small world is adequate if it is equivalent to the expected utility of the corresponding grand world act. As Savage showed, if the axioms of subjective expected utility we listed in Chapter 1 are satisfied, then the small world utility function always satisfies this requirement.

This is not so for the requirement on probabilities. As Savage himself realised, the probability distribution resulting from the analysis of a small world decision matrix may fail to agree with that obtained from the grand world. In particular, there are two different ways in which probabilities of small world states can be computed: first, since the small world state space is a partition of the grand world state space, the probability distribution over the small world states can be obtained by calculating the marginal distribution of the distribution over the grand world states. For instance, suppose that in Table 2.3, $p(\text{Good and Fresh}) = \frac{1}{6}$ and $p\,(\text{Good and Stale}) = \frac{1}{6}$. Then the probability of the state 'Good' in Table 2.2 should be $\bar{p}\,(\text{Good}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$. Second, one can elicit the probability distribution over the small world states from the agent's preferences over small world acts, as explained in Chapter 1. However, these two methods of computing the distribution over states may fail to yield the same result. It is due to this fact that Savage postulated the existence of an *infinitely* refined grand world, such that small world probabilities are correct if they cohere with those computed from the infinitely refined grand world. In other words, the correct probability distribution for the agent to use is 'pegged' by the grand world model, which allows Savage to call a small world adequate whenever the probability distribution obtained from it coheres with that which is had from the grand world. As Shafer (1986) remarks, however, whether or not one's small world probabilities do indeed cohere with their equivalents in the grand world is impossible to verify.

We have seen, then, that Savage's notion of the grand world is in fact very demanding and hard to conceptualise. But maybe such a strong notion of the grand world is not required to talk about uncertainty. As we will argue in the following, there are certain types of uncertainty which are incompatible with a small world treatment, but which can nevertheless be characterised precisely. These types of uncertainty can be seen as features of a more modest grand world than Savage's. In the next section, we will start with a general overview of the topic of uncertainty, and then focus more specifically on the taxonomy of uncertainty this chapter advocates.

	State		
Acts	s_1	$\dots$	s_n
f	x_{f1}	$\dots$	x_{fn}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
h	x_{h1}	$\dots$	x_{hn}

Table 2.4: Savage's decision problem.

2.3 Types of Uncertainty

A good starting point for our discussion of uncertainty is Luce and Raiffa's (1957, p.13) classification, which distinguishes between situations of *certainty*, i.e. cases where each action leads invariably to a specific outcome, *risk*, which are cases where an action leads to one of a set of possible specific outcomes, where each outcome occurs with an objectively known probability, and *uncertainty*, namely cases where actions have sets of possible consequences, but where the probabilities of these outcomes are completely unknown. Similarly, Knight (1921, p.19) defines risk and uncertainty as follows:

> *"But Uncertainty must be taken in a sense radically distinct from the familiar notion of Risk, from which it has never been properly separated. [...] The essential fact is that "risk" means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character. [...] It will appear that a measurable uncertainty, or "risk" proper, as we shall use the term, is so far different from an unmeasurable one that it is not in fact an uncertainty at all."*

Under Luce and Raiffa's definition, uncertainty refers to the case where probabilities are "completely unknown", whereas under Knight's definition, uncertainty refers to cases where uncertainty is "unmeasurable". Both situations will be called *ignorance* in the present context, the absence of any probabilistic information.

This chapter advocates a more wide-ranging classification of uncertainty than those suggested by Luce and Raiffa (1957) and Knight (1921). To this end, consider Savage's simple setting of a small world decision matrix, as characterised in Table 2.4. In general, the most basic form of uncertainty an agent faces is that of what to do. In order to use Savage's framework, an agent must be able to reduce that uncertainty to uncertainty over what the true state is. This source of uncertainty, which we call *state uncertainty*, is the only source of uncertainty compatible with Savage's theory, and it pertains exclusively to an exogenously given state space. Savage's theory then exacts that the agent ought to form subjective beliefs over the state space whose objective probability is unknown, such that the agent can then compare acts based on their subjective expected utility.

However, this view of uncertainty is restrictive, as it precludes other sources of uncertainty. In particular, this thesis advocates a distinction between different types of uncertainty along the following dimensions (see Bradley and Drechsler, forthcoming):

1. Type. The first distinction relates the type of uncertainty to the *nature* of the *judgement* being made. We distinguish three basic types of uncertainty: conceptual, empirical and ethical, corresponding to three types of question we can ask about them.

1. **Conceptual** uncertainty is uncertainty about what is possible or about what *could* be the case. For instance, in thinking about how to represent a decision problem we might be unsure as to what the possible states of the world are or what possible consequences could follow from the choice of an action. This uncertainty thus concerns the make-up of the space of states and consequences, and hence what actions are logically possible. (In the most extreme case of conceptual uncertainty, the agent is *unaware* of certain states and/or consequences).

2. **Factual / empirical** uncertainty is uncertainty about what *is* the case (or has been or would be the case). It arises in connection with our descriptive judgements. Such uncertainty can be present even if all conceptual uncertainty is resolved, since we may be sure about what the relevant possible states are, but unsure as to which is the one that actually holds.

3. **Ethical**[2] uncertainty is uncertainty about what is desirable or what *should* be the case. It arises in connection with our evaluative judgements. Ethical

[2]The term *ethical* uncertainty may be understood by some readers as implying uncertainty pertaining to moral values. This is not the use intended in this thesis: ethical uncertainty refers to uncertainty pertaining to value *in general*, not just moral value. The terminology goes back to Ramsey's (1926) seminal paper "Truth and Probability".

uncertainty can be present even if all conceptual and empirical uncertainty is resolved: we may be sure what the state of the world is, but unsure what value to attach to the consequences that follow from performing an action when that state is the prevailing one.

2. Severity. A separate dimension relates to the *difficulty* the agent has in making a judgement about the uncertain prospects they face. We classify severity by reference to the situation which gives rise to it. In order of decreasing severity:

1. **Ignorance**: When the agent has no judgement-relevant information.
2. **Ambiguity**: When their information allows for some assignment of beliefs, but is insufficient to assign precise probabilities to *all* prospects.
3. **Mild uncertainty**: When the agent has sufficient information to assign a precise probability to all prospects.
4. **Certainty**: When the value of the judgement is given.

The case of "mild uncertainty" comprises both the cases where the agent can assign subjective or objective probabilities to prospects. "Risk", which is commonly understood as the availability of objective probabilities, may be regarded as the limiting case of mild uncertainty.

In the above classification, we may understand the "type" dimension of uncertainty as listing different *sources* of uncertainty. An agent who faces conceptual uncertainty is uncertainty about how best to *model* a given decision problem; the agent is unsure what states and consequences (and, hence, what acts) are feasible. In contrast, under empirical / factual uncertainty, the agent is unsure not about how to model a given decision situation, but rather about the situation itself – the agent's uncertainty concerns the way the world is, and how it will evolve; the agent is therefore unsure about *objective facts.* Finally, under ethical uncertainty, the agent is uncertain with respect to what values best reflect their beliefs and / or desires; these are *subjective facts.* Within the "type" dimension of uncertainty, we do not intend to imply that any particular type of uncertainty poses greater or lesser difficulty to the agent than another. The "severity" dimension of uncertainty concerns the degree, or extent, to which an agent is uncertain. Severity ranges from ignorance, where the agent has no information concerning the likelihood of events, to certainty, where the agent knows that a particular event is true.

Of course, the "severity" dimension measures uncertainty on a scale, such that ignorance is a situation with greater severity of uncertainty than, for instance, mild uncertainty.

In the above classification, the dimensions "type" and "severity" are to be thought of as orthogonal; that is to say, an agent may face any combination of type of uncertainty and severity of uncertainty. For instance, we can associate ambiguity, the case where the agent is unable to assign a precise probability to every state, with conceptual, factual / empirical and ethical uncertainty as follows: an agent may perceive a decision problem as ambiguous when they are uncertain as to what states are feasible. As a result, the agent may not be able to assign probabilities to the states they are aware of; the agent then faces ambiguity of the conceptual kind. Secondly, an agent may perceive ambiguity as a result of their ignorance of the generating distribution; the agent then faces ambiguity of the empirical / factual type (for instance, Ellsberg's experiment pertains to ambiguity of the empirical / factual kind). Finally, an agent may perceive ambiguity as a result of ethical uncertainty, i.e. whenever the agent is unsure as to what probability distribution best represents their belief; the agent then faces ambiguity of the ethical uncertainty type. Similarly, we can associate any other type of uncertainty with any severity.

The kind of uncertainty Savage's theory applies to, namely *state uncertainty*, is *mild uncertainty* of the *empirical / factual* type: it pertains to the question what the true state is; the agent then holds a subjectively known probability distribution over the state space. However, it is the aim of this chapter to characterise, along the dimensions given above, three additional types of uncertainty. Firstly, *ambiguity*, the case where the agent's uncertainty is more *severe*. Secondly, *option uncertainty*, the case where agents are uncertain as to what consequence follows from the exercise of an act at a particular state, they therefore envisage several consequences as possible at every state. Finally, *state space uncertainty*, which refers to the case where the state space is not exhaustive, permitting for unforeseen contingencies.

In this chapter, we will characterise each of these types of uncertainty along the dimensions identified above. This raises, however, the further question to what extent ambiguity, option and state-space uncertainty are genuinely separate: can we reduce option uncertainty to ambiguity? Can we convert state space uncer-

	red	black	yellow
e_1	\$100	\$0	\$0
e_2	\$0	\$100	\$0
e_3	\$100	\$0	\$100
e_4	\$0	\$100	\$100

Table 2.5: THE ELLSBERG PARADOX

tainty into ambiguity or option uncertainty? In the following, we will argue that there is indeed some scope for reduction, but in each case a type of uncertainty exists which cannot be reduced. Also, we will argue that even though reduction is feasible in principle, total uncertainty is preserved. The uncertainty surrounding any decision problem has to be addressed on some level of the analysis, since it cannot be done away with.

Before we proceed to a more detailed treatment of these types of uncertainty, let us investigate the question how this account of uncertainty relates to Savage's conception of small and grand worlds. In the previous section, we have seen that Savage's notion of a grand world is a very demanding one, in the sense that it is an infinitely refined version of the small world, which makes it a scenario even Savage himself found hard to conceptualise. We have also noted that the three types of uncertainty discussed here are incompatible with the small world setting. However, nor are they cases of the grand world Savage had in mind: as we shall see, these are minimal extensions to the small world setting, rather than features of an infinitely refined model. For clarity of exposition, we will therefore call these scenarios *large world* decision situations, since they are, one might argue, intermediate between the small and the grand world. Let us now turn to a detailed analysis of these types of uncertainty.

2.4 Ambiguity

In Chapter 1, we have introduced the Ellsberg paradox as a violation of Savage's axiom P2, the sure-thing principle, and showed that ambiguity aversion can explain the Ellsberg paradox. Let us now proceed to take this further, by asking what characteristics ambiguity has. For ease of reference, the Ellsberg paradox is reproduced in Table 2.5.

First, note that in the Ellsberg paradox, a small world decision matrix of the form given in Table 2.4 is available to the agent. In the Ellsberg setup, the agent is given well-defined acts to choose from, where these yield precise outcomes at each state, and are therefore functions from an exogenously given state space ('red', 'black', 'yellow') to a set of outcomes ('\$0', '\$100'). In each state, a unique outcome follows with certainty, and the state space is exhaustive, such that no unforeseen contingencies can occur. Given that the agent has a small world decision matrix at hand, subjective expected utility maximisation would be feasible in principle. We will argue in the following that this is not so under option- and state space uncertainty, each of which is incompatible with a small world representation.

Secondly, note that ambiguity concerns the refinement of the agent's probabilistic information *relative* to the refinement of the state space. An agent will perceive ambiguity only if the state space is perceived to be more fine-grained than the probabilistic information the agent has. For instance, in the Ellsberg paradox, the agent knows objective and precise probabilities for the events $\{\text{'red'}, \text{'black or yellow'}\}$, namely $p(\text{red}) = \frac{1}{3}$ and $p(\text{black or yellow}) = \frac{2}{3}$. As we can see, the agent's probabilistic information is coarser grained than the state space $\{\text{'red'}, \text{'black'}, \text{'yellow'}\}$.

It is also easy to see that the absence of precise probabilities over the states 'black' and 'yellow' would be irrelevant if all acts yielded the same consequences in the states 'black' and 'yellow'. For instance, if all acts e_i with $i \in \{1, 2, 3, 4\}$ yielded $e_i(\text{black}) = e_i(\text{yellow}) = \$100$, then the agent would be indifferent between betting on black and betting on yellow, and so the expected utility of the acts e_i would be independent of the probability distribution over 'black' and 'yellow'. At the risk of belabouring the obvious, ambiguity becomes decision-relevant only if the state space is more fine-grained than the probabilistic information the agent has *and* if it is the case that which consequence comes about hinges on the so specified state space.

Thirdly, ambiguity regarding objectively given probability distributions is *empirical/factual uncertainty*, in the sense that it is uncertainty over the question what *is* the "right" probability distribution over the state space. In Ellsberg's experiment, there is an (objective) fact of the matter of what the correct probability distribution over the state space is, and this fact is determined before the agent faces the decision problem. Were the agent to play the Ellsberg gambles several times, he could observe the relative frequencies of black and yellow balls, so that the ambiguity would disappear.

One can imagine, however, situations where the agent perceives *subjective ambiguity*, in the following sense: when faced, for instance, with the task of assigning a unique probability to the event that the horse *Silver Charm* wins the Kentucky Derby, the agent might not hold a sharp subjective belief regarding this event. Perhaps the agent finds his beliefs best represented by a probability interval. In this case, ambiguity would be *ethical uncertainty*, since the agent is now unsure which subjective belief he *should* hold.

Let us now consider the question whether ambiguity is genuinely distinct from *mild uncertainty*, the case where the agent has access to either a subjective or objective probability distribution over the state space. Indeed, Bayesians would argue that there is no need to treat ambiguity as a separate category of uncertainty, since all uncertainty can and should be quantified in a single additive probability distribution over the state space, such that ambiguity is reduced to mild uncertainty. There are at least two ways in which this stance can be made precise. The first would be to argue, as de Finetti (1977) does, that ambiguity is meaningless, since all probabilities are equally well known to ourselves (Camerer and Weber, 1992). Under this *strong subjectivist* view, any consistent assignment of probabilities to the events 'black' and 'yellow' in Ellsberg's paradox will be defensible.

A second interpretation of ambiguity is that it can be expressed as a *second order probability distribution*. For instance, in Ellsberg's paradox the probability of drawing a black ball is in the range $[0; \frac{2}{3}]$. Then, we can assign a second order probability distribution over the values in this interval, interpreted as the likelihood that each of the possible distributions is the correct one[3]. Since in Ellsberg's problem there is no information about the likelihood of each distribution, using the principle of insufficient reason[4] would lead one to assigning a

[3] Indeed, de Finetti (1937) shows that if one interprets objective probabilities as limiting relative frequencies, then every subjective probability is a second-order probability of the objective probability distribution.

[4] The principle of insufficient reason holds that when there are n mutually exclusive and collective exhaustive events, and the agent's information regarding their likelihood is symmetric, then the agent should assign probability $\frac{1}{n}$ to each of the events.

uniform distribution over the values in the interval. By computing the expected value of the second order distribution, one can reduce ambiguity to mild uncertainty; then, the second order information is quantified within the first order distribution. This procedure would *reduce* ambiguity to mild uncertainty, thereby enabling (subjective) expected utility reasoning.

The arguments against the strong subjectivist view as well as the reductionist stance are well-known. There are two arguments which are typically made. The first draws on the example we gave in Chapter 1 of a person betting on a coin for which the chance of heads and tails are known to be equal versus betting on a coin with unknown probabilities. Most people would prefer to bet on the first coin, since the probability assignment was made on the basis of observed frequencies rather than on the basis of the symmetry of the agent's ignorance. Schmeidler (1989) expresses this as follows:

> *The probability attached to an uncertain event does not reflect the heuristic amount of information that led to the assignment of that probability.*

Schmeidler's argument is based on observations about human cognition: in many situations, there is insufficient information for the agent to form a unique subjective probability distribution. Returning to the example we gave in the introduction of this chapter, suppose we ask an agent to assess the likelihood of the two states 'Iran is building nuclear weapons' and 'Iran is not building nuclear weapons'. Bayesian reasoning would require the agent to assign a unique point in the real unit interval to these states, yet it does not seem to be a requirement of rationality to do so. In such situations, too much information seems to be lost by doing so.

A second argument against the subjectivist / reductionist stance comes from the descriptive observation that agents appear to be averse to ambiguity. This is shown most clearly in Ellsberg's two urn example. In the experiment, the first urn contains 50 black and 50 red balls, and the second urn contains 100 balls which are all either red or black. In experiments, people prefer betting on a red ball from the first urn to betting on a red ball from the second urn, and also prefer betting on a black ball drawn from the first urn to betting on a black ball drawn from a second urn. However, people are indifferent between betting on red or black from the first urn, and also indifferent between betting on red or black from the second urn. These results would suggest that agents are averse to ambiguity, as they prefer betting on known chances rather than unknown ones. Since some would argue that these results are errors of reasoning, tests of the robustness of these results have been conducted, in which subjects had the opportunity to reverse their choices. After conducting such tests, MacCrimmon (1968) argues that the original (ambiguity-averse) choices were indeed mistakes, and Slovic and Tversky (1974) argue that they were not.

In Chapter 3 of this thesis, we will investigate the question whether ambiguity aversion is irrational in greater detail. The present discussion was intended to show, however, that the reductionist stance can be granted only at the cost of both cognitive unease and descriptive inaccuracy. However, arguing against the Bayesian view requires relaxing the rationality requirements on the agent; it is not easy to see, however, how this can be done in a principled way. Chapter 3 will attempt to give an answer to these questions.

2.5 Option Uncertainty

In Savage's small-world representation of a decision problem actions are associated with definite consequences, one for each state of the world. These consequences are, in Savage's words,"sure experiences of the deciding person" (Savage, 1954), and the description of them includes all decision-relevant aspects. But in real decision problems we are often unsure about what consequence follows from a particular action at a particular state, and this uncertainty affects our decision-making. For instance, we may be uncertain whether taking an umbrella will certainly have the consequence of keeping us dry in the event of rain. Perhaps the umbrella has holes, or the wind will blow it inside out or the rain will be blown in from the sides. Uncertainty of this kind is an endemic feature of decision making, for it is rarely the case that we can predict consequences of our actions in every detail. For most decision situations, the precise consequence will be irrelevant. However, in some cases the details of the consequences will matter to the decision maker to the extent that his choice of act hinges on these details. We will call situations where the agent does not know what consequence follows from an action at a particular state situations of *option uncertainty*, and we will discuss its implications in the following.

There are at least three different reduction strategies one might pursue in connection to option uncertainty. Firstly, reducing option uncertainty by refinement of the state space, secondly, by treating acts and consequences as primitives and viewing states as functions of acts and consequences, and thirdly, by interpreting option uncertainty as uncertainty over the value of consequences. A final view on option uncertainty is to model it directly as an extension of Savage's framework for small worlds, namely by re-defining acts as correspondences from states to sets of consequences. In the following, I will explain each of these views and discuss their respective merits.

Let us begin with what is perhaps the most common response a decision theorist would make to option uncertainty, namely the view that uncertainty over the consequences of actions can be addressed by refinement of the state space until all contingencies are taken care of. This view was, in fact, advocated by Savage, as the discussion of small worlds contained in Chapter 1 showed. In connection with this view, consider the following example: suppose I am throwing a ball, and the consequence of this action is that it lands in a particular place, but I am uncertain as to where exactly it will land. Then the reductionist might argue that given sufficient information regarding the speed and direction of wind, the air pressure, the mass of the ball, the angle at which the ball was propelled, and so forth, it will be possible to predict where exactly the ball will land. However, there are at least two reasons why this strategy will not work on all occasions. Firstly because according to our best scientific theories the world is not purely deterministic. Only if it were deterministic would it be the case that the precise conditions under which a ball is thrown do determine where exactly it will land; in the absence of a deterministic set-up, however, such claims cannot be made. A second reason why the reduction strategy might fail is that even if we are in a purely deterministic set-up, it may be subjectively impossible for the decision maker to conceive of and weigh up all the relevant contingencies which need to be taken account of in order to predict where the ball will land. The reasoning capacities necessary to specify such a fine-grained state space are most likely well beyond a human being's cognitive capacities. Savage (1954, p.16) himself conceded that it is

"utterly beyond our power to plan a picnic or to play a game of chance

	State		
Acts	$s(a_i, x_{1,i})$	$\dots$	$s(a_i, x_{m,i})$
a_1	x_{11}	$\dots$	x_{1n}
$\vdots$	$\vdots$	$\ddots$	$\vdots$
a_m	x_{m1}	$\dots$	x_{mn}

Table 2.6: A SMALL WORLD WITH STATES AS FUNCTIONS OF ACTS AND CONSEQUENCES.

in accordance with the principle [of considering all relevant contingencies], even when the world of states and set of available acts to be envisaged are artificially reduced to the narrowest reasonable limits".

Finally, even if we granted that a decision maker were able to conceive of such fine-grained states, it would most likely be impossible for the agent to assess their likelihood, in which case the agent is faced with a decision problem under ambiguity. The reduction strategy then does not eliminate uncertainty, but much rather converts it into uncertainty over the likelihood of states.

A second view on option uncertainty is to take acts and consequences as primitives, and to define states as functions of these[5]. Letting a_i denote acts and x_{ij} the consequences of act a_i, we can write $s(a_i, x_{ij})$ for the state that maps action a_i into consequence x_{ij} (see Table 2.6). A state then specifies the conditions sufficient to bring about a consequence with certainty. This conception of states has been proposed, for instance by Fishburn (1970) in the economic literature, and Lewis (1981) in the philosophical one. Fishburn's model is in fact designed to treat not option, but state space uncertainty (which we discuss in the next section), namely the case where agents have incomplete knowledge regarding the state space, and hence rationally construct it from acts and consequences. Then, a state just gives the conditions under which a particular utility level is achieved (Dekel, Lipman, and Rustichini, 1998). Similarly, Lewis interprets states as "dependency hypotheses" – maximally specific propositions about the conditions under which an act brings about a particular consequence. Many causal decision theorists follow Stalnaker's (1981) suggestion that states should be interpreted as a conjunction of conditional sentences of the form 'If action a_1 were performed then consequence x_{11} would follow; if action a_m were performed

[5]In Savage's theory, states and consequences are primitive notions, and acts are defined in terms of these.

	State		
Acts	s_1	$\dots$	s_n
a_1	$x(a_1, s_1)$	$\dots$	$x(a_1, s_n)$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
a_m	$x(a_m, s_1)$	$\dots$	$x(a_m, s_n)$

Table 2.7: A SMALL WORLD WITH COARSENED CONSEQUENCES.

then consequence x_{m1} would follow; if ...'. Then, the state of the world just specifies the conditions under which the conjunction of these conditional sentences is true (Bradley, *Decision Theory with a Human Face*, forthcoming).

This reduction strategy will convert uncertainty over outcomes into uncertainty regarding what state will suffice to bring about a particular consequence with certainty. Using Stalnaker's definition of states, the agent would be uncertain under what conditions the conjunction of conditional sentences that describe it are true. However, this reduction strategy comes at the cost of an increase in the severity dimension of uncertainty, since the agent's probabilistic information may now be coarser-grained than the state space[6]. The agent then faces a decision problem under ambiguity rather than option uncertainty.

A third reduction strategy would be to coarsen the description of the consequence sufficiently to be certain that it will follow from a particular act at a particular state. This is the strategy Savage advocated, for he remarks (Savage, 1954, p.84): *"I therefore suggest that we must expect acts with actually uncertain consequences to play the role of sure consequences in typical isolated decision situations"*. Pursuit of this strategy leads to a small-world representation as in Table 2.7, where $x(a_1, s_1)$ is the consequence of act a_1 at state s_1. The description of $x(a_1, s_1)$ is now assumed to be less than maximally specific.

Coarsening the consequences until they are sure to follow in a particular state will convert option uncertainty into *ethical uncertainty*, since now, we may not be sure what value to attach to a consequence which is so described. For instance, consider the act of taking an umbrella in a rainy state. Then we can be sure that the umbrella will keep our head from getting wet, but it may or may not protect

[6]Notice that by an increase in the severity dimension, we do not wish to imply that there is greater uncertainty in the decision problem overall once the reduction is performed. We do wish to imply that whilst total uncertainty is conserved, the reduction strategy implies that the agent will have greater difficulty in assigning probabilities to states.

	State		
Acts	s_1	$\ldots$	s_n
a_1	$\{x_{11}^1, x_{11}^2, ..., x_{11}^s\}$	$\ldots$	$\{x_{1n}^1, x_{1n}^2, ..., x_{1n}^t\}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
a_m	$\{x_{m1}^1, x_{m1}^2, ..., x_{m1}^u\}$	$\ldots$	$\{x_{mn}^1, x_{mn}^2, ..., x_{mn}^v\}$

Table 2.8: A small world with acts as correspondences from states into consequences.

out feet. Then the value of the consequence of taking an umbrella in a rainy state may depend on factors (such as our feet getting wet) which by assumption we are unable to identify. Whatever (single) utility value we attach to such a coarse consequence will be surrounded by uncertainty.

A final view on option uncertainty is to enumerate all feasible consequences which the agent thinks might follow from an act at a particular state. Then no further uncertainty surrounds these fine-grained consequences, such that each consequence is to be understood as a "sure experience of the deciding person". An act then yields a set of possible consequences at a particular state, where only one of the set of consequences will be the true consequence. This strategy leads to a small world matrix as in Table 2.8. This case has received some attention in the literature: the case of option uncertainty can be modelled in a Savage framework by replacing Savage's notion of acts as functions from states to consequences by a notion of acts which understands these as correspondences from states into sets of consequences. A model which pursues this strategy has been given by Ghirardato (2001), and this model will be discussed in greater detail in Chapter 4. An epistemic approach to what we call option uncertainty has been taken by Mukerji (1997).

In summary, we have seen that 'pushing' option uncertainty into the state space by refinement leads to an increase in the severity dimension of uncertainty, such that the agent must make decisions under ambiguity rather than mild uncertainty[7]. A similar argument holds for the case where states are re-defined as function of acts and consequences. Conversion of option uncertainty to ethical

[7]Notice that by an increase in the severity dimension, we do not wish to imply that there is greater uncertainty in the decision problem overall once the reduction is performed. We do wish to imply that whilst total uncertainty is conserved, the reduction strategy implies that the agent will have greater difficulty in assigning probabilities to states.

uncertainty implies that the utility value of consequences becomes uncertain. The final strategy illustrated here does not reduce option uncertainty to any other kind of uncertainty, and represents feasible consequences individually; the uncertainty is then addressed directly. From this discussion we can see that no reduction strategy eliminates uncertainty, but rather just moves the uncertainty around in the decision matrix, so to speak – at one level of the analysis, the uncertainty must be addressed.

The belief that reduction would eliminate uncertainty is, however, not the only fallacy the reductionist may commit. A second one would be to think that the reduction will leave the decision problem unchanged. We have seen earlier in this Chapter in our discussion of Savage's conception of small and grand worlds that the probability distribution over the state space may change with refinement, such that the expected utility of acts computed using a small world model may not cohere with that computed from its refinement. Since the reduction strategies work using a refinement of the state space, the same effect may occur. Moreover, a large body of empirical evidence on framing effects demonstrates that preferences are generally not invariant under different representations of decision problems (see, e.g. Tversky and Kahneman, 1981). Framing effects may occur in the case at hand particularly since the reduction strategies convert option uncertainty into different *kinds* of uncertainty – ambiguity or ethical uncertainty – which may yield different psychological responses from agents. Chapter 3 of this thesis will investigate the topic of option uncertainty further.

2.6 State Space Uncertainty

In Savage's framework, the state space is a primitive of the theory, and is exogenously given. Indeed, the principle that the agent is supposed to conceive of all relevant contingencies can be seen as a basic tenet of Bayesianism (Gilboa, 2004, p.17). In real decision problems, however, a state space the elements of which are mutually exclusive and collectively exhaustive may not be given, or may be hard to construct. There are, in the real world, events which most people would argue are *unforeseen contingencies*, eventualities that even the educated decision maker fails to anticipate. For instance, natural disasters, such as the recent tsunami and subsequent nuclear meltdown in Japan are events most agents would have omitted as a potential contingency in their decision problem.

There are a number of ways in which the term “unforeseen contingencies” has been used in the literature. We will here distinguish two understandings of the term: First the case where the state space is insufficiently fine-grained, and secondly the case where a state is omitted from the state space entirely. In the following, we will investigate each of these interpretations and discuss to what extent they can be reduced to ambiguity or option uncertainty. But before we do so, let us briefly consider the connection between Savage’s framework and unforeseen contingencies. In particular, whilst Savage’s state space permits for the former case of state space uncertainty, it rules out the latter.

In the beginning of this chapter, we explained the connection between small and grand worlds, where the grand world state space is an exhaustive list of all feasible contingencies. Then, in the grand world, states are complete descriptions of all contingencies, in the sense that at a so-described state, a particular consequence follows with certainty. By construction, the grand world state space is exhaustive. In the small world, the agent only considers a partition of the grand world state space. Then, of necessity there is variation in individual small world states which the model does not capture. This variation can be either irrelevant or unforeseen by the agent. It is only through this variation in the small world states that unforeseen contingencies can occur in a Savage framework. Unforeseen contingencies can occur in Savage’s framework when the small world state space is insufficiently fine-grained. In contrast, since the state space in Savage’s theory is assumed to be exhaustive, there is no single state which can be elided entirely.

Let us now focus on the former case, namely where unforeseen contingencies come about through omission of decision-relevant details in the description of the states, and consider the question whether this is separate from ambiguity. Suppose, for instance, that I am interested in whether I should take an umbrella with me or not, and in my deliberation I consider the state space $\mathcal{S} = \{\text{‘sunny’},\text{‘rainy’}\}$. However, my decision would in fact best be represented using a state space which includes details about whether it is windy or not, as follows: $\mathcal{S}' = \{\text{‘sunny} \wedge \text{windy’}, \text{‘sunny} \wedge \neg\text{windy’}, \text{‘rainy} \wedge \text{windy’}, \text{‘rainy} \wedge \neg\text{windy’}\}$. Then, by using the state space $\mathcal{S}$ rather than $\mathcal{S}'$, I treated the states ‘sunny $\wedge$ windy’ and ‘sunny $\wedge\neg$windy’, as a single state, which I called ‘sunny’. But this does not seem substantively different from ambiguity, for the following reason: the consequences of acts hinge on the more fine-grained state space, whilst the states themselves are coarser-grained. Then, the agent's beliefs are formed only over the coarser state space, 'rainy' and 'sunny', rather than the full state space. In our discussion of ambiguity, we argued that one of the characteristics of ambiguity is that it is perceived when the agent's probabilistic information is coarser than the state space on which consequences hinge. This, however, is the case when we consider state space uncertainty as the omission of decision-relevant details from the description of the states: If unforeseen contingencies come about through an insufficiently fine-grained state space, then the agent holds relatively coarse-grained probabilistic information relative to the state space on which consequences depend. So this case does not seem substantively different from ambiguity. Notice, however, that this conception of unforeseen contingencies is popular in the literature. For instance, it is used by Ghirardato (2001), Modica et al. (1998), Skiadas (1997) and Walker and Dietz (forthcoming).

A second conception of state space uncertainty is the case where the agent fails to foresee a state entirely, rather than eliding details of its description. It may help to first clarify what it means to fail to foresee a state. In particular, the interpretation intended here is *not* the case where the agent (erroneously) attaches the probability zero to a possible event, since this interpretation would be compatible with the subjectivist view, under which a decision is optimal if it is made consistently with the agent's personal beliefs and desires. If the agent believes that a state is impossible and acts accordingly, then his decision is by definition optimal. Hence, this case does not require a new model. Furthermore, an unforeseen contingency is also not a state which the agent is unaware of: that would be the extreme case where the agent does not know what the concept of that state means. Much rather, an unforeseen contingency is a case where the agent has just not thought to include the given event in the state space (Dekel, Lipman and Rustichini, 1998). In this sense, state space uncertainty is empirical/factual uncertainty as we have previously characterised it: it concerns what states are possible. The agent faces, however, no ethical or conceptual uncertainty, since he does not face uncertainty over values, and is in principle aware of all contingencies.

There are, again, two ways in which an agent can elide an event. First, such that the agent has anticipated the outcome of the event, but not the event itself, and secondly, such that neither the state nor the corresponding outcome were considered. In the first case, the agent uses a state space, $\mathcal{S} = \{s_1, ..., s_n\}$ which enumerates all contingencies he can think of. However, the agent's state space $\mathcal{S}$ omits the state s_{n+1} with consequence $x_{i,n+1}$. Suppose now that the consequence $x_{i,n+1}$ was entertained by the agent as a potential result of a different state, say s_n. Then at state s_n, the agent perceived both $x_{i,n}$ and $x_{i,n+1}$ as possible consequences. This, however, seems like a case of option uncertainty rather than unforeseen contingencies.

Finally, the agent can fail to foresee both the state s_{n+1} and its consequence $x_{i,n+1}$. It is this case which most intuitively captures the notion of an unforeseen contingency, as exemplified by the nuclear meltdown in Japan. In the Japan example, the agent is not unaware of the possibility of such an event – nuclear meltdowns have happened before and hence the agent can be assumed to be familiar with the notion of a nuclear meltdown – and he also would have given the contingency of a nuclear meltdown a positive probability, had he thought of it, but he just didn't think to include it in his decision problem.

This case of uncertainty cannot easily be reduced to any other type of uncertainty. Also, it clearly cannot be integrated within Savage's framework; it is ruled out by the assumption of an exogenously given state space. In a decision situation with unforeseen contingencies of this kind, the agent cannot construct a small world decision matrix, and the "look before you leap" principle becomes uninformative. The agent must then "cross the bridge when he comes to it". Given that the agent cannot construct a small world decision matrix in the case of unforeseen contingencies, he will also not be able to decide optimally: the concept of optimality is defined relative to a decision matrix.

Although the case of unforeseen contingencies through an excessively coarse-grained state space is, perhaps, more popular in the literature than that where states are elided, there exist some models which, implicitly or explicitly, treat this latter case. In particular, Gilboa and Schmeidler's (1995) theory of case-based decisions is usually credited with suitability to such scenarios. In their model, decision makers evaluate prospects according to both their similarity to previously encountered problems and their utility. The model can account for unforeseen contingencies to the extent that these bear resemblance to previous decision problems.

The most prominent model for unforeseen contingencies is, perhaps, Kreps' (1992) model, which introduces unforeseen contingencies into Savage's framework. Kreps argues that we recognise an agent's anticipation of unforeseen contingencies by their preference for flexibility; this allows Kreps to infer which states of the world the agent subjectively considers possible. Whilst retaining most of Savage's theory, the main departure from Savage's theory is that in Kreps' model the state space is subjective. Dekel, Lipman and Rustichini (2001) extend Kreps (1992) to the case where the subjective state space can be derived from the agent's preferences, giving a more solid interpretation of the state space. A further extension has been suggested by Epstein and Seo (2009), who derive a unique state space from preferences; the authors provide axiomatic foundations for these preferences, and show that the state space is uniquely determined by the agent's ranking of menus. These theories, however, don't directly engage with the normative question of how an agent should deal with the possibility of unforeseen contingencies. There is therefore considerable scope for further investigation; however, this thesis will not pursue the question of rational choice under state space uncertainty any further.

2.7 Conclusion

In this chapter, we have first introduced Savage's notions of small and grand worlds, and argued that Savage's notion of a grand world is very demanding. We claimed that such a demanding notion of grand worlds in not necessary to talk about uncertainty. The Chapter then proceeded to identify different dimensions of uncertainty and characterised different kinds of uncertainty along those dimensions. In particular, uncertainty may vary in type (conceptual, empirical/factual and ethical) and severity (ignorance, ambiguity, mild uncertainty and certainty). The chapter argued that the kinds of uncertainty identified here, namely ambiguity, option uncertainty and state space uncertainty, differ along these dimensions.

In particular, we argued that ambiguity is perceived by the agent if the state space is more finely grained than the agent's probabilistic information. We argued that ambiguity is factual/empirical uncertainty, and considered arguments for and against reducing ambiguity to risk. We concluded that reducing ambiguity to risk comes at the cost of cognitive unease and descriptive inaccuracy.

Furthermore, we characterised four different conceptions of option uncertainty, decision situations where consequences of acts are not unique. Option uncertainty can be 'pushed' into the state space, but this comes at the cost of an increase the severity dimension of uncertainty; the agent is then faced with a decision problem under ambiguity. Option uncertainty can also be converted into ethical uncertainty by coarsening consequences; there is then uncertainty over the true utility of a consequence. Finally, we considered treating option uncertainty by re-defining acts as correspondences, and argued that this is the most fruitful approach for treating such decision problems.

Two different views on state space uncertainty were considered: an insufficiently fine-grained state space, and an incomplete state space. We argued that the first case can be reduced to ambiguity. Furthermore, we argued that the case where an agent fails to foresee a contingency, but does foresee its consequence, can be treated as option uncertainty. State space uncertainty of the kind where both a contingency and its consequence was elided by the agent was characterised as a "large world" problem, where agents can no longer maximise subjective expected utility.

The main thesis of this chapter is that the reduction arguments, if they are granted, come at the cost of an increase of severity dimension of uncertainty. This is so since most reduction strategies transfer uncertainty into the state space, thereby requiring the agent to hold very fine-grained beliefs. Rather than eliminating uncertainty, reduction converts one kind of uncertainty into another, whilst total uncertainty is conserved.

All three types of uncertainty – ambiguity, option and state space uncertainty – can be modelled as extensions to Savage's small world framework. Indeed, in each case, such a model exists: ambiguity has been characterised in a Savage framework by Sarin and Wakker (2004), a model of option uncertainty within Savage's framework has been given by Ghirardato (2001) and state space uncertainty has been modelled as an extension to Savage's framework by Kreps (1992). Shifting the perspective to a grand world model of the kind Savage had in mind may then not be necessary. Savage's grand world requires an infinitely fine-grained state space, so that its consequences are experiences of the person rather than descriptions of his circumstances. This model is so remote from practical applications that its use is reduced to that of a theoretical construct. Yet, recourse to the grand world is not required to make normative claims about decision making under uncertainty, since by extensions of the small world framework, much insight can be gained.

Chapter 3 will give a more detailed treatment of ambiguity, arguing that ambiguity may be objectively given. We investigate the normative implications of this claim, and contend that Savage's subjective expected utility must be weakened in cases of objective ambiguity. Chapter 4 turns to a model of option uncertainty, and uses the concept to explain status quo bias.

Chapter 3

Objective Ambiguity

3.1 Introduction

Consider the following example, due to Gilboa and Marinacci (2012): John and Lisa are considering buying insurance against the risk of developing a heart disease. In order to decide which insurance policy is appropriate for them, they would like to know the probability that they will develop such a disease within the next ten years. Both are 70 years old, smoke, and do not have a blood pressure problem. Their cholesterol level is at 310 mg/dL, and their HDL-C is at 45mg/dL. They each have a systolic blood pressure of 130. On the internet, they type their data into calculators which estimate the risk of developing a heart disease, and construct the table below.

As the table demonstrates, the different probability calculators don't agree on the likelihood of John and Lisa developing a heart disease within ten years. In

	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%

fact, the probability of John developing a heart disease lies within the interval $[25\%, 53\%]$ and the probability of Lisa developing a heart disease lies within the interval $[11\%, 27\%]$.

The Bayesian paradigm, as expounded in Chapter 1, requires agents to form a unique subjective prior probability over the state space. In the example, this would mean that John and Lisa would each be required to form a unique probability judgement over the states {'develop a heart disease within ten years', 'do not develop a heart disease within ten years'}. However, as ample empirical evidence demonstrates, agents do not, generally, hold preferences which are consistent with beliefs which are representable using a unique and additive prior probability distribution over the state space. This is usually attributed to the presence of *ambiguity*, the concept that probabilities of some events may be vague. In situations which are ambiguous, agents may hold preferences which are incompatible with the existence of a unique and additive probability distribution over the state space, thereby revealing *ambiguity aversion*. A preference pattern is called ambiguity averse when agents express a preference for acts which pay out a given amount with a known probability over acts which pay out a given amount with an unknown probability.

This chapter addresses the normative question whether, and in which situations, ambiguity aversion is rational. To answer this question, we first offer a definition of ambiguity. In particular, we will argue that ambiguity may be *objectively* given. To model objective ambiguity precisely, we relax the assumption that only states, acts, outcomes and preferences are observable. In particular, we extend Savage's framework to include, additionally to the mentioned elements, an objective likelihood ranking $\trianglerighteq$ defined on the algebra of events 2^S. Then $A \trianglerighteq B$ can be read as "A is objectively at least as likely as B". In the unambiguous case, for all events $A, B \in 2^S$ either $A \trianglerighteq B$ or $B \trianglerighteq A$ – the relation $\trianglerighteq$ is then complete on the set of all events 2^S. In contrast, in an objectively ambiguous decision problem, there will be events $C, D \in 2^S$ such that neither $C \trianglerighteq D$, nor $D \trianglerighteq C$. Events which cannot be compared via the objective likelihood ranking $\trianglerighteq$ will be called ambiguous. The exogenous likelihood ranking $\trianglerighteq$ can be used to derive definitions of a set of objectively unambiguous events $\Lambda \subseteq 2^S$, and of a set of unambiguous acts $\mathcal{A}^{\mathrm{ua}} \subseteq \mathcal{A}$.

In order to render the account of objective likelihood consistent with Savage's framework, we assume that the agent's subjective likelihood ordering on events, $\succeq^*$, which can be derived from the comparative probability axiom of Savage's theory (axiom P4 of Chapter 1) coheres with the exogenous likelihood ranking: whenever A is objectively more likely than B under $\trianglerighteq$, then A is also subjectively more likely than B under $\succeq^*$. This coherence is a particular case of David Lewis' (1986) *Principal Principle*, which requires that subjective beliefs agree with objective chances. The Principal Principle is, in our view, normatively convincing, and, as we argue, potentially coherent with Savage's own view of objective probability.

An objective definition of ambiguity allows for careful distinctions between ambiguity and ambiguity attitude. This distinction may be hard to make precise under subjective definitions. On many subjective notions of ambiguity a decision problem is identified as ambiguous whenever the agent's preferences violate the sure-thing principle, and as unambiguous otherwise. On subjective definitions, ambiguity is therefore revealed through preference, rather than given exogenously.

Due to the fact that on subjective definitions of ambiguity, ambiguity is identified only when the sure-thing principle is violated, subjective definitions may either over- or underestimate the presence of ambiguity. Overestimations of the presence of ambiguity arise when departures from subjective expected utility theory which do not arise as a result of ambiguity are attributed to ambiguity. This may be the case when risk-based violations of the sure-thing principle, as in the case of Allais' (1953) paradox, are erroneously attributed to ambiguity. Underestimations of the presence of ambiguity arise when ambiguity is not identified although the decision problem is ambiguous. This would be the case whenever an agent does not violate the sure-thing principle in an ambiguous decision problem; for instance, whenever the agent does not violate the sure-thing principle in Ellsberg's paradox. An objective notion of ambiguity aids in overcoming these issues, allowing for careful distinctions between the objective decision situation and the agent's behaviour in light of the decision situation.

Furthermore, on subjective notions of ambiguity it is impossible to assess in which scenarios the agent's failure to observe the sure-thing principle is a rational violation of the theory and in which scenarios it is not. As we will argue, violating the sure thing principle in situations of risk is not rational, whereas violating the principle in situations of ambiguity should be seen as permissible. The typical violation of the sure-thing principle situations of risk is given in Allais' (1953) paradox. We will argue that violations of the sure-thing principle in Allais' paradox are best understood as arising through a framing effect, which makes it difficult for the agent to apply Savage's theory successfully. In those cases where agents are presented with the Allais paradox specified in a small world decision matrix, they no longer wish to violate the sure-thing principle.

In contrast, violations of the sure-thing principle in situations of objective ambiguity are, as we argue, permissible. In particular, we will argue that ambiguity may force the agent's subjective likelihood ordering over events to be incomplete, so that agents are not willing to judge whether an event A is more likely than B or vice versa. Agents may respond to the incompleteness of their subjective likelihood ordering over events by hedging uncertainty, thereby violating the sure-thing principle. However, it seems permissible both to hold an incomplete likelihood ordering over events under objective ambiguity, and to respond to this incompleteness by preferring acts whose payoffs occur with known probabilities to those whose payoffs occur with unknown probabilities.

It follows from our discussion that the sure-thing principle should be assumed to hold on the set of unambiguous events, whereas it is permissible to violate the sure-thing principle when acts are compared which are measurable with respect to ambiguous events. On the view defended here both Schmeidlers's Choquet expected utility model, and Gilboa and Schmeidler's (1989) Maxmin expected utility model are too permissive, in that deviations from the sure-thing principle are admissible not only when acts are measurable with respect to ambiguous events, but also when they are not. A normative model of ambiguity should permit ambiguity aversion only in those situations which are objectively ambiguous.

The chapter is structured as follows: Section 3.2 turns to possible definitions of ambiguity, identifying the limitations of subjective definitions, and suggesting an objective notion of ambiguity. Section 3.3 defines a notion of ambiguity attitude consistent with our objective view of ambiguity. Section 3.4 compares the approach to ambiguity advocated here to the related literature. Section 3.5 contrasts the Allais and Ellsberg paradoxes, and argues that the sure-thing principle is valid in the former, but not required in the latter problem. Section 3.6 concludes.

3.2 Defining ambiguity

Let us start by reconsidering the Ellsberg paradox, since much of the motivation for studying ambiguity derives from the empirical finding of ambiguity aversion in this example. Consider Table 3.1. In the Ellsberg paradox, an urn contains 90 balls, 30 of which are red and the remaining are black or yellow in an unknown distribution. Hence, the probability of drawing a red ball is $1/3$, and the probability of drawing a red or yellow ball is contained within the interval $[0, 2/3]$ respectively. Intuitively, the event 'red' is *unambiguous* as it obtains with a known probability, whereas the events 'black' and 'yellow' are *ambiguous*, as they obtain with an unknown probability. As we have observed in Chapter 1, agents generally prefer act e_1 to e_2 and e_4 to e_3.

The preference pattern $e_1 \succ e_2$ and $e_4 \succ e_3$ is inconsistent with the existence of a unique and additive subjective probability distribution, for the following reason: if the agent attributed a subjective belief to the state 'black' of $1/3$ or more, then the agent would prefer gamble e_2 to e_1, as both have the same payoff, but the payoff of gamble e_1 occurs with a known probability of $1/3$ whereas the payoff of gamble e_2 occurs with a probability which is not precisely known. Therefore, the preference of e_1 over e_2 reveals that the agent must have attributed a subjective probability of *less* than $1/3$ to the event 'black'. In contrast, the preference of e_4 over e_3 reveals that the agent must believe the event 'black' to be *more* likely than $1/3$: both gambles e_3 and e_4 pay out \$100 in the state 'yellow', but e_3 additionally pays out \$100 in the event 'red' and e_4 pays out \$100 in the event yellow. Hence, if the agent attributed a probability of less than $1/3$ to black, gamble e_3 would be preferred. Since the agent prefers e_4 to e_3, they must have attributed a probability greater than $1/3$ to the event 'black'.

The preference pattern $e_1 \succ e_2$ and $e_4 \succ e_3$ can be interpreted as arising out of *ambiguity aversion*, as the inconsistency in the agent's assignment of probabilities to the event 'black' can be explained as resulting from aversion to the uncertainty over the true distribution of black and yellow balls. In particular, the agent prefers those gambles for which the payoffs obtain with a known probability to gambles where payoffs obtain with an unknown probability: The payoff of e_1 obtains with a known probability of $1/3$, and the payoffs of e_4 obtain with a probability of $2/3$, whereas the payoff of e_2 obtains with a probability within the interval $[0, 2/3]$ and

	red	*black*	*yellow*
e_1	\$100	\$0	\$0
e_2	\$0	\$100	\$0
e_3	\$100	\$0	\$100
e_4	\$0	\$100	\$100

Table 3.1: Ellsberg's three colour problem.

the payoff of e_3 obtains with a probability within the interval $[1/3, 1]$.

3.2.1 Subjective definitions of ambiguity

Most existing definitions of ambiguity define the concept subjectively, that is, as a property of the agent's preference relation. In particular, subjective definitions of ambiguity equate the existence of ambiguity with the revelation of ambiguity through particular preference patterns, such as the ones exhibited in Ellsberg's paradox. Subjective definitions proceed from an *observability assumption*, namely that to an external observer, only states, acts and outcomes, as well as the agent's preferences over acts, are observable (Ghirardato, Maccheroni and Marinacci, 2004). The observability assumption implies that ambiguity is defined within the constraints of a small world model, as exemplified in Table 2.4.

The motivation for modelling ambiguity subjectively is particularly well illustrated in Zhang's (2002) exposition of Ellsberg's four colour problem, shown in Table 3.2. In the example, there are 100 balls in an urn, which may be black, red, grey or white. It is known that 50 out of the 100 balls are either black or red, and that there are also a total of 50 black or grey balls. The probabilities of 'black or red' and 'black or grey' are then $1/2$ respectively. In Ellsberg's four colour problem, individuals express a typical preference pattern of $f_1 \succ f_2$, $f_4 \succ f_3$, and $f_5 \succ f_6$.

The reasoning behind these preference patterns is plausibly the following: f_1 is preferred to f_2 since the chances of obtaining \$100 are the same (the probability of the events 'black' and 'red' are each contained within the interval $[0, 1/2]$), whereas f_1 additionally yields a payoff of \$1 in the event of a black ball being drawn. Acts f_3 and f_4 are identical to f_1 and f_2, with the exception that both f_3 and f_4 pay out \$100 in the event 'grey'. Hence, the sure-thing principle would require that $f_1 \succ f_2 \Rightarrow f_3 \succ f_4$ or, respectively, $f_2 \succ f_1 \Rightarrow f_4 \succ f_3$. However, act f_4 hedges

	black	*red*	*grey*	*white*
f_1	\$1	\$100	\$0	\$0
f_2	\$100	\$0	\$0	\$0
f_3	\$1	\$100	\$100	\$0
f_4	\$100	\$0	\$100	\$0
f_5	\$1	\$100	\$100	\$100
f_6	\$100	\$0	\$100	\$100

Table 3.2: Ellsberg's four colour problem.

the uncertainty over the distribution of balls in the urn: it pays out \$100 whenever a black or grey ball is drawn, which is known to occur with a probability of $1/2$. In contrast, act f_3 does not hedge uncertainty, since it pays out \$100 whenever a red or grey ball is drawn – however, the probability of the event 'red or grey' is not known precisely. The preference pattern $f_1 \succ f_2$ but $f_4 \succ f_3$ violates the sure-thing principle, consistently with a hedging rationale. Acts f_5 and f_6 are, again, identical to f_1 and f_2 respectively, with the exception that f_5 and f_6 both pay out \$100 on the event 'grey or white'. Just like previously, the sure-thing principle requires that $f_1 \succ f_2 \Rightarrow f_5 \succ f_6$ or, respectively, $f_2 \succ f_1 \Rightarrow f_6 \succ f_5$. Coherently with the sure-thing principle, agents express a preference of f_5 over f_6, presumably for the same reason that act f_1 is preferred to f_2, namely because f_5 offers an additional chance of obtaining \$1.

Zhang argues that what this example demonstrates is that we are able to assess whether or not an event is ambiguous by observing preferences. In the example, the agent's preferences between f_1 and f_2 are reversed when a common outcome is replaced on the event 'grey', yielding acts f_3 and f_4. However, when a common outcome is replaced on the event 'grey or white', yielding acts f_5 and f_6 the original preference of $f_1 \succ f_2$ does not reverse. The difference between replacing an outcome on only 'grey' and on 'grey or white' lies in the fact that the agent is not able to assess the probability of the event 'grey', but they are able to determine the probability of 'grey or white'. The probability of 'grey' lies within the interval $[0, 1/2]$, it is hence ambiguous. In contrast, the probability of 'grey or white' is exactly $1/2$, since 'grey and white' is the complement of 'black and red', which is known to occur with a probability of $1/2$; the probability of 'grey or white' is thus unambiguous. Intuitively, then, replacing an outcome on an ambiguous event may reverse preferences, whereas replacing an outcome on an unambiguous event will not. In the special case where all events are unambiguous, Zhang argues, the sure-thing principle of expected utility is always satisfied; only ambiguity aversion will lead to violations of the sure-thing principle. This establishes the intuition behind subjective notions of ambiguity: events are ambiguous whenever the sure-thing principle is not satisfied on that event.

This reasoning leads Zhang to identify the absence of ambiguity with the criterion that the sure-thing principle obtains. In particular, Zhang proposes the following definition of *unambiguous events*:

Definition *(Zhang, 2002): An event A is unambiguous if (i) For all acts $f(\cdot), f^*(\cdot)$ and outcomes $x, y \in X$:*

$$\begin{bmatrix} f(s) & \text{if} & s \in A^c \\ x & \text{if} & s \in A \end{bmatrix} \succeq \begin{bmatrix} f^*(s) & \text{if} & s \in A^c \\ x & \text{if} & s \in A \end{bmatrix}$$
$$\Rightarrow \begin{bmatrix} f(s) & \text{if} & s \in A^c \\ y & \text{if} & s \in A \end{bmatrix} \succeq \begin{bmatrix} f^*(s) & \text{if} & s \in A^c \\ y & \text{if} & s \in A \end{bmatrix}.$$

and if (ii) For all acts $f(\cdot), f^*(\cdot)$ and outcomes $x, y \in X$:

$$\begin{bmatrix} f(s) & \text{if} & s \in A \\ x & \text{if} & s \in A^c \end{bmatrix} \succeq \begin{bmatrix} f^*(s) & \text{if} & s \in A \\ x & \text{if} & s \in A^c \end{bmatrix}$$
$$\Rightarrow \begin{bmatrix} f(s) & \text{if} & s \in A \\ y & \text{if} & s \in A^c \end{bmatrix} \succeq \begin{bmatrix} f^*(s) & \text{if} & s \in A \\ y & \text{if} & s \in A^c \end{bmatrix}.$$

Otherwise, A is called ambiguous.

Part (i) of Zhang's definition holds that whenever event A is unambiguous, then the sure-thing principle should hold on the partition (A, A^c). The intuition behind condition (i) follows that behind Ellsberg's four-colour problem explained above: whenever the event A is unambiguous, then preferences are separable across (A, A^c). Notice that condition (i) is a special case of the sure-thing principle where the subact on A is constant; the full sure-thing principle imposes separability also for non-constant subacts $g(s), g^*(s)$ instead of the constant subacts x, y. The constancy of the acts on the event A is important as the condition obtained when the outcomes x, y are replaced with $g(s), g^*(s)$ may not be true: we can not make any claims about the ambiguity of A when A leads to different outcomes across its states. Part (ii) holds that the case where A is replaced with A^c everywhere in (i) is true as well. This conditions is imposed because an event is unambiguous if and only if its complement is also unambiguous (Epstein and Zhang, 2001).

A second observation which follows from Ellsberg's four colour example is that the set of unambiguous events will not satisfy the requirements of an algebra. In particular, a σ-algebra $\mathcal{B}$ satisfies the property of closure under intersection, that is, if two events A, B are each contained in $\mathcal{B}$, it must also be the case that $A \cap B$ is contained in $\mathcal{B}$. However, from the four-colour example we know that this property is not satisfied for unambiguous events: even though the events A = 'black or red' and B = 'black or grey' are each unambiguous (each obtain with a probability of $1/2$), the intersection of A and B, namely event C = 'red' is ambiguous (the probability of 'red' is contained within the interval $[0, 1/2]$). Zhang thus argues that the set of unambiguous events must form a λ-system, which shares the properties of a σ-algebra with the exception that a λ-system need not satisfy closure under intersection. A λ-system of events $\mathcal{A}$ with typical element A is defined as follows:

(i) $S \in \mathcal{A}$

(ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

(iii) If $A_n \in \mathcal{A}$ for $n = 1, ...,$ and $A_i \cap A_j = \emptyset$, then $\forall i \neq j$, $\cup_n A_n \in \mathcal{A}$.

Property (i) holds that the sure event must be an element of $\mathcal{A}$. This requirement should be satisfied for the set of unambiguous events, as the state space itself is unambiguous: it occurs with probability one. Requirement (ii) is called closure under complementation, and it holds that if an event A is in $\mathcal{A}$, then so is its complement A^c. This requirement should also be satisfied for the set of unambiguous events, since whenever the probability of some event A is known, then the probability of its complement is just one minus the probability of A. Finally, requirement (iii) is called closure under countable disjunctive unions, and it holds that whenever the intersection of two unambiguous events A_i and A_j is empty, then the union of A_i and A_j is also contained in $\mathcal{A}$. This requirement should also be satisfied for the set of unambiguous events, since whenever the probability of events A_i and A_j are known, and there is no state $s \in S$ contained in both A_i and A_j, then the probability of the union of A_i and A_j is just the sum of their individual probabilities.

We have seen so far that Zhang's (2002) definition of ambiguity identifies the presence of ambiguity with the sure-thing principle being violated. However, on Zhang's definition of ambiguity, the presence of ambiguity may be either over- or underestimated. It may be *overestimated* for the following reason: an agent may violate the sure-thing principle not because of ambiguity aversion, but because of risk-aversion. However, on Zhang's definition, *every* violation of the sure-thing principle is attributed to the presence of ambiguity. Zhang's definition of ambiguity may then yield *too small* a set of unambiguous events, as even those events where the sure-thing principle is violated due to risk-aversion will count as ambiguous. Furthermore, Zhang's definition of ambiguity may *underestimate* the presence of ambiguity when it is the case that the agent assigns precise probabilities to all events *in spite* of the fact that an event is ambiguous. For instance, an agent who assigns precise subjective probabilities to the intuitively ambiguous events 'black' and 'yellow' in Ellsberg's three-colour problem will not violate the sure-thing principle. On Zhang's definition, the events 'black' and 'yellow' would therefore be unambiguous. Therefore, Zhang's definition of ambiguity may yield *too small* a set of unambiguous events. In summary, Zhang's definition of ambiguity works only in the special case where agents never violate the sure-thing principle out of risk-aversion, and always violate the sure-thing principle in situations of ambiguity.

There exist a number of alternative subjective definitions of ambiguity, for instance, Epstein and Zhang (2001), Ghirardato and Marinacci (2002), Ghirardato, Maccheroni and Marinacci (2004). However, none of the subjective definitions of ambiguity is fully satisfactory in discerning cases of ambiguity from cases where no ambiguity is present. We will explore the issues with subjective definitions of ambiguity further in the next subsection.

3.2.2 Problems with subjective definitions

Subjective approaches to defining the notion of ambiguity within Savage's framework (or the Anscombe-Aumann framework) proceed from the basic intuition that departures from (subjective) expected utility theory are induced by the presence of ambiguity. Hence, violations of the theory are attributed to the presence of ambiguity, such that the presence of ambiguity is identified with violations of the sure-thing principle (respectively the independence axiom). This is in line with the observability condition, which permits as observable only the information contained within a small world framework. However, there are three interrelated problems subjective definitions of ambiguity suffer from:

1. *Separation between ambiguity and ambiguity attitude.* Subjective definitions of ambiguity identify the presence of ambiguity with ambiguity-averse behaviour. The link between ambiguity and ambiguity aversion is direct: on subjective definitions of ambiguity, there is ambiguity whenever the agent reveals ambiguity aversion. Ghirardato, Maccheroni and Marinacci (2004, p.137) defend this close link between ambiguity and preference-related traits as follows: "as we are ultimately interested in modelling the ambiguity *as it affects behavior*, we do not believe this to be a serious problem from an economic viewpoint". This argument seems unconvincing, for it seems unclear how one can model ambiguity as it affects behaviour if ambiguity itself is identical with ambiguity-averse behaviour. To put this point differently, if it is impossible to distinguish ambiguity and ambiguity-averse behaviour, then no claims can be made regarding the effects *of* ambiguity *on* behaviour, ambiguity just *is* ambiguity-averse behaviour. It is then impossible to address the question of what kinds of behaviour are rational in light of ambiguity, a point which we shall address in section 3.4.

2. *Overestimation of the presence of ambiguity.* As we have seen in our discussion of Zhang's (2002) definition of ambiguity, subjective definitions of ambiguity may overestimate the presence of ambiguity. This will occur whenever risk-based violations of expected utility theory are spuriously attributed to the presence of ambiguity. Zhang's definition conflates these two separate violations of expected utility theory. The view that subjective definitions of ambiguity may overestimate the presence of ambiguity proceeds from the intuition that ambiguity may be present independently of the revelation of ambiguity through preference, as argued in (1) above.

3. *Underestimation of the presence of ambiguity.* As we have seen in our discussion of Zhang's (2002) definition of ambiguity, subjective definitions of ambiguity may underestimate the presence of ambiguity. This occurs whenever an agent's preferences are consistent with expected utility theory in spite of the presence of ambiguity. As in point (2) above, subjective definitions of ambiguity can be said to underestimate ambiguity when ambiguity

is conceived of as separate from ambiguity-averse preference patterns.

By relaxing the observability assumption such that ambiguity is modelled exogenously, one can obtain a definition of ambiguity which allows for a meaningful distinction between ambiguity and behavioural traits whilst neither over-, nor underestimating the presence of ambiguity. An exogenous definition of ambiguity permits for a natural distinction between ambiguity and ambiguity aversion.

Relaxing the observability assumption requires, however, a departure from Savage's framework; it is presumably for this reason that to date objective notions of ambiguity have, to the best of our knowledge, found little attention in the literature on ambiguity cast within Savage's framework (and the reformulation of Savage's framework contained in the Anscombe-Aumann framework). Savage's framework permits as observable only states, consequences, acts and preferences over acts. Defining ambiguity objectively would require that additionally, the objective information the agent holds is modelled precisely. However, Savage's framework is too restrictive to admit exogenously given objective information.

The constraints of Savage's framework can be overcome by admitting exogenously given objective probabilities and simultaneously assuming David Lewis' (1986) *Principal Principle*, which requires that subjective beliefs should cohere with objective chances. Formally:

$$C(A|P(A) = x) = x \tag{3.1}$$

where C stands for a subjective probability (i.e. a "credence"), $P(A)$ is the objective probability of event A (i.e. a "chance"), and x is the value of the probability of A. The principal principle is intuitively plausible: if we know that the objective probability of A is x and do not hold any evidence contradicting this, we should believe the event A to be as likely as its objective probability. Applied to Ellsberg's three colour problem, this means that the agent should assign a subjective probability of $1/3$ to the event 'red' and a subjective probability of $2/3$ to the event 'black or yellow'. The Principal Principle seems like a natural extension to Savage's theory. We will thus assume in the following that the observables are not restricted to states, consequences, acts and preferences, but furthermore include an exogenously given likelihood ordering over the algebra of events 2^S.

We assume furthermore that the Principal Principle is valid, hence rendering the account of objective probabilistic information coherent with Savage's framework.

Before turning to the definition of objective ambiguity advocated here, it is worth pointing out that this extension of Savage's framework may be consistent with Savage's own view of objective notions of probability, for he remarks (Savage, 1954, p.51):

> *Thus far, in this book, I have not argued against the possibility of defining some useful notion of objective probability, but have contented myself with presenting a particular notion of personal probability. Therefore, at this point it might be tempting to seek a dualistic theory admitting both objective and personal probabilities in some kind of articulation with one another.*

And furthermore (1954, p.60):

> *Again, objectivistic views can be regarded as personalistic views according to which comparisons of probability can be made only for very special pairs of events, and then only according to such criteria that all (right-minded) people agree in their comparisons.*

Extending Savage's theory by the Principal Principle would appear to provide just such a "dualistic theory admitting both objective and personal probabilities". In the framework thus obtained, all probabilities are subjective, but some (or even all) are informed by primitively given objective probabilities.

3.2.3 An objective definition of ambiguity

Assume then, that the agent's decision problem consists of a small world decision matrix including additional objective information. More specifically, we assume henceforth that the agent's decision problem consists of a set of states S, a set of consequences X, a set of acts $\mathcal{A} : S \to X$ and an exogenously given likelihood relation $\trianglerighteq$ on the σ-algebra of events 2^S. For events A, B in the set of events 2^S, $A \trianglerighteq B$ can be read "A is at least as likely as B". The agent will then form preferences over acts in light of the consequences of acts, and will form subjective beliefs informed by the exogenously given likelihood relation $\trianglerighteq$. Let us first define a situation of risk, i.e. one where all events are unambiguous.

Definition: Objective risk. *Let $\trianglerighteq$ be a likelihood relation defined on 2^S. Then a decision problem is called unambiguous if $\trianglerighteq$ is complete on 2^S.*

$\trianglerighteq$ is said to be complete if for all $A, B \in 2^S$, either $A \trianglerighteq B$ or $B \trianglerighteq A$. As ambiguous events cannot be compared in likelihood, whenever the exogenously given likelihood relation $\trianglerighteq$ ranks all events in 2^S in their likelihood, then all events must be unambiguous.

Let us investigate this definition in light of Savage's framework. Within Savage's framework, we can define a *subjective* likelihood relation $\succeq^*$ from axiom P4, the comparative probability axiom. In particular, axiom 4 holds that:

[P4] (Comparative Probability): *For all events A, B and outcomes $x^* \succeq x$ and $y^* \succeq y$:*

$$\begin{bmatrix} x^* & \text{if} & A \\ x & \text{if} & \neg A \end{bmatrix} \succeq \begin{bmatrix} x^* & \text{if} & B \\ x & \text{if} & \neg B \end{bmatrix}$$
$$\Rightarrow \begin{bmatrix} y^* & \text{if} & A \\ y & \text{if} & \neg A \end{bmatrix} \succeq \begin{bmatrix} y^* & \text{if} & B \\ y & \text{if} & \neg B \end{bmatrix}.$$

From this axiom, we can construct $\succeq^*$ as follows: For events $A, B \in S$ and consequences $x, y \in X$ such that $x \succ y$,

$$A \succeq^* B \Leftrightarrow \begin{bmatrix} x & \text{if} & A \\ y & \text{if} & \neg A \end{bmatrix} \succeq \begin{bmatrix} x & \text{if} & B \\ y & \text{if} & \neg B \end{bmatrix}$$

Intuitively, given that the agent prefers outcome x to outcome y, the agent prefers an act which yields x if event A occurs and y if $\neg A$ occurs to an act which yields x when B occurs and y if $\neg B$ occurs whenever they subjectively think that event A is more likely than event B; this can be expressed in the subjective likelihood ordering $\succeq^*$.

Suppose now that the agent holds, additionally to states, consequences and acts, the exogenous objective likelihood ordering $\trianglerighteq$. Then it suffices to assume, via the Principal Principle, that $\succeq^*$ and $\trianglerighteq$ agree:

$$A \succeq^* B \Leftrightarrow A \trianglerighteq B. \tag{3.2}$$

Equation (3.2) adapts the Principal Principle to Savage's framework: The agent ranks event A as *subjectively* more likely than B via $\succeq^*$ whenever the objective likelihood ordering, $\trianglerighteq$, ranks event A as more likely than B. For instance, by equation (3.2) an agent is required to subjectively rank the event 'black or yellow' of Ellsberg's three colour problem as more likely than 'red', since 'black or yellow' is objectively more likely than 'red'. Let us now turn to the more complex case of ambiguity.

Definition: Objective ambiguity. *Let $\trianglerighteq$ be a likelihood relation defined on 2^S. Then a decision problem is called ambiguous if $\trianglerighteq$ is incomplete on 2^S.*

The relation $\trianglerighteq$ is said to be incomplete when there exist events $C, D \in 2^S$, such that neither $C \trianglerighteq D$ nor $D \trianglerighteq C$. Intuitively, events are objectively ambiguous whenever there are two events C and D such that $\trianglerighteq$ does not rank these in terms of their likelihood. It is important to note that $\trianglerighteq$ will not be able to rank two events C, D in terms of their likelihood even if just one of the events, say C, is ambiguous. Consider, for instance, Ellsberg's three colour problem. We know that the event 'red' occurs with a probability of $1/3$, and that the event 'black or yellow' occurs with a probability of $2/3$. So the events 'red' and 'black or yellow' are unambiguous, and can be compared in likelihood: 'black or yellow' is more likely than 'red'. However, suppose we now want to compare the unambiguous event 'red' with the ambiguous event 'black'. Even though we know the probability of 'red', we cannot compare 'red' and 'black' in terms of their likelihood. Of course, two ambiguous events can also not be compared in terms of likelihood: the events 'black' and 'yellow' are both ambiguous, and also cannot be compared.

From the observation that $\trianglerighteq$ is complete only when two events *both* of which are unambiguous are compared, it is easy to define the set of unambiguous events. In particular, we call the set of unambiguous events Λ, since they will form a λ-system. Of course, Λ is a subset of 2^S, the set of all events.

Definition: Set of unambiguous events. *Let $\trianglerighteq$ be a likelihood relation defined on 2^S. Then the set of unambiguous events Λ is given by the largest subset of 2^S such that $\trianglerighteq$ is complete on Λ.*

Thus, the set Λ is the largest subset of 2^S such that for all events A, B in Λ, either $A \trianglerighteq B$ or $B \trianglerighteq A$. It is easy to see this intuitively: if the set Λ is the largest subset of 2^S such that $\trianglerighteq$ is complete, then adding any further ambiguous event E to Λ would make $\trianglerighteq$ incomplete. Suppose that an ambiguous event E is added to Λ. Then for any event F in Λ neither $E \trianglerighteq F$ nor $F \trianglerighteq E$. This immediately yields the definition of ambiguous events: the set of ambiguous events is just given by $2^S \setminus \Lambda$. [1]

Let us illustrate the definition of unambiguous events using the Ellsberg three-colour problem. There, the state space S is given by $S = \{R, B, Y\}$, where R denotes 'red', B denotes 'black' and Y denotes 'yellow'. The σ-algebra of events is given by $2^S = \{S, \emptyset, R, B, Y, RB, RY, BY\}$, where RB denotes 'red or black', and so forth. Then Λ, the set of unambiguous events, is given by $\Lambda = \{S, \emptyset, R, BY\}$. The relation $\trianglerighteq$ is complete for all elements of Λ, since any two elements in Λ can be compared in likelihood. The set of ambiguous events is given by $2^S \setminus \Lambda = \{B, Y, RB, RY\}$, and none of the elements of $2^S \setminus \Lambda$ can be compared in likelihood.

In the case of ambiguity, equation (3.2) must be modified such that now the subjective likelihood ordering of events, $\succeq^*$ agrees with the objective likelihood ordering $\trianglerighteq$ only on the set of unambiguous events Λ:

$$A \trianglerighteq B \Rightarrow A \succeq^* B. \tag{3.3}$$

Equation (3.3) holds that whenever the objective likelihood ordering $\trianglerighteq$ ranks an event A as more likely than B with $A, B \in \Lambda$, then the subjective likelihood ordering $\succeq^*$ must agree. For instance, applied to Ellsberg's three colour problem, equation (3.3) holds that given that the event 'black or yellow' is objectively more likely than the event 'red', then the agent must also subjectively hold the event 'black or yellow' as more likely than 'red'. So far, equation (3.3) merely imposes the consistency of the subjective likelihood ranking $\succeq^*$ with the objective likelihood ranking $\trianglerighteq$. It does not make any claims with respect to the subjective likelihood rankings between events on which $\trianglerighteq$ is incomplete, namely the set $2^S \setminus \Lambda$. For instance, the relation $\trianglerighteq$ does not rank the events 'black' and 'yellow' in likelihood; 'black' and 'yellow' are elements of $2^S \setminus \Lambda$. The agent's subjective likelihood ranking between 'black' and 'yellow' is thus not determined by $\trianglerighteq$.

[1] More precisely, the relation $\trianglerighteq$ is a subset of $(2^S)^2$. The set of tuples (A, B) such that $\trianglerighteq$ is complete is then given by the set $\Lambda^2 \subseteq (2^S)^2$, whereas the set of tuples (A, B) such that $\trianglerighteq$ is incomplete is given by $(2^S)^2 \setminus \Lambda^2$.

Then, if the agent's preferences satisfy Savage's axioms, the agent will hold a complete subjective likelihood ordering of all events in 2^S, where the subjective likelihood ranking $\succeq^*$ agrees with $\trianglerighteq$ on those events where $\trianglerighteq$ is complete. In contrast, if the agent's preferences do not satisfy Savage's axioms, then $\succeq^*$ may not be complete; the agent's beliefs are then not representable using a probability measure on $(S, 2^S)$. As we shall argue below, under objective ambiguity the agent's preferences are not required to satisfy Savage's postulates, such that it is permissible that the agent does not hold a complete subjective likelihood ordering $\succeq^*$.

It is also worth pointing out that the interpretation of Savage's comparative probability axiom changes in light of equation (3.3). In particular, in a typical small world scenario, Savage's comparative probability axiom provides a definition of the agent's subjective likelihood relation $\succeq^*$. In contrast, given that we are considering the case where the agent has access to an exogenously given objective likelihood ordering $\trianglerighteq$, Savage's comparative probability axiom is here interpreted as a *consistency condition* between the agent's preferences and the exogenously given likelihood ranking $\trianglerighteq$. In particular, the agent's preferences between acts are consistent with the objective likelihood ranking $\trianglerighteq$ if preferences reveal that the agent holds an event A to be subjectively more like than an event B whenever A is objectively more likely than B.

The definition of unambiguous events leads to a natural definition of unambiguous acts:

Definition: Set of unambiguous acts. *Let $\Lambda \subseteq 2^S$ be a λ-system of events such that $\trianglerighteq$ is complete on Λ. Then the set of unambiguous acts $\mathcal{A}^{ua} \subseteq \mathcal{A}$ is given by the set of Λ-measurable acts.*

In the above definition, $\mathcal{A}^{\mathrm{ua}}$ is the set of unambiguous acts, which is a subset of the set of all acts $\mathcal{A}$. That is, for an unambiguous act h the typical outcome x will obtain on an event in Λ. More formally, letting $h^{-1}(x)$ designate the set of states where act h yields outcome x, when h is unambiguous, then $h^{-1}(x) \in \Lambda$. In contrast, for an ambiguous act $e \in \mathcal{A} \setminus \mathcal{A}^{\mathrm{ua}}$ the typical outcome x will obtain on an event $e^{-1}(x)$ which may be an element of the set of ambiguous events $2^S \setminus \Lambda$.

Before we proceed, let us consider the role of the exogenously given likelihood ordering $\trianglerighteq$. One possible interpretation of it is that it contains an objective likelihood ordering, such as the one given in the Ellsberg problem. This interpretation of $\trianglerighteq$ is particularly suitable to the present context, since we are interested in addressing the question whether the agent's beliefs are required to be representable using a unique and additive probability distribution in light of objective ambiguity. However, one could conceive of $\trianglerighteq$ as any arbitrary exogenously given likelihood ordering of events. But if we do not interpret $\trianglerighteq$ as an objective likelihood ordering, then a different justification than the Principal Principle would need to be provided to motivate the coherence between the agent's subjective likelihood ordering $\succeq^*$ and $\trianglerighteq$. In the following, we will think of $\trianglerighteq$ as containing objective information.

3.3 Defining ambiguity attitude

In the previous section, we have given an exogenous characterisation of ambiguity. Let us now turn to the question of what ambiguity attitude is. Ambiguity attitude refers to the agent's disposition toward the presence of ambiguity. Ambiguity attitude, as opposed to ambiguity as such, is always a property of the agent's preference relation, since it concerns the agent's subjective stance to the presence of ambiguity. Three types of attitude are possible towards the presence of ambiguity: ambiguity-neutrality, ambiguity-aversion and ambiguity-attraction. We follow Epstein (1999) in characterising ambiguity attitude by first defining relative ambiguity aversion, the notion that one preference relation is more ambiguity averse than another, and then deriving a notion of absolute ambiguity aversion.

Epstein's account of ambiguity aversion coheres with our notion of objective ambiguity in the sense that Epstein assumes an exogenously given set of unambiguous acts $\mathcal{A}^{\mathrm{ua}}$. Our notion of objective ambiguity above can be seen as giving a foundation for the use of an exogenously given set of unambiguous acts; our account of ambiguity is then complementary to Epstein's notion of ambiguity aversion. It is important to point out the Epstein's definition of ambiguity aversion is unique in that it assumes a set of unambiguous acts; other existing definitions of ambiguity aversion, such as the one contained in Ghirardato and Marinacci (2002), do not make reference to a set of unambiguous acts. Epstein suggests the following notion of comparative ambiguity aversion:

Definition: Relative ambiguity aversion *(Epstein, 1999): Given two orderings $\succeq_1$ and $\succeq_2$, say that $\succeq_2$ is more ambiguity averse than $\succeq_1$ if for every unambiguous act $h \in \mathcal{A}^{ua}$ and every act $e \in \mathcal{A}$:*

$$h \succeq_1 (\succ_1) e \Rightarrow h \succeq_2 (\succ_2) e \tag{3.4}$$

In the above definition, acts h and e differ in the sense that act h is measurable with respect to the set of unambiguous events Λ, whereas e is not. Then whenever the less ambiguity-averse preference relation $\succeq_1$ prefers (strictly prefers) the unambiguous act h to the ambiguous act e due to the greater certainty h offers, then the more ambiguity-averse relation $\succeq_2$ also prefers (strictly prefers) the unambiguous act h to the ambiguous act e.

One feature of the above definition is that it implies that $\succeq_1$ and $\succeq_2$ are both representable using an identical utility function (Ghirardato, 2004). To see why this is so, consider the special case where both acts h and e which are ranked in preference by $\succeq_1$ and $\succeq_2$ respectively are unambiguous; i.e. assume that $h, e \in \mathcal{A}^{\text{ua}}$. Then equation (3.4) above holds biconditionally, i.e. $h \succeq_1 (\succ_1) e \Leftrightarrow h \succeq_2 (\succ_2) e$. Let us show, by contradiction, that $h \succeq_2 (\succ_2) e \Rightarrow h \succeq_1 (\succ_1) e$. To this end, assume that it is not the case that $h \succeq_1 (\succ_1) e$. This means that $h \prec_1 (\preceq_1) e$. By implication of equation (3.4), this means that $h \prec_2 (\preceq_2) e$, contradicting $h \succeq_2 (\succ_2) e$. Since $\succeq_1$ and $\succeq_2$ agree on the ranking of unambiguous acts, they can be represented using the same utility function. Moreover, if the set of unambiguous events $\mathcal{A}^{\text{ua}}$ is sufficiently rich, then the preferences $\succeq_1$ and $\succeq_2$ on $\mathcal{A}^{\text{ua}}$ can be used to compute the degree of probabilistic risk aversion of preferences, and $\succeq_1$ and $\succeq_2$ will exhibit the same degree of risk aversion.

Having now defined a notion of relative ambiguity aversion coherent with our framework, let us turn to the notion of absolute ambiguity aversion. In order to define an absolute notion of ambiguity aversion, it is necessary to define an ambiguity-neutral preference relation, relative to which another preference relation will be more ambiguity averse. To this end, Epstein refers to Machina and Schmeidler's (1992) notion of *probabilistic sophistication.* Let us briefly introduce the concept of probabilistic sophistication in order to clarify why probabilistically sophisticated preferences can be seen as ambiguity-neutral. In particular, an agent whose preferences are probabilistically sophisticated will hold a unique and additive probability measure π on the state space $(S, 2^S)$. Then every act e will be viewed, by a probabilistically sophisticated agent, as a lottery over outcomes in X. Then the agent's utility function is a function of the distribution over outcomes induced by an act.

Let us explain this using Ellsberg's three-colour problem, and suppose that the agent holds the probabilities $\pi(\text{'red'}) = \pi(\text{'black'}) = \pi(\text{'yellow'}) = 1/3$. Consider now the act e_1 which pays out \$100 in the state 'red' and \$0 otherwise. The act e_1 can then be viewed as a lottery over outcomes, as it pays out \$100 with a probability of $1/3$ and it pays out \$0 with a probability of $2/3$. Call this distribution Θ. A probabilistically sophisticated agent will then hold a utility function which is some function W of the distribution over outcomes Θ induced by acts. More formally, consider a probability distribution π on $(S, 2^S)$ and an act e. Denote the distribution over outcomes induced by e relative to π by $\Theta_{\pi,e}$. Then an agent whose preferences are probabilistically sophisticated will hold a utility function $U(\cdot)$ such that $U(e) = W(\Theta_{\pi,e})$, where W is some strictly increasing function.

A preference relation is said to be probabilistically sophisticated if it ranks acts in utility purely in light of the probability measure on $\Theta_{\pi,e}$ they induce, thus transforming all acts into lotteries. A probabilistically sophisticated decision maker will hold beliefs satisfying the axioms of probability theory, so that beliefs can be represented using a probability measure. While a probabilistically sophisticated decision maker may be risk-averse, they will always be ambiguity-neutral. In contrast, the beliefs of an ambiguity-averse agent will not satisfy the axioms of probability theory, and can therefore not be represented using a probability measure. Epstein makes use of the notion of probabilistic sophistication to give an absolute, rather than relative, notion of ambiguity aversion:

Definition: Ambiguity aversion *(Epstein, 1999): Given two orderings $\succeq_{ps}$ and $\succeq$, say that $\succeq$ is more ambiguity averse than $\succeq_{ps}$ if for every unambiguous act $h \in \mathcal{A}^{ua}$ and every act $e \in \mathcal{A}$:*

$$h \succeq_{\text{ps}} (\succ_{\text{ps}}) e \Rightarrow h \succeq (\succ) e \tag{3.5}$$

According to Epstein's definition of ambiguity aversion, an agent is ambiguity averse if they are more so than a probabilistically sophisticated agent. Equation (3.5) holds that whenever a probabilistically sophisticated decision maker prefers an unambiguous act to an ambiguous one, then the ambiguity-averse decision maker also ranks the unambiguous act as preferable to the ambiguous one. Notice in particular that as observed earlier, $\succeq_{\text{ps}}$ and $\succeq$ are representable by the same utility function. Hence, $\succeq_{\text{ps}}$ and $\succeq$ disagree on the ranking of acts only when the two preference relations do not agree on the likelihood ranking over events.

Let us consider what this section, in conjunction with our notion of objective ambiguity, has achieved. We criticised subjective accounts of ambiguity on the grounds that they do not allow for a meaningful distinction between ambiguity and ambiguity attitude. This problem is addressed by defining ambiguity objectively, since it is no longer the case that any particular attitude to ambiguity reveals the presence of ambiguity. Furthermore, the richer framework suggested here allows for precise distinctions between violations of the sure-thing principle resulting from risk-related behaviour and violations resulting from ambiguity-related behaviour.

3.4 Related Literature

Whilst in the above we have presented the necessity of an objective definition of ambiguity as a response to Zhang's (2002) definition of ambiguity, similar criticisms can be made of alternative subjective definitions of ambiguity. In particular, this section compares our objective definition of ambiguity with the definitions provided by Klibanoff, Marinacci and Mukerji (2005), Ghirardato and Marinacci (2002) and Ghirardato, Maccheroni and Marinacci (2004).

Klibanoff, Marinacci and Mukerji (2005) provide a representation result which allows for a separation between ambiguity and ambiguity attitude. In particular, in Klibanoff, Marinacci and Mukerji's model agents hold preferences over lotteries, defined as functions which are measurable with respect to a partition of the state space for which objective probabilities are given; these preferences are assumed to satisfy the von Neumann-Morgenstern expected utility axioms. Also, agents hold preferences over so-called "second-order acts", the payoffs of which are contingent on which prior in a given set of priors is true. These latter preferences are assumed to satisfy Savage's axioms. These assumptions, together with a consistency condition forcing the agents' preferences between lotteries and second order acts to be consistent, allow the authors to prove the representation result. The representation allows for a separation between risk and ambiguity attitudes, where risk attitude can be measured by the curvature of the utility function, and ambiguity attitude is measured by a function which attaches a particular weight to each of the possible priors in a given set of priors.

In Klibanoff, Marinacci and Mukerji's model, agents will neither over- nor underestimate the extent to which a given decision problem is ambiguous, since ambiguity is identified by the agent not through violations of the sure-thing principle, but rather directly through violations of the weak comparative probability axiom; agents will identify an event to be ambiguous whenever their preferences reveal that an event A is both more and less likely than another event B. Furthermore, Klibanoff, Marinacci and Mukerji's result successfully distinguishes risk from ambiguity attitude, as risk attitude is revealed from preferences over lotteries, whilst ambiguity attitude is revealed from preferences over second order acts. However, Klibanoff, Marinacci and Mukerji implicitly assume the validity of the Principal Principle, by assuming that preferences over lotteries satisfy the von Neumann-Morgenstern axioms. An agent who holds von Neumann Morgenstern preferences over lotteries will conform their subjective credences to objective chances. A further potential weakness of the model concerns the interpretation of ambiguity in the model. Note that in Klibanoff, Marinacci and Mukerji's model ambiguity is a subjective feature revealed by the agent's preferences. As Al Najjar and Weinstein (2009, p.275) point out, it is not clear what it means for an agent to use the "wrong" prior in a subjective setting. Under the subjectivist view advanced by de Finetti (1974), "probabilities do not exist", so that there is no objective distribution which subjective credences may match or fail to match. Ambiguity, in Klibanoff, Marinacci and Mukerji's model is then, strictly speaking, uncertainty over something that does not exist. This issue regarding the role of ambiguity aversion is addressed once we understand ambiguity as an objective feature of the decision problem, in which case ambiguity is uncertainty over the true objective probability distribution.

Ghirardato and Marinacci (2002) as well as Ghirardato, Maccheroni and Marinacci (2004) pursue a similar approach to Zhang (2002) and Epstein and Zhang (2001) in defining ambiguity, in the sense that ambiguity is identified with violations of the sure thing principle. In contrast to Zhang (2002) and Epstein and

Zhang (2001), the authors identify the presence of ambiguity not with departures from probabilistically sophisticated behaviour, but rather with departures from subjective expected utility maximisation more generally. However, violations of subjective expected utility theory may occur, as pointed out above, either because of risk-based violations of the theory, or because of ambiguity-based violations of the theory. Hence, the approach suggested by Ghirardato, Maccheroni and Marinacci will have a systematic tendency to overestimate the presence of ambiguity. The authors concede that "we prefer to attribute all departures from independence to the presence of ambiguity. However, the reader may prefer to use a different name for what we call 'ambiguity.'" (Ghirardato, Maccheroni and Marinacci, 2004, p.138). Such problems are avoided on an objective notion of ambiguity, where the extent to which a given decision problem is ambiguous is exogenously given.

3.5 Rationality under ambiguity

We now turn to the question whether ambiguity aversion may be rational. In the following, we will argue that violations of the sure-thing principle are justified only in cases of ambiguity, but not in cases of risk. In order to argue this stance, we will compare the typical behaviour of agents exhibited in the Allais paradox with that displayed by agents in the Ellsberg paradox. Let us begin with risk-based violations of the sure-thing principle.

3.5.1 Risk-based violations of the sure-thing principle

Perhaps the most compelling challenge against the sure-thing principle under risk is given in Allais' (1953) paradox, illustrated in Table 3.3. In the Allais' paradox, subjects are first given the choice between gambles a_1 and a_2, where gamble a_1 pays out \$1 million for sure, whereas gamble a_2 pays out \$5 million with a probability of 10%, \$1 million with a probability of 89% and nothing with a probability of 1%. Most people prefer a_1 to a_2. Subjects are next asked to compare gambles a_3 and a_4, where a_3 pays out \$ 1 million with an 11% chance, and nothing with an 89% chance, whereas gamble a_4 pays out \$5 million with a 10% chance and nothing with a 90% chance. Now most subjects prefer gamble a_4

	#1	#2-#11	#12-100
a_1	\$1M	\$1M	\$1M
a_2	\$0	\$5M	\$1M
a_3	\$1M	\$1M	\$0
a_4	\$0	\$5M	\$0

Table 3.3: THE ALLAIS PARADOX

to a_3. These preferences violate the sure-thing principle, since gambles a_1 and a_2 are identical to gambles a_3 and a_4 respectively with the exception that gambles a_1 and a_2 pay out \$1 million with an 89% chance, and gambles a_3 and a_4 pay out nothing with an 89% chance.

The Allais paradox is a case of decision making under risk, since the agent is provided with objective probabilities for all events. There is no ambiguity in Allais' paradox, since the likelihoods of all events are objectively known. Hence, Allais' paradox is a case of a small world decision problem: the agent holds well-defined states, consequences, and acts defined as functions from the state space to consequences. As illustrated in Chapter 1, many have aimed to reconcile Allais' paradox with Savage's theory, as the sure-thing principle is extremely plausible as a normative principle: When two acts yield the same outcome in some state, then that outcome should not matter to the agent's decision. This chapter shares this intuition, and will give its own attempt to reconcile the paradox with Savage's theory.

In particular, it is useful to think of the sure-thing principle as an axiom which applies to a small world decision matrix, consisting of states, consequences, and acts. Allais' paradox, as presented in Table 3.3 satisfies this requirement. However, the options in Allais' problem are not usually presented in the form of Table 3.3. In Allais' original paper, the options are presented as follows (Allais, 1953, p.527):

(1) *Do you prefer situation A or B?*

SITUATION A: *Certainty of receiving 100 million.*

$$\text{SITUATION B:} \begin{cases} 10\% & \textit{chance of winning 500 million} \\ 89\% & \textit{chance of winning 100 million} \\ 1\% & \textit{chance of winning nothing} \end{cases}$$

(2) *Do you prefer situation C or D?*

$$\text{SITUATION C:} \begin{cases} 11\% & \textit{chance of winning 100 million} \\ 89\% & \textit{chance of winning nothing} \end{cases}$$

$$\text{SITUATION D:} \begin{cases} 10\% & \textit{chance of winning 500 million} \\ 90\% & \textit{chance of winning nothing} \end{cases}$$

There is a difference between presenting the decision maker with the problem as in Allais (1953), and presenting them with the decision matrix in Table 3.3. The difference is that in Table 3.3 the agent is provided with an exogenously given state space, as Savage's theory requires. In the problem as presented in Allais (1953), however, the agent is not provided with an exogenously given state space, so that the agent has to *construct* the state space. Whilst this task is fairly straightforward when comparing Situations A and B, it is less so for Situations C and D. In particular, it seems plausible that the fact that a state space is not given to the agent might confuse the agent, such that the state space they construct for Situations A and B does not agree with the state space they construct for Situations C and D. In this case, the sure-thing principle does not apply.

For instance, when comparing Situations A and B the agent may reason that the relevant state space is given by the states in Table 3.3, namely $S = \{\#1, \#2 - 11, \#12 - 100\}$. This state space allows the agent to compare the outcomes of Situations A and B in a structured way. However, it is a much less obvious how a state space should be constructed for situations C and D: the probabilities with which payoffs are obtained for Situation C are 11% and 89% respectively, whereas the probabilities with which outcomes obtain for Situation D are 10 % and 90%. It is, prima facie, not evident how the state space should be partitioned to allow for a useful comparison between Situations C and D. The agent might then reason that the difference between obtaining an outcome with an 89% probability or a 90% probability seems irrelevant, and similarly, that it is irrelevant whether an outcome obtains with a probability of 10% or 11%. Perhaps the agent may resolve this issue by simply constructing the state space $S = \{$'Win', 'Lose'$\}$, and comparing Situations C and D on that basis; the agent then decides for Situation D. Given that the state spaces used to compare Situations A / B and Situations C / D differ, the sure-thing principle no longer applies.

More importantly, when the paradox is presented as in Allais (1953), it is not straightforward to identify the common outcome in Situations A / B and C / D respectively. In contrast, it is much easier to see the common outcome in Table 3.3. Agents who are presented with the problem as given in Allais (1953) may plausibly agree with the sure-thing principle in the abstract, but may not think that Situations A / B and C / D have a common outcome which would be irrelevant to their decision.

This hypothesis is supported by the evidence in Carlin (1990). Carlin tested the Allais paradox using a different frame, namely one where payoffs depend on the numbers on a wheel. The setup of the Allais problem in Carlin (1990) therefore provides a state space, given by the numbers of a wheel. Carlin finds that the number of violations of the sure-thing principle is greatly reduced once the Allais paradox is so-presented: only 20 out of 142 respondents made the typical choices exhibited in Allais' paradox. Similarly, Conlisk (1989) shows that when the Allais gambles are formulated in a fashion that brings out the independence aspect, then violations of the sure-thing principle are greatly reduced.

These results suggest that once the agent holds an exogenously given state space, then they become aware of the irrelevance of a particular outcome which is common to two acts. In the absence of a given state space, agents may find it difficult to construct a state space. However, this is, as such, not a problem for Savage's theory, which presumes an exogenously given state space. Once a state space is provided, agents find that they wish to conform their choices with the sure-thing principle.

A further argument to the effect that the sure-thing principle is valid under conditions of risk has been made by Samuelson (1952). In particular, Samuelson argues that whenever outcomes of an act are complementary, then one cannot assume that the two outcomes of an act affect preferences independently. This may be the case for decisions the outcomes of which are nonstochastic. However, when the outcomes of acts are risky, as is the case for lotteries, then the outcomes of acts are never complementary: as only one state in the set of states will be true, only one outcome will result of a particular act. Hence, lottery outcomes are not complementary and should therefore affect preference independently. As we are here concerned with preferences under situations of risk, complementarities between outcomes will not arise and hence, the sure-thing principle is justified.

3.5.2 Ambiguity-based violations of the sure-thing principle

Recently, Al-Najjar and Weinstein (2009) have argued that the ambiguity aversion literature is lacking in normative content: the project of the research area is predominantly that of reconciling the descriptive evidence expressed in Ellsberg's paradox with the normative theory of Savage. Yet, as Al-Najjar and Weinstein argue, it is not clear why ambiguity aversion would constitute a normatively more convincing response to Ellsberg's paradox than any other alternative theory, such as a heuristic explanation of Ellsberg's paradox. In this section, we will defend the view that ambiguity aversion is normatively permissible, and that hence, agents should be permitted to violate the sure-thing principle under ambiguity.

Let us first contrast the two competing views. On the view that Savage's theory holds in situations of ambiguity, ambiguity does not constitute a separate case from the typical small world case we explained in Chapter 1. Under this view, the agent should adhere to the sure-thing principle in situations of ambiguity, and the agent's beliefs should be representable using a unique and additive probability distribution over the state space. On the opposing view, decision problems featuring ambiguity cannot be represented using Savage's theory for small worlds. The agent then need not adhere to the sure-thing principle, and the agent's beliefs may not be representable using a unique and additive probability distribution over the state space. We will here defend this latter point of view.

Let us now characterise these two competing views in light of our framework of objective ambiguity above. We have argued that when objective probabilities are available, the agent must conform their subjective beliefs to the exogenously given objective probabilities using the Principal Principle. Under ambiguity, however, the objective likelihood ordering $\trianglerighteq$ may not be complete on the algebra of events 2^S; the objective likelihood ranking will be complete only for a subset Λ of 2^S. From equation (3.3) we know that the agent will also have a subjective likelihood ordering which is complete on Λ. However, unless the agent subjectively ranks those events in likelihood which cannot be compared via $\trianglerighteq$, say events $A, B \in 2^S \setminus \Lambda$, the agent's subjective likelihood ranking will be incomplete also.

On the view that agents are required to treat ambiguous decision problems as small worlds, agents are then obliged to rank ambiguous events in likelihood, such that the incomplete subjective likelihood ordering $\succeq^*$ is completed using subjective beliefs. The so-obtained likelihood ranking $\succeq^*$ will then be complete on Λ as a result of the use of the Principal Principle, and will also be complete of $2^S \setminus \Lambda$ in virtue of the use of Savage's axioms. In contrast, on the opposing view defended here, the subjective likelihood ordering $\succeq^*$ is required to be complete on Λ on account of the use the Principal Principle, whereas it is not required to be complete on $2^S \setminus \Lambda$, since the agent is not required to adhere to Savage's framework when judging ambiguous events.

Consider two objectively ambiguous events A and B, such as the events 'black' and 'yellow' in Ellsberg's three-colour problem. Both events have an objective probability within the range $[0, 2/3]$. The requirement that $\succeq^*$ be complete implies that the agent must either hold that 'black' is more likely than 'yellow', or 'yellow' is more likely than 'black', or the two events are exactly equally likely. However, there seems to be no basis for such a judgement, as the agent does not know which of the two events is more likely, and has no evidence supporting the subjective likelihood ranking. The agent may therefore wish to withhold a judgement regarding the relative likelihoods of the events 'black' and 'yellow'. This seems normatively justified, if the agent does not hold sufficient information to make such a judgement.

The argument that $\succeq^*$ is not required to be complete amounts to claiming that Savage's comparative probability axiom has no normative force in situations of objective ambiguity; agents should not be required to rank ambiguous events in likelihood. It should then be normatively admissible that on the set of ambiguous events $2^S \setminus \Lambda$ the agent's subjective likelihood ordering is incomplete, such that for $A, B \in 2^s \setminus \Lambda$ neither $A \succeq^* B$ nor $B \succeq^* A$.

Let us draw out the implications of this claim on preferences using Ellsberg's three-colour problem. Given that the agent cannot judge whether 'red' in more likely than 'black' or vice versa, the agent cannot rank acts e_1 and e_2, as the two acts differ only on the events in which the payoffs occur. Similarly, gamble e_3 pays out \$100 in the event 'red or yellow', whereas the event e_4 pays out \$100 in the event 'black or yellow'. Again, the agent cannot compare acts e_3 and e_4 since their subjective likelihood ranking $\succeq^*$ cannot rank the events 'red or yellow' and 'black or yellow' in likelihood.

Agents may respond to this incomparability of acts resulting from the incompleteness of the subjective likelihood ordering by preferring acts which hedge uncertainty to acts which do not. This can be interpreted as a principle of caution, where agents prefer betting on unambiguous acts to betting on ambiguous ones. In the Ellsberg paradox, agents therefore prefer the unambiguous act e_1 to the ambiguous act e_2, and also prefer e_4 to e_3, as e_1 and e_4 are measurable with respect to unambiguous events, whereas e_2 and e_3 are not.

Let us investigate the question whether making decisions on the grounds of caution is a rational strategy in light of ambiguity. Return to the example in the introduction of this chapter, where John and Lisa are trying to decide whether or not to buy insurance against the risk of developing a heart disease. John and Lisa cannot assess the likelihood of developing such a disease. Suppose, however, that they know that if they were to develop a heart disease, they would not be able to afford treatment unless they are insured. It certainly seems rational to be cautious in this decision problem.

An agent who (i) holds an incomplete likelihood ordering over the set of events and who (ii) responds to the incompleteness by hedging uncertainty will violate the sure-thing principle on sets of ambiguous events. However, as we have argued above, an agent should not be required to hold a complete likelihood ordering over objectively ambiguous events, and may rationally respond to the incompleteness by hedging uncertainty. Therefore, it should be permissible for an agent to violate the sure-thing principle in situations of objective ambiguity.

3.5.3 Implications for axiomatic characterisations

It follows from our discussion that axiomatic characterisations should distinguish cases of risk from ambiguity. Situations of risk can and should be modelled using Savage's framework for small worlds, and imposing the sure-thing principle on preferences is justified. In contrast, in objectively ambiguous decision problems, the sure-thing principle is not compelling from a normative point of view.

There exist, in the ambiguity aversion literature, two different ways of allowing for ambiguity in axiomatic frameworks. In particular, the first tenet of Bayesianism holds that agent must form a unique and additive probability distribution over the state space. The ambiguity literature relaxes the first tenet of Bayesianism by either relaxing the additivity, or the uniqueness of the probability distribution representing beliefs. The ambiguity literature works predominantly within the Anscombe-Aumann framework. Ambiguity is introduced into the framework by weakening the independence axiom of the Anscombe-Aumann framework, which holds that for any three acts $f, g, h \in \mathcal{A}$ and a constant $\alpha \in [0, 1]$, $f \succeq g \Leftrightarrow \alpha f + (1-\alpha)h \succeq \alpha g + (1-\alpha)h$.

Schmeidler's (1989) nonadditive probability decision model (also known as the Choquet expected utility, or CEU model) allows for ambiguity by restricting the independence axiom so as to allow for nonadditive beliefs. Schmeidler proceeds from the intuition that in situations of ambiguity, agents use opportunities to hedge uncertainty. Consider the following example: act f pays out \$1 in state s_1 and \$0 in state s_2, and act g pays out \$0 in state s_1 and \$1 in state s_2. Suppose also that you do not know whether state s_1 or s_2 is more likely; the states are ambiguous. Then you might prefer an act $h = {}^1\!/\!_2 f + {}^1\!/\!_2 g$ to an act $h' = {}^1\!/\!_2 g + {}^1\!/\!_2 g = g$, since act h pays out \$${}^1\!/\!_2$ in both states, whereas act h' pays out \$0 in state s_1 and \$1 in state s_2. Whilst act h fully hedges uncertainty, act h' does not. However, one can easily verify that a preference pattern of $f \sim g$ and $h \succ h'$ violates the independence axiom. Proceeding from the intuition that the hedging rationale leads to violations of independence, Schmeidler restricts the independence axiom to acts which are *comonotonic*, where comonotonic acts offer no hedging opportunities. In particular, two acts f and g are said to be comonotonic whenever there are no two states s_1 and s_2 in S such that $f(s_1) \prec f(s_2)$ and $g(s_1) \succ g(s_2)$. Intuitively, for any two states two comonotonic acts increase (decrease) in the same direction, therefore offering no hedging opportunities.

Let us investigate the connection of Schmeidler's model to the present framework of objective ambiguity. In particular, Schmeidler's model permits hedging in that the independence axiom is restricted to comonotonic acts. However, Schmeidler's model therefore also permits hedging in situations which are objectively unambiguous; for instance, an agent may wish to hedge uncertainty in situations of pure risk. Hence, from our normative point of view Schmeidler's model is *too permissive*: it allows for ambiguity aversion in objectively unambiguous decision problems. In order to render Schmeidler's framework consistent with our notion of objective ambiguity, agents would be required to satisfy the independence axiom for all acts which are measurable with respect to unambiguous events, and for comonotonic acts which are measurable with respect to objectively ambiguous events. The agent is permitted to violate independence only on acts which are not measurable with respect to unambiguous events and which are not comonotonic.

A second prominent model of ambiguity is Gilboa and Schmeidler's (1989) Maxmin expected utility, or MEU, model. The authors permit for ambiguity aversion by relaxing the uniqueness of the probability distribution representing beliefs; the agent then entertains several probability distributions. In particular, Gilboa and Schmeidler weaken the independence axiom such that it applies only to mixtures with constant acts x; Gilboa and Schmeidler call the so-obtained axiom Certainty-independence (or C-independence).The C-independence axiom then reads for all $f, g \in \mathcal{A}$ and $x \in X$, $f \succeq g \Leftrightarrow \alpha f + (1-\alpha)x \succeq \alpha g + (1-\alpha)x$. Intuitively, mixtures with constant acts do not permit for hedging. Additionally, Gilboa and Schmeidler impose an uncertainty aversion axiom, which holds that for any two acts $f, g \in \mathcal{A}$ and a constant $\alpha \in [0,1]$, $\alpha f + (1-\alpha)g \succeq f$. The uncertainty aversion axiom imposes a weak preference for acts which hedge uncertainty.

Let us investigate Gilboa and Schmeidler's model in light of our framework. Note first that we have argued above that ambiguity aversion is rationally *permissible* in situations of objective ambiguity; we have not argued that ambiguity aversion is required. Gilboa and Schmeidler's uncertainty aversion axiom contrasts with our normative view in the sense that agents whose preferences satisfy Gilboa and Schmeidler's axioms will always hedge uncertainty. This seems *too restrictive*, as a normative model should not prescribe any particular attitude to objective ambiguity. For instance, the preference patterns consistent with the sure-thing principle in Ellsberg's three-colour problem, namely $e_1 \succ e_2$ ($e_2 \succ e_1$) and $e_3 \succ e_4$ ($e_4 \succ e_3$) should be rationally permissible. Secondly, the agents modelled in Gilboa and Schmeidler's model may violate independence also on the set of unambiguous acts, a position which seems *too permissive*. Just as for Schmeidler's CEU model above, the independence axiom should be assumed to hold for mixtures between all acts which are measurable with respect to the set of unambiguous events; these acts may not be constant. Furthermore, the C-independence axiom should be assumed for all acts; agents may then violate independence only for acts which are measurable with respect to ambiguous events.

3.6 Conclusion

This chapter has introduced a definition of objective ambiguity by introducing an exogenously given objective likelihood order $\trianglerighteq$ into Savage's framework, and requiring that subjective beliefs cohere with objective chances. We have used this definition to define a notion of unambiguous events and unambiguous acts. Within our framework it is possible to distinguish thoroughly between ambiguity and ambiguity attitude. Our notion of ambiguity coheres with Epstein's (1999) notion of ambiguity attitude, which makes reference to an exogenously given set of acts.

We have argued that whilst in situations of risk, Savage's theory for small worlds can and should be employed, in situations of objective ambiguity violations of both Savage's sure-thing principle and comparative probability axiom are rationally permissible. We have used this normative view to argue that both Schmeidler's (1989) Choquet expected utility and Gilboa and Schmeidler's (1989) Maxmin expected utility model are too permissive, as they admit violations of the sure-thing principle also in cases where there is no objective ambiguity.

The account we have given of objective ambiguity opens up numerous possibilities for further research. It may be interesting, for instance, to contrast decision problems which are objectively more ambiguous with others which are less so. Our framework leads to a natural interpretation of the notion that one decision problem is more ambiguous than another, in that a more complete exogenous likelihood relation $\trianglerighteq$ will be associated with a lesser degree of ambiguity. Furthermore, it may be instructive to study the relation of Schmeidler's (1989) CEU model and Gilboa and Schmeidler's (1989) MEU model to the case of objective ambiguity characterised here further, in particular with respect to the connection of the exogenous likelihood ranking $\trianglerighteq$ to the nonadditive or nonunique beliefs exhibited by agents in the CEU and MEU models respectively. Finally, it may be interesting to study the connection between objective ambiguity and ambiguity attitude empirically.

Chapter 4

Option Uncertainty Aversion: Explaining Status Quo Bias

4.1 Introduction

In our daily lives, we often find it hard to assess what consequences follow from our actions; indeed, it is perhaps in the minority of cases that we can be sure that the exercise of an action will result in some consequence, even at a particular state of the world. For instance, when European politicians decided to establish a common currency contrary to the advice of economists who argued that the Eurozone is not an optimal currency area, they may well have evaluated the decision problem on the basis of its expected beneficial consequences, rather than on worst-case reasoning. However, as the current European sovereign debt crisis demonstrates, uncertainty aversion may not be completely irrational when much is at stake.

As we have argued in Chapter 2, there are decision situations where the state of the world may not fully determine the consequence of an action, such that the agent can envisage a variety of consequences at every state; we labeled these decision problems cases of "option uncertainty". Under option uncertainty, consequences cannot be treated as sure experiences of the deciding person, and Savage's (1954, p.84) claim that "we must expect acts with actually uncertain consequences to play the role of sure consequences in typical isolated decision situations" seems to lack justification. We have argued that this type of uncertainty is a distinct kind of uncertainty which should be treated separately from other types of uncertainty, such as ambiguity.

This chapter will give a more detailed account of option uncertainty, by modelling it formally. Rather than defining acts as functions from states of the world to consequences as Savage does, we will generalise this notion by defining acts as correspondences from states of the world into consequences. Each act is then associated with a set of consequences at every state. Ghirardato's (2001) axiomatic framework does just this, and we will therefore use this model as the basis of our analysis.

Ghirardato argues that there are three different ways in which what we call option uncertainty can come about: first if the decision maker has an underspecified choice set, i.e. every act really is a set of acts. Then, the agent perceives several possible consequences at any state, since the description of the act is not specific enough to yield a unique consequence at every state. This case can be interpreted as a case of coarse consequences, as discussed in Chapter 2. Secondly, the state space may be insufficiently fine grained, such that the consequence of each act at each state is not unique. We have argued in Chapter 2 that this is a case of what we called state space uncertainty, and showed how this can be reduced to ambiguity. The third and final case is that where the consequences of actions are insufficiently fine-grained, so that they do not constitute "sure experiences of the deciding person", i.e. determinate psychological states. It is the first and third case which are closest to what we mean by option uncertainty as we have characterised it in Chapter 2.

Ghirardato's model extends Savage's axioms to the case of acts which are defined as correspondences from states to consequences (axioms 1 – 7 below), and additionally imposes two very weak axioms (axioms 8 and 9 below) to model the agent's attitude towards the uncertainty over consequences. Axiom 8 captures a normatively appealing dominance condition: Assume that an act is constant, i.e. it yields the same consequence in every state, and that it is moreover *crisp*, meaning that consequences at every state are unique. Then if the unique consequence is judged better than any of the consequences of a constant act with uncertain consequences, then the former act should be preferred by the agent. Axiom 9 holds that for every set of consequences that results at a particular state, there is a single consequence which is better than it, and a single consequence which is worse than it. This axiom implies the boundedness of the utility function. With these two additional axioms, Ghirardato's model extends Savage's framework to account for, and model attitudes to, option uncertainty. In particular, Ghirardato shows that an agent whose preferences satisfy axioms $1-9$ will act as if maximising their expected utility relative to a probability measure over the set of states and a convex combination of the least and most favourable utility values of individual consequences.

A crucial aspect of the model is the assumption of *conscious unawareness* on the part of the agent, namely the fact that the agent is aware of the limits of his information. In particular, Ghirardato (2001, p. 250) points out that "in the absence of such awareness there would be little interesting that a decision theorist (or an economist, for that matter), could say". It is in virtue of this premise that the agent can be assumed to respond rationally to the lack of full information. This chapter retains this assumption.

Besides Ghirardato (2001), a model which shares the assumption of conscious unawareness and may also be interpreted as featuring *option uncertainty* is contained in Walker and Dietz' (forthcoming). Their model bears a strong resemblance to the present contribution, in that the state space does not resolve all uncertainty. The main difference between Ghirardato (2001) and Walker and Dietz (forthcoming) is that whilst in Ghirardato's model, the agent does not hold beliefs regarding the likelihoods of individual consequences in a given consequence set at a particular state, in Walker and Dietz' model the agent does. It is in this sense that the agents modelled in Walker and Dietz (forthcoming) can be seen as more rational than those in Ghirardato's model.

Using Ghirardato's model as a starting point, we extend the framework in order to model the *Status quo bias*. The status quo bias was originally observed as an empirical phenomenon by Samuelson and Zeckhauser (1988), and it holds that when there is a status quo, agents generally dislike giving it up for other alternatives. In particular, we assume that the status quo is an act which has no option uncertainty; in Ghirardato's model, a crisp act.

This conception of the status quo is particularly convincing when we interpret option uncertainty as *ethical uncertainty*, namely uncertainty over what value best reflects the agent's desire for a particular outcome. We argue that in the case of the status quo, the agent may find it easier to resolve ethical uncertainty than for other available alternatives. We assume further that agents are averse to option uncertainty, in the sense that for any given set of consequences X which an act yields at a particular state, the agent gives a relatively large weight to their least preferred element in the set of consequences. In the model proposed here, the agent is not uncertainty averse with respect to the status quo, but is uncertainty averse regarding other available alternatives; it is in this sense that the status quo is "privileged" over other alternatives. These two assumptions, that the status quo act is crisp and that the agent is averse to option uncertainty, imply that the agent will reveal a bias toward the status quo. Moreover, once one grants that option uncertainty aversion may be rational, and that the agent may be justified in conceiving of the status quo act as crisp due to uncertainty-reducing information regarding the status quo, then the agent's bias towards the status quo is rational.

The paper closest in spirit to the present chapter is Bewley (2002). Bewley argues that the presence of uncertainty may imply that the agent holds incomplete preferences: if the agent is uncertain regarding the likelihoods of particular outcomes, they may not be able to evaluate which of two acts is preferable. Additionally, Bewley makes an inertia assumption to the effect that an agent will stay at the status quo unless there is a different act which dominates the status quo for all possible priors. The difference between the present account and Bewley's is that Bewley assumes incompleteness and inertia, then deriving a representation which requires that an act will be preferred to the status quo only if it dominates the status quo for all priors in a given set of priors – a unanimity representation. The main difference between Bewley's representation and ours is that whilst Bewley assumes, via inertia, that agents will be biased toward the status quo, our model merely assumes crispness of the status quo and uncertainty aversion, thereby deriving status quo bias.

A second model connected to ours is Loomes, Orr and Sugden (2009), which gives an account of status quo bias within reference-dependent subjective expected utility theory (RDSEU). Loomes et al. explain status quo bias via taste uncertainty, namely the case where agents are uncertain with respect to the utility they derive from the consumption of a particular good. Taste uncertainty can be understood as a particular kind of ethical uncertainty, namely that where the utility value of a particular consequence is fully determined by the agent's taste; in this sense, Loomes et al.'s model is more restricted in scope than the present account. Furthermore, the reliance of Loomes et al.' account on reference dependence can be seen as a limitation of the model, since under reference-dependence, acts are evaluated in terms of the utility differential they generate with respect to some neutral reference point. In this sense, Loomes et al.'s model has a stronger flavour of bounded rationality than the model proffered here.

The chapter is structured as follows: Section 4.2 will give Ghirardato's model, and investigate the question to what extent Ghirardato's model embodies requirements of rationality. Section 4.3 will discuss the notions of option uncertainty aversion and relative option uncertainty aversion in Ghirardato's framework, and will argue that uncertainty aversion may be rational. Section 4.4 will model the status quo bias formally, showing that a more option uncertainty averse agent will have a tendency to prefer the status quo. We then argues that status quo bias may be rational. Section 4.5 concludes.

4.2 A model of option uncertainty

Ghirardato's (2001) model is an extension of Savage's framework, which we presented in Chapter 1. For ease of reference, the notation used in Chapter 1 is continued here. Just like in Savage's model, there is a set of states of the world, denoted S, the elements of which are mutually exclusive and collectively exhaustive. A typical element of S is denoted s. Furthermore, the set of consequences is denoted $\mathcal{X}$, elements of which are denoted x. Thereby, all elements of $\mathcal{X}$ are to be thought of as "sure experiences of the deciding person", i.e. fully specified payoffs over which no uncertainty can arise. The novelty of Ghirardato's model is the introduction of an algebra $\mathcal{A}$ of subsets of $\mathcal{X}$ containing all singleton elements, a typical element of which is denoted X. Let $\mathcal{U}$ denote the set of all nonempty subsets in $\mathcal{A}$, i.e. $\mathcal{U} \equiv \mathcal{A} \setminus \emptyset$. Agents are assumed to envision a set X as the result of their actions. Notice also that the notation x will be used to denote both an element of X and an element of $\mathcal{X}$.

Unlike Savage's model, where acts are functions from states of the world to consequences, in the present model acts are functions from states of the world S into $\mathcal{U}$. The set of acts $\mathcal{F}$ is then defined as $\mathcal{F} \equiv \mathcal{U}^S$. The so-defined acts are correspondences from S into $\mathcal{X}$. In summary:

States of the world: $S = \{..., s, ...\}$.

Events: $\mathcal{E} := 2^S = \{..., A, B, E, F, ...\}$.

Consequences: $\mathcal{X} = \{..., x, ...\}$.

Algebra of subset of $\mathcal{X}$: $\mathcal{A} = \{..., X, ...\}$

Set of nonempty subsets of $\mathcal{A}$: $\mathcal{U} = \{..., X, ...\}$

Acts: $\mathcal{F} := \mathcal{U}^S = \{..., f(\cdot), g(\cdot), ...\}$.

The agent is assumed to have a preference relation $\succeq$ on $\mathcal{F}$, with asymmetric and symmetric components $\succ$ and $\sim$. Finally, we will say that an event A is *null* if $f \sim g$ for every $f, g \in \mathcal{F}$ which differ only on A. The first axiom corresponds to Savage's postulate P1:

[Axiom 1] (Weak Order): *$\succeq$ is a weak order on $\mathcal{F}$.*
(i) (Completeness): Either $f \succeq g$, or $g \succeq f$.
(ii) (Transitivity): If $f \succeq g$ and $g \succeq h$, then $f \succeq h$.

The set of acts considered here is, however, much larger than that in Savage's model: Not only does the weak order assumption apply to all functions from the set of states to the set of consequences, but also to all functions from states to the set of all non-empty subsets of the set of consequences. The cardinality of the set $\mathcal{F}$ is therefore much larger than that of the set $\mathcal{A}$ of Chapter 1. This makes Axiom 1 a very demanding requirement.

By implication of Axiom 1, the agent is also able to rank all *constant acts*, where here a constant act is an act which pays out the same set $X \in \mathcal{U}$ of consequences in every state; the agent can therefore rank all elements of $\mathcal{U}$ in order of preference. Furthermore, Ghirardato introduces the term *crisp act* for those acts for which $f(s)$ is a singleton at every state, and denotes these $\mathcal{F}_c \subseteq \mathcal{F}$. The set of Savage acts of Chapter 1 is identical to the set of *crisp acts* in Ghirardato's model.

The following condition is an extension of Savage's postulate **P2** to the larger set $\mathcal{F}$:

[Axiom 2] (Sure-Thing Principle): *For all events $A \subseteq S$ and all acts $f(\cdot), f^*(\cdot), g(\cdot)$ and $h(\cdot) \in \mathcal{F}$:*

$$\begin{bmatrix} f^*(s) & \text{if} & s \in A \\ g(s) & \text{if} & s \notin A \end{bmatrix} \succeq \begin{bmatrix} f(s) & \text{if} & s \in A \\ g(s) & \text{if} & s \notin A \end{bmatrix}$$
$$\Rightarrow \begin{bmatrix} f^*(s) & \text{if} & s \in A \\ h(s) & \text{if} & s \notin A \end{bmatrix} \succeq \begin{bmatrix} f(s) & \text{if} & s \in A \\ h(s) & \text{if} & s \notin A \end{bmatrix}.$$

Just like Savage's P2, the axiom requires that the agent's preferences be separable across events (see Chapter 1). Of course, the objections against Savage's P2 apply also to Axiom 2 above.

The following condition is an extension of Savage's P3 to the larger set $\mathcal{F}$. Its interpretation mirrors that of P3 in Chapter 1: The agent's evaluation of sets of consequences $X, Y \in \mathcal{U}$ should not hinge on the state they obtain in. Of course, while in Savage's framework a constant act is one which results in the same (unique) consequence at every state, in Ghirardato's model a constant act pays out the same consequence set X in every state.

[Axiom 3] (Eventwise Monotonicity): *For all non-null events $A \subseteq S$, consequence sets $X, Y \in \mathcal{U}$ and acts $f(\cdot) \in \mathcal{F}$:*

$$\begin{bmatrix} X & \text{if} & s \in A \\ f(s) & \text{if} & s \notin A \end{bmatrix} \succeq \begin{bmatrix} Y & \text{if} & s \in A \\ f(s) & \text{if} & s \notin A \end{bmatrix} \Leftrightarrow X \succeq Y.$$

As in Savage's model, the agent's beliefs can be elicited from their preferences over acts, yielding a likelihood relation over events. This is expressed in the following axiom:

[Axiom 4] (Comparative Probability): *For all events $A, B \subseteq S$ and consequence sets $X, Y, X', Y' \in \mathcal{U}$ such that $X \succ Y$ and $X' \succ Y'$:*

$$\begin{bmatrix} X & \text{if} & A \\ Y & \text{if} & \neg A \end{bmatrix} \succeq \begin{bmatrix} X & \text{if} & B \\ Y & \text{if} & \neg B \end{bmatrix}$$
$$\Leftrightarrow \begin{bmatrix} X' & \text{if} & A \\ Y' & \text{if} & \neg A \end{bmatrix} \succeq \begin{bmatrix} X' & \text{if} & B \\ Y' & \text{if} & \neg B \end{bmatrix}.$$

Like in Savage's model, Axiom 4 allows us to defined a relative likelihood ranking $\succeq^*$ of events as follows: For all events $A, B \subseteq S$ and consequences $x, y \in \mathcal{X}$ such that $x \succ y$,

$$A \succeq^* B \Leftrightarrow \begin{bmatrix} x & \text{if} & A \\ y & \text{if} & \neg A \end{bmatrix} \succeq \begin{bmatrix} x & \text{if} & B \\ y & \text{if} & \neg B \end{bmatrix}.$$

The following nondegeneracy condition is identical to Savage's P5:

[**Axiom 5**] (**Nondegeneracy**): *There exist x and $y \in X$ such that $x \succ y$.*

Axiom 6, the Archimedean axiom, also parallels Savage's P6. It is used to ensure continuity of the preference relation.

[**Axiom 6**] (**Small Event Continuity**): *If $f, g \in \mathcal{F}$ are acts such that $f \succ g$ and $x \in X$ then there is a finite partition Π of S such that, for every $A \in \Pi$:*

$$f \succ \begin{bmatrix} x & \text{if} & s \in A \\ g(s) & \text{if} & s \notin A \end{bmatrix} \text{and} \begin{bmatrix} x & \text{if} & s \in A \\ f(s) & \text{if} & s \notin A \end{bmatrix} \succ g$$

Finally, axiom 7 imposes a dominance condition on preferences, holding that if an act f is worse than any of the consequences of another act g conditionally on event A, then act g should not be preferred:

[**Axiom 7**] (**Uniform Monotonicity**): *For all events $A \subseteq S$ and all acts $f, g \in \mathcal{F}$, if*

$$\begin{bmatrix} f(s) & \text{if} & s \in A \\ h(s) & \text{if} & s \notin A \end{bmatrix} \succeq (\preceq) \begin{bmatrix} X & \text{if} & s \in A \\ h(s) & \text{if} & s \notin A \end{bmatrix}$$

for all $h(\cdot)$ and each $X \in g(A)$, then:

$$\Rightarrow \begin{bmatrix} f(s) & \text{if} & s \in A \\ h'(s) & \text{if} & s \notin A \end{bmatrix} \succeq (\preceq) \begin{bmatrix} g(s) & \text{if} & s \in A \\ h'(s) & \text{if} & s \notin A \end{bmatrix}$$

for all $h'(\cdot)$.

Ghirardato shows that since Axiom 1 – 7 are extensions of Savage's P1 – P7, a similar representation to Savage's can be obtained. In particular, there exists a function $V : \mathcal{U} \to \mathbb{R}$ and a probability measure P on $(S, 2^S)$ such that, for every act $f, g \in \mathcal{F}$,

$$f \succeq g \Longleftrightarrow \int_S V(f(s))Pds \geq \int_S V(g(s))Pds \tag{4.1}$$

However, this representation does not capture the agent's attitude to the uncertainty over consequences: the representation above treats the consequence $f(s)$ of an act f as if it were a unique, sure consequence, rather than a set of individual consequences $X = \{x_1, ..., x_m\}$. Two further axioms are required to model the agent's attitude towards this uncertainty. In particular, Ghirardato introduces the following dominance condition:

[**Axiom 8**] **(Contingencywise dominance)**: *Given $X \in \mathcal{U}$ and $f \in \mathcal{F}_c$, suppose that for every $x \in X$, $f(s) \succ x$(resp. $x \succ f$). Then $f \succeq X$ (resp. $X \succeq f$).*

The interpretation of axiom 8 is that if a crisp act f which yields a unique consequence at every state is strictly better than *any* of the consequences of a non-crisp act yielding a set X at every state, then the crisp act should be preferred. The axiom is normatively appealing: an act with uncertain consequences none of which are better than the consequences of a crisp act should not be preferred by the agent.

One further axiom is needed for the representation. It holds that for every set of consequences X, there is a singleton consequence x which is better than it, and a singleton consequence which is worse than it. This is expressed in the following condition:

[**Axiom 9**] **(Outcome Boundedness)**: *For any $X \in \mathcal{U}$, there are $x, y \in \mathcal{X}$ such that $x \succeq X \succeq y$.*

Finally, Ghirardato introduces the following notation: given a real-valued function $u : \mathcal{X} \to \mathbb{R}$, let $\mathcal{U}^I \subseteq \mathcal{U}$ denote the set of non-empty sets X such that $\inf_{x \in X} u(x) \leq \sup_{x \in X} u(x)$. Ghirardato is then able to prove the following result:

Theorem 1 *(Ghirardato, 2001): If $\succeq$ satisfies Axioms 1 – 9, then there is a convex-ranged probability measure P on $(S, 2^S)$, a non-constant bounded utility function $u : \mathcal{X} \to \mathbb{R}$ and a function $\alpha : \mathcal{U} \to [0, 1]$ such that, if we define $V : \mathcal{F} \to \mathbb{R}$ by*

$$V(f) \equiv \int_{\mathcal{U}} \left[\alpha(X) \inf_{x \in X} u(x) + (1 - \alpha(X)) \sup_{x \in X} u(x) \right] \varphi_f(dX) \tag{4.2}$$

with $\varphi_f(U) \equiv P(\{s \in S : f(s) \in U\})$ for $U \subseteq \mathcal{U}$, then for all $f, g \in \mathcal{F}$

$$f \succeq g \Longleftrightarrow V(f) \geq V(g).$$

P is unique, u is unique up to positive affine transformation, and α is uniquely defined on $\mathcal{U}$.

Let us now consider the features of this representation. First, note that the utility function in equation (4.2) is a function of the fine-grained consequences x, which, as we mentioned earlier, are ultimate consequences or sure experiences of the deciding person. It is this dependency of the agent's utility on ultimate consequences which makes this a representation which explicitly models the agent's attitude to the uncertainty over consequences. Notice the contrast with the expression in equation (4.1), where the agent's utility V depends only on $f(s)$. Expression (4.1) therefore does not model the agent's attitude to uncertainty.

In Theorem 1, $\inf_{x \in X} u(x)$ designates the least utility value for any particular consequence within the consequence set X at a state s, and $\sup_{x \in X} u(x)$ refers to the highest utility value for a particular consequence in a consequence set X at a state s. The agent evaluates acts by considering a convex combination of the least and greatest utility values resulting at any state, so that the utility value associated with a consequence set X lies within the range spanned by the least and greatest utility value of the final consequences contained in X. The agent then weighs the so-computed utility values of consequence sets X at the states by the likelihood of each state being true, and ranks acts according to their expected utility thus obtained.

The factor $\alpha(X)$ reflects how strongly the agent weighs the least as opposed to the highest utility value for any consequence set. Ghirardato interprets $\alpha(X)$ as reflecting the agent's optimism or pessimism regarding the uncertainty: When $\alpha(X)$ is equal to one, the agent evaluates the set X purely on the basis of the least utility it could yield. A value of $\alpha(X)$ equal to zero would reflect optimism towards the uncertainty over the true consequence in the set X: the agent then evaluates the set X only in light of the highest utility value it could realise. Thereby, $\alpha(X)$ depends on the particular consequence set X the agent evaluates: for instance, the agent may be very pessimistic regarding the uncertainty over the true consequence within the set X, so that $\alpha(X) = 1$, but may at the same time be very optimistic regarding the uncertainty over the true consequence contained in the consequence set Y, so that $\alpha(Y) = 0$.

Finally, note that the function P on $(S, 2^S)$ is a probability measure; the agent thus holds additive beliefs over the states in the state space.

4.2.1 Option uncertainty and rationality

Let us now turn to the question to what extent the axioms of Ghirardato's model above can be interpreted as requirements of rationality in situations of option uncertainty. Axioms 1 – 7 closely follow Savage's model, they are indeed direct extensions of the concepts to the larger set of acts $\mathcal{F}$ considered here. So, prima facie, all criticisms one may raise against Savage's axioms apply with equal force to axioms 1 – 7.

However, the move from acts as functions from states of the world to consequences to correspondences from states into consequences implies that the restrictions on preferences in axioms 1 – 7 are now applied to a much larger set of acts than that considered in Savage's model. For instance, the completeness requirement of axiom 1 above requires the agent to rank all acts in $\mathcal{F}$. In Savage's framework the total number of acts is $\mathcal{X}^S$. In contrast, in Ghirardato's framework the total number of acts will be $2^{\mathcal{X} \times S}$, an order of magnitude larger than the set of constant acts in Savage's framework. For instance, if $|\mathcal{X}| = |S| = 2$, the total number of acts in Savage's framework is $\mathcal{X}^S = 4$, whereas in Ghirardato's framework, if $|\mathcal{X}| = |S| = 2$, the total number of acts is $2^{\mathcal{X} \times S} = 16$. With three consequences and three states, there are 27 Savage acts, and 512 Ghirardato acts. Axiom 1 above is, therefore, a much stronger requirement, and perhaps a less convincing one, than Savage's P1.

In contrast, axiom 8, contingencywise dominance, is very convincing as a requirement of rationality: It holds that an agent must check for dominance when choosing between acts. Axiom 9, outcome boundedness, is required to ensure that an given consequence set X will be ranked in utility between the utility values of the least and greatest utility values of its elements. The axiom precludes uncertainty averse preferences of the following kind: suppose that an agent ranks two constant and crisp acts f and g as indifferent, but prefers each of acts f and g to a constant act h which is defined as $h(s) = \{f(s), g(s)\}$ for all states s in S. For instance, act f could be an act which results in the consequence 'dinner' at every state, and act g could be an act which has as its consequence 'drinks' at every state; the set of all consequences is then $\mathcal{X} = \{$'dinner','drinks'$\}$. Although one might be indifferent between 'dinner' and 'drinks', one might prefer either of them to the uncertain prospect h which yields the consequence X=$\{$'dinner', 'drinks'$\}$ at every state. Holding $f \sim g$ but $f \succ h$ (resp. $g \succ h$) would violate axiom 9, since there would be no single consequence $x \in \mathcal{X}$ such that $X \succeq x$. In Ghirardato's model, h will be ranked as indifferent to acts f and g in virtue of axiom 9, although one might think that the greater uncertainty contained in h would make it less preferable than acts f and g. In this sense, axiom 9 is a rationality condition, ruling out this particular kind of uncertainty aversion. However, Ghirardato's representation permits for another type of uncertainty aversion, which we will discuss in greater detail in the following section.

4.3 Option uncertainty aversion

We now turn to the notion of option uncertainty aversion consistent with Ghirardato's model. Notice first that in Theorem 1 above the factor $\alpha(X)$ is a variable depending on the particular set X the agent evaluates. This means that the agent may be very pessimistic in evaluating a set of consequences X, and at the same time very optimistic with respect to option uncertainty when evaluating the set Y. It is thus interesting to consider the case where the factor α is constant, so that the agent expresses the same degree of uncertainty aversion in all evaluations between acts. In order to model the case where α is constant, Ghirardato introduces the following additional axiom:

[Axiom 10] (Option Uncertainty Attitude Robustness): *For every finite set $X \in \mathcal{U}$ and $x \in X^c$, suppose that $X' = X \cup \{x\}$, and that $\overline{x}$ and $\overline{x}'$ (respectively, $\underline{x}$ and $\underline{x}'$) are the $\succeq$-maximal (respectively, $\succeq$-minimal) elements of X and X'*

respectively, and that $\overline{x} \succ \underline{x}$. Then, for every $A \subseteq S$

$$X \sim \begin{bmatrix} \underline{x} & \text{if} & s \in A \\ \overline{x} & \text{if} & s \notin A \end{bmatrix} \Longleftrightarrow X' \sim \begin{bmatrix} \underline{x}' & \text{if} & s \in A \\ \overline{x}' & \text{if} & s \notin A \end{bmatrix}$$

Thereby, X^c denotes the complement of X. The axiom holds that if an agent is indifferent between a constant act which pays out X at every state and an act which pays out the preference-minimal element of X in event A and the preference-maximal element of X under the complement of A, then they should also be indifferent between a constant act X' which is larger than X and an act which pays out the preference-minimal element in X' under the event A and the preference-maximal element of X' in the complement of A. An agent's attitude to option uncertainty should not be affected by adding an element to the set of consequences they currently envision. Assuming axioms 1 – 10 then yields a variant of Theorem 1 with constant α, provided that the set of consequences $\mathcal{X}$ is finite:

Lemma 1 *(Ghirardato, 2001): If $\succeq$ satisfies Axioms 1 – 10 and if $\mathcal{X}$ is finite, then there is a convex-ranged probability measure P on $(S, 2^S)$, a non-constant bounded utility function $u : \mathcal{X} \to \mathbb{R}$ and a constant $\alpha \in [0,1]$ such that, if we define $V : \mathcal{F} \to \mathbb{R}$ by*

$$V(f) \equiv \int_{\mathcal{U}} \left[\alpha \inf_{x \in X} u(x) + (1-\alpha) \sup_{x \in X} u(x) \right] \varphi_f(dX) \tag{4.3}$$

with $\varphi_f(U) \equiv P(\{s \in S : f(s) \in U\})$ for $U \subseteq \mathcal{U}$, then for all $f, g \in \mathcal{F}$

$$f \succeq g \Longleftrightarrow V(f) \geq V(g).$$

P is unique, u is unique up to positive affine transformation.

Suppose now that an agent's preferences satisfy axioms 1 – 10, and are hence representable using the utility function $V(\cdot)$ given in Lemma 1.

Let us now turn to uncertainty attitude. As observed above, an agent is extremely averse (i.e. pessimistic) with respect to option uncertainty if they attach a utility value to a set of consequences X at a particular state equivalent to the least utility value of any of its elements. In contrast, the agent is option uncertainty loving (i.e. optimistic) if they attach a utility value to a set of consequences X at a particular state equivalent to the highest utility value feasible for its elements. Ghirardato thus defines option uncertainty pessimism and optimism as follows:

[**Axiom 11**] (**Option Uncertainty Pessimism / Optimism**): *Given $X, Y \in \mathcal{U}$, suppose that for every $x \in X$ there is a $y \in Y$ such that $x \succeq y$ (respectively for every $y \in Y$ there exists an $x \in X$ such that $x \succeq y$), then $X \succeq Y$.*

This definition fixes the notion of aversion to option uncertainty Ghirardato's model permits. Of course, when imposing axioms 1 – 10 and option uncertainty pessimism, the agent's preferences will be representable as given in Lemma 1 with a constant value of α equal to one; the value function then becomes $V(f) \equiv \int_{\mathcal{U}} \inf_{x \in X} u(x)\varphi_f(dX)$. Conversely, imposing axioms 1 – 10 and option uncertainty optimism implies that α is constant at zero, yielding the value function $V(f) \equiv \int_{\mathcal{U}} \sup_{x \in X} u(x)\varphi_f(dX)$.

It may be interesting to ask, then, under what circumstances it is the case in Ghirardato's model that one preference relation is more option uncertainty averse than another. In particular, assume that axioms 1 – 10 hold, so that the agent expresses a constant level of option uncertainty aversion with respect to all consequence sets X. It follows from axioms 1 – 10 that preferences are representable as in Lemma 1. Assume that both preference relations $\succeq_1$ and $\succeq_2$ agree on the ranking of constant acts $x, y \in \mathcal{X}$, so that $x \succeq_1 y \Leftrightarrow x \succeq_2 y$. Assume further that both $\succeq_1$ and $\succeq_2$ agree on the likelihood ranking of events, such that $A \succeq_1^* B \Leftrightarrow A \succeq_2^* B$. Then it follows from Lemma 1 that a higher value of α implies a greater degree of option uncertainty aversion: the larger α, the more the agent will evaluate sets of consequences in light of their least element. This yields the following notion of relative option uncertainty aversion:

[**Definition 1**] (**Relative Option Uncertainty Aversion**): *Consider two preference relations $\succeq_1$ and $\succeq_2$ such that $\succeq_1$ and $\succeq_2$ satisfy axioms 1 – 10 and assume that $\forall x \in \mathcal{X}$, $x \succeq_1 y \Leftrightarrow x \succeq_2 y$, and $\forall A, B \in 2^S$, $A \succeq_1^* B \Leftrightarrow A \succeq_2^* B$. Then $\succeq_2$ is more uncertainty averse than $\succeq_1$ if and only if $\alpha(\succeq_2) > \alpha(\succeq_1)$.*

4.3.1 Option uncertainty aversion and rationality

Let us now address the question whether option uncertainty aversion, as characterised in Ghirardato's model, is rational. First, let us look at the role of α, which imposes that the agent evaluates sets of consequences as linear combinations of their least and most preferable elements. This evaluation rule does not take into account the size of the set X, in spite of the fact that the size of the set X under consideration may matter in determining the degree of the agent's aversion to option uncertainty. For instance, consider two outcome sets $X = \{\$0, \$1000\}$ and $Y = \{\$0, \$1, \$2, ..., \$1000\}$. The first outcome set has only two outcomes, namely winning nothing and winning \$1000. In contrast, the second outcome set includes all intermediate outcomes between \$0 and \$1000. It would seem reasonable that an agent would be much more option uncertainty averse with respect to the set X than Y, as Y offers a number of payoffs which are better than winning nothing, whereas the set X offers only one outcome which is better than winning nothing. It is clear from the example that taking linear combinations of least and most favourable elements of a particular consequence set fails to take into account the additional information concerning the nature of all intermediate outcomes, which may be relevant to the agent's attitude to option uncertainty.

Let us now ask whether axiom 10 is a requirement of rationality. Intuitively, it seems plausible that an agent should be required to exhibit the same attitude to uncertainty for all acts in a given set of acts. If the agent's decision problem consists in choosing a particular medium of transportation to travel from A to B, it would seem irrational if the agent is very uncertainty averse with respect to flying, but not with respect to riding a motorbike. So it seems that the agent's attitude to uncertainty should be constant. However, axiom 10 may not reflect our intuitive notion of what it means to have a constant attitude to uncertainty. An agent whose uncertainty aversion can be modelled by a coefficient $\alpha = 2/3$ will appear only mildly uncertainty averse when this constant is applied to the set X above, but will appear extremely uncertainty averse when the same constant $\alpha = 2/3$ is applied to the set Y above. It appears that a linear α fails to reflect an intuitive notion of constancy of uncertainty attitude. Within the constraints of the model, however, it is not straightforward to construct a more plausible notion of a constant attitude to uncertainty.

We have not, so far, addressed the normative question to what extent option uncertainty aversion as such is rational. As we have seen, Ghirardato's model is consistent with all attitudes to option uncertainty ranging from option uncertainty pessimism to optimism, where option uncertainty aversion is understood in the particular sense of attaching a relatively large weight to the least preferable outcome of a particular outcome set X. This seems reasonable: the normative model should not prescribe a particular attitude to option uncertainty, as, depending on the particular decision problem, uncertainty aversion or attraction may be rational. An argument to this effect for the case of ambiguity has been given by Nehring (2009), who argues that an agent's uncertainty-averse decisions should be seen as decisions aiming at robustness in light of uncertainty, rather than at avoiding uncertainty. A similar argument can be made for the case of option uncertainty: an agent who exhibits uncertainty aversion will aim to make decisions such that their decision will yield favourable results even if one of the worse (or the worst) possible consequence of their action is true. Choosing such that the decision is robust to uncertainty is a rational strategy for coping with uncertainty.

One important psychological reason why people may be particularly averse to option uncertainty is that they may feel personally responsible for the outcomes of their actions. For instance, when faced with the decision whether or not to launch an attack on Iran, the Head of State of Israel may evaluate the alternatives in an uncertainty averse way, given that many lives may be lost as a direct consequence of his decision. Taking a decision which turns out to have bad consequences may then be associated with particularly severe regret over not having chosen a different alternative with less severe worst-case outcomes. More specifically, suppose an act f yields a consequence set X with a least element $\underline{x}$ at a particular state s, and an act g yields a consequence set Y with least element $\underline{y}$ at the same state s. Supposing that $\underline{x} \prec \underline{y}$, the agent may feel severe regret for having chosen act f when nature chooses state s and $\underline{x}$ occurs, since they may attribute the fact that $\underline{x}$ occurred to their choice of f over g. They may then regret not having chosen act g, which has a more favourable least element at s. Anticipating the potential for regret may lead agents to be more uncertainty averse than they would be without the feeling of personal responsibility for consequences.

	red	black	yellow
o_1	\$20	\$0	\$0
o_2	\$0	\$20	\$0
o_3	\$#black	\$0	\$0

Table 4.1: Option uncertainty in the Ellsberg gambles

4.3.2 Descriptive evidence on option uncertainty aversion

Option uncertainty is a type of uncertainty which has found little attention in the literature. However, there is an empirical study by Eliaz and Ortoleva (2011) the results of which can be interpreted as a test for option uncertainty aversion. Amongst other tests, Eliaz and Ortoleva introduce a variant of the Ellberg paradox where the prize the agent receives is conditional on the composition of the urn. Eliaz and Ortoleva assume that there are 60 balls in total, 20 of which are red, and the distribution over black and yellow balls is unknown. In Table 4.1, gambles o_1 and o_2 are the standard Ellsberg gambles, and gamble o_3 pays out a prize in dollars equivalent to the number of black balls in the urn. Gamble o_3 can be interpreted as an act featuring option uncertainty, since the payoff of gamble o_3 ranges between \$0, in case there are no black balls in the urn, and \$40, if 40 out of the 60 balls in total are black. Assuming the principle of insufficient reason (see section 2.4), the agent would hold that there are 20 black and 20 yellow balls in the urn. Under this assumption, the expected payoff of gamble o_3 is \$20 with a probability of $1/3$; this makes the expected payoff of o_3 just equivalent to that of gamble o_1. Eliaz and Ortoleva find that in the experiment 67 out of 80 subjects prefer gamble o_1 to o_2, consistently with the results of Ellsberg (1961). However, it is furthermore the case that 68 out of 80 subjects prefer o_1 to o_3, consistently with option uncertainty aversion.

4.4 Introducing status quo bias

Ghirardato's model provides a convincing framework for modelling the concept of option uncertainty introduced in Chapter 2. Before modelling the status quo bias formally, let us consider some possible conceptions of status quo bias, so as to clarify the nature of the account of status quo bias given here.

First, it is important to distinguish between the status quo bias, endowment effect, status quo reference effect and omission bias, since these effects are interrelated. The *status quo bias* refers to cases where agents favour, for no evident reason, the status quo over other available alternatives. Samuelson and Zeckhauser (1988) first observed the bias, and tested it using the following example (Samuelson and Zeckhauser, 1988, p.12):

> *You are a serious reader of the financial pages but until recently have had few funds to invest. That is when you inherited a large sum of money from your great uncle. You are considering different portfolios. Your choices are: (a) invest in a moderate risk company [...], (b) a high risk company [...], (c) treasury bills [...], (d) municipal bonds.*

Samuelson and Zeckhauser then presented a separate group of people with the following choices, which explicitly give a status quo:

> *You are a serious reader of the financial pages but until recently have had few funds to invest. That is when you inherited a portfolio of cash and securities from your great uncle. A significant portion of this portfolio is invested in moderate-risk Company A. You are deliberating whether to leave the portfolio intact or to change it by investing in other securities. (The tax and broker commission consequences of any change are insignificant.) Your choices are: (a) invest in a moderate risk company [...], (b) a high risk company [...], (c) treasury bills [...], (d) municipal bonds.*

The authors observed that the status quo option (in this case, investing in moderate risk company A) becomes significantly more popular if it is singled out as the status quo. Status quo bias has been observed in a variety of decision problems, ranging from investment decisions (see, e.g. Patel, Zeckhauser and Hendricks, 1991, Rubaltelli et al., 2005, and Kempf and Ruenzi, 2006) to moral decision making (see Bostrom and Ord, 2006, and Tetlock and Boettger, 1994) and medical decision making (see Kahneman, Knetsch and Thaler, 1991, and Johnson and Goldstein, 2004).

A separate but related anomaly in decision making is the *endowment effect*, which was first observed by Thaler (1980), and investigated in the context of prospect theory by Kahneman, Knetsch and Thaler (1991). This refers to cases where an agent holds a particular good which they would like to neither buy nor sell: it seems to them that the good is too expensive to buy, and too inexpensive to sell. The endowment effect is observed empirically as a gap between willingness to pay (WTP) and willingness to accept (WTA). An agent who succumbs to the endowment effect will generally be biased toward the status quo, namely that of holding the good, rather than buying or selling it. For instance, the endowment effect has been observed by Knetsch and Sinden (1984) using the following experiment: experimental subjects were given either a lottery ticket or \$2. After some time, all subjects were given the opportunity to trade the lottery ticket for \$2 or vice versa; yet very few subjects chose to switch. The endowment effect can be seen as a special case of status quo bias, where the status quo consists in holding a particular consumption good. We conceive here of the status quo as an act, which need not, but may, consist in holding a particular good. The endowment effect can be explained in the framework proposed below if we understand option uncertainty as *ethical uncertainty*, namely the case where the agent is unsure which utility value best reflects their desire for a particular good. In particular, it may be the case that the agent understands the utility they derive from holding a good better for the status quo than for other alternatives.

A further anomaly in close connection to status quo bias is the *reference effect*, which holds that agents evaluate alternatives in comparison to a given reference point (Tversky and Kahneman, 1991). According to reference-based theories (such as prospect theory), alternatives are evaluated in light of the gain or loss they offer relative to a neutral reference point. The status quo is then a natural reference point with which other alternatives are compared; status quo bias can be explained as resulting from loss aversion relative to the reference point. An expected utility model in this spirit has been suggested by Loomes, Orr and Sugden (2009), who analyse status quo bias in a reference-dependent subjective expected utility (RDSEU) model. The account of status quo bias proffered here differs from reference-based accounts in the sense that the agent is able to compare all acts; they will not exclusively compare acts to the status quo. In this sense, the option uncertainty framework is more general than reference-dependent accounts. However, in our model the agent is averse to option uncertainty, which can – in the extreme case of option uncertainty pessimism – be interpreted as a form of loss aversion.

Finally, it is important to distinguish between the status quo bias and *omission bias*. Thereby, omission bias refers to the case where an agent fails to act, rather than deliberating whether or not to act and deciding in favour of the status quo (Ritov and Baron, 1990). For instance, Johnson and Goldstein (2004) show that in countries where the default legislation is that all citizens who do not opt out are organ donors (e.g. France, Austria, Belgium), there are significantly more organ donors than in countries where the default legislation is that citizens do not donate organs (e.g. UK, Germany, Denmark). These data are perhaps best understood as exemplifying omission bias, since they reflect a failure to act, rather than a conscious choice for not acting. A second example which may be best understood as omission bias is given in the following example from Kahneman, Knetsch and Thaler (1991, p.199):

> *One final example of a presumed status quo bias comes courtesy of the JEP staff. Among Carl Shapiro's comments on this column was this gem: "You may be interested to know that when the AEA was considering letting members elect to drop one of the three Association journals and get a credit, prominent economists involved in that decision clearly took the view that fewer members would choose to drop a journal if the default was presented as all three journals (rather than the default being 2 journals with an extra charge for getting all three). We're talking economists here."*

Again, it seems plausible to assume that members elect do not generally entertain the option of dropping journals, so that status quo bias arises out of a failure to see the possibility of action, rather than a conscious choice in favour of the status quo. In the model proposed below, omission bias may be comprehensible as an unconscious choice between the status quo and alternative acts, where the status quo is implicitly ranked as better than other alternatives. This behaviour may be motivated by reasoning to the effect that unless deviating from the status quo will lead to unambiguously better outcomes, the status quo is chosen over other alternatives. In this sense, omission bias can be understood as the special case of status quo bias where alternatives to the status quo are not consciously entertained by the agent, due to the certainty of the status quo relative to other alternatives. In particular, our model does not require that the status quo be consciously chosen over other alternatives.

4.4.1 Modelling status quo bias

Turn now to the characterisation of the status quo. In particular, let $f \in \mathcal{F}$ denote the status quo act, and assume that $f \in \mathcal{F}_c$:

Assumption 1: *Let $f \in \mathcal{F}_c$ be the status quo.*

Assumption 1 holds that the status quo is a crisp act, i.e. one where option uncertainty plays no role. Let us investigate assumption 1 in light of the different interpretations of option uncertainty we have given in Chapter 2. We have argued that option uncertainty can be understood either as a non-uniqueness of consequences at particular states, or alternatively as ethical uncertainty, namely uncertainty over the utility value an agent attaches to a particular consequence at a given state. In each case, assumption 1 has different implications.

In the case where option uncertainty is interpreted as non-uniqueness of consequences at states, assumption 1 holds that for the status quo, a unique consequence exists at every state. On assumption 1, the agent finds it easier to assess what consequences follow from the exercise of the status quo than for other alternatives. This may be the case when the agent holds superior, uncertainty-reducing information with respect to the status quo, such that the agent understands the status quo act better than they understand other alternatives. For instance, if the status quo act is to live in a particular area, than the experience of having lived in the area before may enhance the agent's knowledge of the consequences of continuing to live in the same area. In contrast, there may be much greater uncertainty in determining the consequences of living in an unknown area; one's neighbours may not be nice, or one's way to work from a different location may require using a route which has a traffic jam every day. Assumption 1 would hold, when option uncertainty is interpreted as the non-uniqueness of consequences at particular states, that such factual uncertainties are fully resolved for the status quo.

Assumption 1 is particularly convincing, however, under the interpretation of option uncertainty as ethical uncertainty. In the case of ethical uncertainty, assumption 1 holds that the agent is able to determine the utility value of the consequences of the status quo act with exactitude, whilst they may not be able to do so for alternative acts. For instance, if I am a regular costumer of Starbucks, I find it easier to assess what the coffee will taste like than when I go to the new Italian coffee shop around the corner. Secondly, as pointed out in Bradley and Drechsler (forthcoming), ethical uncertainty may concern not the factual properties of a particular consequence (e.g. the taste of the coffee), but rather the utility value one would derive from that consequence (i.e. how much enjoyment the taste of the coffee produces). For instance, even though one may know all the specifications of a particular car, say, one may not be able to assess to what extent these specifications are desirable. Assumption 1 seems justified particularly when we interpret ethical uncertainty in this latter sense, as prior experience with a given commodity may resolve uncertainty of this kind.

This leads to the main claim of this chapter, namely that an agent who is more option uncertainty averse will be more biased toward the status quo: under a more uncertainty-averse preference relation $\succeq_2$ the set of acts judged at least as good as the status quo, i.e. $\mathcal{D} := \{g \in \mathcal{F} : g \succeq f\}$, will be smaller or equal to the set of elements judged at least as good as the status quo under a less uncertainty averse relation $\succeq_1$. This can be expressed as follows: for any two preference relations $\succeq_1$ and $\succeq_2$ such that $\succeq_2$ is more uncertainty averse than $\succeq_1$ according to definition 1 above,

$$|\mathcal{D}_2| := \{g \in \mathcal{F} : g \succeq_2 f\} \leq |\mathcal{D}_1| := \{g \in \mathcal{F} : g \succeq_1 f\}. \tag{4.4}$$

This finding can explain at least part of the experimental evidence for status quo bias, since uncertainty averse agents will decide more binary decisions between acts in favour of the status quo than less uncertainty averse agents. If many agents are option uncertainty averse and for this reason favour the status quo, this may then show up in experimental evidence as a bias towards the status quo. Formally, compare an agent whose preferences are represented by $\succeq_2$ and whose coefficient of option uncertainty aversion is given by $\alpha(\succeq_2)$ with an agent whose preferences are represented by $\succeq_1$ and whose coefficient of option uncertainty aversion is given by $\alpha(\succeq_1)$. If $\alpha(\succeq_2)$ is larger than $\alpha(\succeq_1)$, then the set of acts $\mathcal{D}_2$ deemed preferable to the status quo under $\succeq_2$ will be smaller than the set of acts judged as better than the status quo under $\succeq_1$. This is expressed in the following theorem, the proof of which is contained in the appendix:

Theorem 2: *Let $\succeq_1$ and $\succeq_2$ satisfy axioms 1 – 10. Assume that $\forall x \in \mathcal{X}$,*

$x \succeq_1 y \Leftrightarrow x \succeq_2 y$, and that $\forall A, B \in 2^S$, $A \succeq_1^* B \Leftrightarrow A \succeq_2^* B$. Assume also that $\mathcal{X}$ is finite. Then

$$\alpha(\succeq_2) \geq \alpha(\succeq_1) \Rightarrow \mathcal{D}_2 \leq \mathcal{D}_1. \tag{4.5}$$

It is easy to see this using a simple example (assume, for simplicity, that all acts are constant): suppose you hold a mobile phone contract with Deutsche Telekom, which you have held for several years. You pay roughly £50 every month. Since you distrust mobile phone companies generally, you are unsure which amount exactly you would pay were you to change providers – there may be small print in another mobile phone contract you do not understand and you can't assess the implications of. So you suppose that your phone bill with Orange would range between £20 and £60 every month. Suppose then that you are mildly uncertainty averse, with an $\alpha = 3/4$. You calculate accordingly that your expected phone bill with Orange would be given by $3/4 \times \pounds 60 + 1/4 \times \pounds 20 = \pounds 50$, so you are indifferent between staying with Deutsche Telekom and switching to Orange. An option uncertainty pessimistic agent would hold an $\alpha = 1$, computing thus that the expected bill with Orange is £60. So the more uncertainty averse agent will prefer remaining with Deutsche Telekom to switching to Orange, revealing a greater attachment to the status quo.

Let us analyse the option uncertainty account of status quo bias using Samuelson and Zeckhauser's example introduced above. The example explicitly mentions that "until recently [you] have had few funds to invest". It does not seem far-fetched to assume that the agent has absolutely no practical experience with investment decisions, if it is the first time in their life where such a decision has to be made. It is also natural to assume that when no status quo is specified, all options seem uncertain to the agent. So it is also reasonable to assume that, having constructed the state space $S = \{\text{'boom'}, \text{'recession'}\}$ the agent envisages several possible consequences at each state for each of the acts a,b,c and d. When a status quo is singled out, however, the agent may reason that this must have been a 'safe' strategy in the past, since their uncle accumulated a significant fortune: the agent now takes the fact that their uncle accumulated a large sum of money as evidence to the effect that the investment in the moderate risk company has been successful. Under the representation where a status quo is specified, the agent therefore envisions unique consequences at every state for the act of investing in a moderate risk company, but continues to entertain several possible consequences at every state for the acts c,d and e. It would then seem reasonable that the agent is biased towards keeping the portfolio as it is.

4.4.2 Status quo bias and rationality

The status quo bias seems, in many instances, irrational: when many potentially better alternatives are available, why should one choose to remain at the status quo? We have given an account here of status quo bias which partially rationalises it. Let us be clear in which way it is rationalised, and in which way it is not.

Firstly, note that we are here explaining status quo bias by reference to uncertainty averse preference. This seems rational to the extent that (i) the status quo is crisp, and (ii) the agent is uncertainty averse. We have argued in section 4.1 above that uncertainty aversion may be rational, and will therefore focus on substantiating (i). The assumption that the status quo is crisp is justified to the extent that the agent holds either uncertainty-reducing objective information, or uncertainty-reducing subjective information. The former applies in the case where there is option uncertainty in the form of non-uniqueness of consequences, whereas the latter applies to the case where option uncertainty takes the form of ethical uncertainty.

Let us focus on the case where consequences are non-unique first. Clearly, it is not always the case that the status quo is in fact crisp in that case. For instance, suppose your status quo act is living in New Orleans. One potential outcome of continuing to live in New Orleans is that your house may be destroyed by a hurricane, since these occur on a regular basis in that part of the world. There is, in fact, great uncertainty over what consequences follow from continuing to live in New Orleans. Nevertheless, an agent may favour continuing to live in New Orleans over moving away due to status quo bias. The model proposed here cannot account for cases where the agent does not hold uncertainty-reducing objective information with respect to the status quo; a bias towards the status quo where the status quo is itself uncertain cannot be explained by our model. This may be seen as a limitation of the account.

However, the assumption that the status quo is crisp does seem justified in many other cases, where the agent's prior experience with the status quo act eliminates option uncertainty. For instance, if an agent has visited a particular holiday destination previously, they may have learnt certain features of the destination, such as the distance to the beach, the quality of the hotel, and so forth, from past experience, therefore eliminating uncertainty of this factual kind. Ultimately, however, whether the assumption that the status quo is crisp is justified depends on the source of option uncertainty in particular decision problems.

Let us turn now to the case of ethical uncertainty. Assumption 1 then holds that the agent is certain with respect to the utility value a particular consequence affords them in the case of the status quo, whereas they may not be certain with respect to the utility values of other alternatives. The ethical uncertainty explanation of status quo bias can explain habitual behaviour in consumption choices. For instance, Samuelson and Zeckhauser give the example of a colleague who chose the same lunch for 26 years, namely a ham and cheese sandwich on rye bread. One day, Samuelson and Zeckhauser's colleague ordered a chicken salad sandwich instead, and has continued to order this for lunch ever since. Prima facie, it may seem as though this habitual behaviour is irrational, since the colleague could have had a healthier, tastier or cheaper diet by alternating their lunch choice. However, one may reason that the choice of the ham and cheese sandwich just reflects the fact that they know that this choice affords them some level of utility, whereas they did not know what utility value they would attribute to alternative lunch choices. Once Samuelson and Zeckhauser's colleague tried the chicken sandwich, this uncertainty was resolved: they were then able to attribute a unique utility value to the consumption of chicken sandwich. However, since the colleague continued to choose the chicken sandwich ever since, it seems reasonable to assume that the colleague continued to use an option uncertainty averse decision rule, evaluating all alternative options in light of their potentially worse utility values.

4.4.3 Related literature

The empirical pervasiveness of the status quo bias has led to the development of a number of models featuring status quo bias. Bewley's (2002) model of Knightian uncertainty models status quo bias in the sense that agents will deviate from the status quo only when another alternative is preferred to the status quo for all priors in a given set of priors. In Bewley's model, status quo bias is linked to ambiguity, whereas our account proceeds from a framework of option uncertainty.

Furthermore, whilst Bewley's model assumes status quo bias by postulating inertia, our account rationalises the bias. However, Bewley (2002) shares the intuition behind this paper in the sense that agents will trade the status quo for another alternative only when the competing act is certainly better than the prevailing one.

Loomes, Orr and Sugden (2009) explain status quo bias within a consumer choice model. Loomes et al. proceed from reference-dependent subjective expected utility theory, where agents are uncertain about the utility which will be yielded by their consumption experience in different taste states of the world. Loomes et al.'s model shows why the bias toward the status quo may be more or less strong depending on the decision environment. However, their model is more restrictive than ours in the sense that it explains status quo bias only in the special case of taste uncertainty, a particular kind of ethical uncertainty. Furthermore, our account differs from Loomes et al.'s in that we do not assume a reference-dependent framework; hence, the agents modelled here can be seen as more rational than those modelled in Loomes et al.'s framework.

A model of status quo bias within a revealed preference framework has been suggested by Masatlioglu and Ok (2005). Masatlioglu and Ok assume the status quo bias axiomatically. In particular, Masatlioglu and Ok's status quo bias axiom holds that if an alternative is chosen when it is not the status quo, it will be chosen uniquely when it is the status quo. Masatlioglu and Ok's model can be seen as a generalisation of revealed preference theory which allows for status quo bias and includes the standard framework as a special case. In contrast to their model, the present account rationalises status quo bias, by giving a rational choice explanation for status quo bias rather than assuming the bias axiomatically.

Finally, in a recent contribution Ortoleva (2010) links status quo bias to ambiguity aversion. In this sense, Ortoleva's model is strongly related to ours, since it provides a link between uncertainty and status quo bias. In contrast to our account, however, Ortoleva assumes status quo bias axiomatically, using a similar approach to Masatlioglu and Ok (2005). Ortoleva then shows formally that an agent who is biased toward the status quo will be more averse to ambiguity than one who is not. Ortoleva's result can be seen as complementary to our own, as we proceed from assuming uncertainty aversion and deriving bias towards the status quo.

4.5 Conclusion

This chapter has argued that once we extend Savage's framework to allow for uncertainty over the true consequence of actions, we can account for the empirical phenomenon of status quo bias, which might otherwise seem irrational. In particular, the chapter is based on Ghirardato's (2001) contribution, which generalises Savage acts such that correspondences, rather than functions, from states of the world to consequences are envisaged by the agent. We interpreted Ghirardato's framework as one of option uncertainty, as introduced in Chapter 2.

We then argued that on two relatively weak assumptions, namely that the status quo is a crisp act, and that agents are option uncertainty averse, one can account for status quo bias. In particular, we argued that the status quo may be perceived as crisp either because the agent holds objective knowledge which reduces the uncertainty surrounding the status quo, or because the agent holds superior subjective information, such that the agent can assess the utility value they would derive from the status quo better than for other alternatives. The interpretation of option uncertainty as ethical uncertainty allows us to explain status quo bias in consumption choices, as well as brand loyalty.

To the best of our knowledge, there is, so far, no model which explains status quo bias in a non-reference dependent set-up; for instance, Loomes, Orr and Sugden (2009) consider the status quo bias within reference-dependent subjective expected utility. Furthermore, those models which treat status quo bias either stipulate it axiomatically (e.g. Ortoleva, 2010, and Masatlioglu and Ok, 2005), or assume the bias behaviorally (see, e.g. Bewley, 2002). The present chapter offers a rational explanation of status quo bias appealing neither to reference-dependence, nor by assuming the bias. The approach taken here demonstrates, above all, that once we grant that the rationality constraints on agents must be weaker under conditions of uncertainty, we are able to give rational explanations of empirical phenomena such as status quo bias.

A.1 Appendix

Proof of Theorem 2:

We would like to show that (A) $\alpha(\succeq_2) \geq \alpha(\succeq_1)$ implies that (B)

$$|\mathcal{D}_2| := \{g \in \mathcal{F} : g \succeq_2 f\} \leq |\mathcal{D}_1| := \{g \in \mathcal{F} : g \succeq_1 f\}.$$

We will prove the claim by contradiction: $\neg B \Rightarrow \neg A$.

Assume that $\neg B$, such that $|\mathcal{D}_2| := \{g \in \mathcal{F} : g \succeq_2 f\} > |\mathcal{D}_1| := \{g \in \mathcal{F} : g \succeq_1 f\}$. Then it must be the case that $\exists g$ such that $g \in D_2(f)$ and $g \notin D_1(f)$. We have then that $g \succeq_2 f$ but $g \prec_1 f$. From $g \succeq_2 f$, by Lemma 1, we have $V_2(g) \geq V_2(f)$. By the assumption of crispness of f, $V(f) = \int_{\mathcal{U}} u(x)\varphi_f(dX)$. Then we have

$$\begin{aligned} V_2(g) &= \int_{\mathcal{U}} \left[\alpha \inf_{x\in X} u(x) + (1-\alpha) \sup_{x \in X} u(x)\right] \varphi_g(dX) \\ \geq V_2(f) &= \int_{\mathcal{U}} u(x)\varphi_f(dX). \end{aligned}$$

From $g \prec_1 f$, by Lemma 1, we have $V_1(g) < V_1(f)$. From the crispness of f, $V(f) = \int_{\mathcal{U}} u(x)\varphi_f(dX)$, so that:

$$\begin{aligned} V_1(g) &= \int_{\mathcal{U}} \left[\alpha \inf_{x\in X} u(x) + (1-\alpha) \sup_{x \in X} u(x)\right] \varphi_g(dX) \\ < V_1(f) &= \int_{\mathcal{U}} u(x)\varphi_f(dX). \end{aligned}$$

By assumption, $\succeq_1$ and $\succeq_2$ agree on the ranking of all elements $x \in \mathcal{X}$; hence, $x \succeq_1 y \Leftrightarrow x \succeq_2 y$. Whence $u_1(x) \geq u_1(y) \Leftrightarrow u_2(x) \geq u_2(y)$. By the uniqueness properties of $u(\cdot)$, it is the case that $\inf_{x\in\mathcal{X}} u_1(x) = \inf_{x\in\mathcal{X}} u_2(x)$ and $\sup_{x\in\mathcal{X}} u_1(x) = \sup_{x\in\mathcal{X}} u_2(x)$. Furthermore, by the assumption that $A \succeq_1^* B \Leftrightarrow A \succeq_2^* B$, we have $\varphi_1 = \varphi_2$. Hence, $V_2(g) \geq V_2(f)$ and $V_1(g) < V_1(f)$ if and only if (C) $\alpha(\succeq_2) < \alpha(\succeq_1)$, contradicting (A) as required. $\square$

Chapter 5

Axiomatising Bounded Rationality: The Priority Heuristic

Mareile Drechsler, Konstantinos Katsikopoulos and Gerd Gigerenzer

Expected utility theory remains to this day the dominant decision theoretic framework in economics. Much of the appeal of expected utility theory lies in its elegant axiomatic characterisations (e.g., von Neumann and Morgenstern, 1944, and Savage, 1954), which lend themselves to a normative reading. It has, however, been shown empirically that the axioms of expected utility are systematically violated by people. For instance, Allais (1953) has demonstrated violations of the independence axiom, and Kahneman and Tversky have empirically identified a number of violations of expected utility theory, including framing effects, the reflection effect, and the fourfold pattern of risk-taking (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992).

An alternative approach to studying human decision making is to study how human beings make choices in the real world. To obtain a more realistic account of human decision making, Selten (2001) and, before, Simon (1991), have called for a theory of bounded rationality that is based on an empirical analysis of the cognitive processes that lead to choice. The formal study of simple heuristics provides one approach towards this end (Gigerenzer and Selten, 2001). A heuristic is a strategy that relies on limited search for information and does not involve the calculation of a maximum or minimum. Instead, it is composed of rules for search, stopping, and decision making consistent with the observation that people often search for information sequentially in time and stop search at some point rather than engaging in exhaustive search. A limitation of this approach is that it has so far not been characterised axiomatically (for an exception, see Rubinstein, 1988).

This chapter gives an axiomatic characterisation of a family of lexicographic theories of choice which include the priority heuristic as a special case. The priority heuristic is a heuristic used to make binary decisions between gambles. The heuristic is remarkable because it predicts the choices between gambles of the majority extremely well (Brandstätter, Gigerenzer and Hertwig, 2006), as well as accounting for a number of violations of expected utility theory, in particular the common consequence and common ratio effects, reflection effects, and the fourfold pattern of risk taking (Katsikopoulos and Gigerenzer, 2008). An axiomatisation will be helpful in at least two ways: first, it will make it possible for theorists to study the relation of the priority heuristic to other axiomatic theories, such as cumulative prospect theory (see Wakker and Tversky, 1993). Second, it will allow for new empirical tests via the axioms of the heuristic. Our axiomatisation is close to Luce (1978).

The representation given here is for a parameterised version of the priority heuristic. While the version of the heuristic with fixed parameters predicts the data nicely, there is a need for a parameterised version as well. For example, parameters are needed in order to account for individual differences, and for inconsistencies in choice (Rieskamp, 2008). The axiomatisation suggested here makes no claims with respect to parameters. Studying this generalisation of the priority heuristic does not mean we advocate a research program in which heuristics are populated with parameters, which are fitted anew to each data set. Rather, we see the generalisation as covering other possible fixed parameters of the priority-heuristic, in the case that independent theory or evidence suggests such fixed values in some situations.

The representation uses semiorders (Luce, 1956), which have the property of having a transitive strict preference part, and an intransitive indifference part. This seems reasonable and consistent with real world evidence, since utility may not be perfectly discriminable. This is argued, for instance, by Armstrong (1950): "The nontransitiveness of indifference must be recognised and explained on any theory of choice, and the only explanation that seems to work is based on the imperfect powers of discrimination of the human mind whereby inequalities become recognisable only when of sufficient magnitude."

We proceed as follows. Section 5.1 introduces the priority heuristic and reviews relevant analytical and empirical results. Section 5.2 gives a brief introduction to measurement theory, the mathematical framework representation results such as the one proffered in this chapter employ. Section 5.3 presents a representation theorem for the heuristic in choices where gambles differ on two attributes (an outcome and a probability). Section 5.4 generalises the result to the case of three attributes (two outcomes and a probability). Section 5.5 concludes with a discussion of the present contribution to the foundations of a theory of bounded rationality in the sense of Selten (2001) and Simon (1991).

5.1 The Priority Heuristic

The priority heuristic is a model of how people make choices between gambles. Its domain are difficult risky-choice problems, that is, problems in which no alternative dominates the other and expected values are close (ratio ≤ 2). A large part of the evidence on people's choice behaviour derives from simple monetary gambles. The priority heuristic proposes that people make choices by using at most three attributes: the minimum outcome, the probability of the minimum outcome, and the maximum outcome. For choosing between two gambles with nonnegative outcomes (then called gains), the priority heuristic has a search rule, stopping rule, and decision rule (Brandstätter, Gigerenzer, and Hertwig, 2006):

Search Rule: Go through attributes in the order: Minimum gain, probability of minimum gain, maximum gain.

Stopping Rule: Stop search if the minimum gains differ by 1/10, or more, of the maximum gain (across the two gambles); otherwise, stop search if probabilities of the minimum gains differ by .1 or more.

Decision Rule: Choose the gamble that is more attractive in the attribute (gain or probability) that stopped search.

The more attractive gamble is the one with the higher (minimum or maximum) gain or with the lower probability of minimum gain. For negative outcomes (the minimum and maximum outcomes are then called losses), the difference in the statement of the heuristic is that "gain" is replaced by "loss". The more attractive loss is the lower one and the more attractive probability of minimum loss is the higher one. Our axiomatisations refer to gambles with gains and it will be obvious how they would be restated for gambles with losses.

Formally, we axiomatise a relation $\succeq$, defined on $A \times B \times C$, where A is the set of minimum outcomes, B is the set of probabilities of minimum outomes, and C is the set of maximum outcomes, such that $(a_1, b_1, c_1) \succeq (a_2, b_2, c_2)$ iff

$$
\begin{array}{rl}
\text{(i)} & a_1 - a_2 > \dfrac{\max\{c_1, c_2\}}{10} \text{ , or} \\
\text{(ii)} & |a_1 - a_2| \leq \dfrac{\max\{c_1, c_2\}}{10} \text{ and} \\
 & b_2 - b_1 > .1 \text{ , or} \\
\text{(iii)} & |a_1 - a_2| \leq \dfrac{\max\{c_1, c_2\}}{10} \text{ and} \\
 & |b_2 - b_1| \leq .1 \text{ and} \\
 & c_1 \geq c_2
\end{array}
$$

The priority heuristic is lexicographic in the sense that an attribute is used for making a choice only if the attributes that precede it in the search order do not allow making a choice (see also Luce, 1956). For more discussion on the heuristic, for example, on why the aspiration levels for stopping search were fixed to .1, see Brandstätter, Gigerenzer, and Hertwig (2006) and Katsikopoulos and Gigerenzer (2008).

To illustrate how the heuristic works, consider one of the problems posed by Allais (1953, p. 527), known as the Allais paradox, where people choose first between gambles A and B:

A:	100,000,000	with probability 1.00
B:	500,000,000	with probability .10
	100,000,000	with probability .89
	0	with probability .01

By subtracting a .89 probability to win 100 million from both gambles A and B, Allais obtained the following gambles, C and D:

C:	100,000,000	with probability .11
	0	with probability .89
D:	500,000,000	with probability .10
	0	with probability .90

The majority of people chose gamble A over B and D over C (MacCrimmon, 1968), and this constitutes a violation of the independence axiom. Expected utility theory does not predict whether A or B will be chosen; it only makes conditional predictions such as "if A is chosen from A and B, then C is chosen from C and D." The priority heuristic, in contrast, makes stronger predictions: It predicts whether A or B will be chosen, and whether C or D will be chosen. Consider the choice between A and B. The maximum gain across the two gambles is 500 million and therefore the aspiration level for gains is 50 million. The difference between the minimum gains equals $100 - 0 = 100$ million, which exceeds the aspiration level, and search is stopped. The gamble with the more attractive minimum gain is A. Thus, the heuristic predicts the majority choice correctly. In the choice between C and D, minimum gains are equal. Thus the next attribute is looked up. The difference between the probabilities of minimum gains equals $.90 - .89 = .01$, which is smaller than the aspiration level for probabilities of .1. Thus the choice is decided by the last attribute, maximum gain, in which gamble D is more attractive. This prediction is again consistent with the choice of the majority.

More generally, Katsikopoulos and Gigerenzer (2008) have mathematically shown that the priority heuristic implies common consequence effects, common ratio effects, reflection effects, and the fourfold pattern of risk attitude. In fact, because the parameters of the heuristic (the order in which attributes are searched, and the aspiration levels that stop attribute search) are fixed, the priority heuristic implies the effects simultaneously.

On the other hand, modifications of expected utility theory, such as cumulative prospect theory (Tversky and Kahneman, 1992), that are consistent with the effects by appropriately setting parameters, cannot simultaneously account for the empirical evidence (Neilson and Stowe, 2002). For instance, the probability weighting functions of cumulative prospect theory, as estimated by Wu and Gonzalez (1996), imply that people will purchase neither lottery tickets nor insurance policies. Neilson and Stowe (2002) also showed that no parameter combinations allow for these two behaviours and a series of choices made by a large majority of participants and reasonable risk premia. Similarly, Blavatskyy (2010) showed that the conventional parameterisations of cumulative prospect theory do not explain the St. Petersburg paradox. Overall, in multi-parameter models, the parameter values fitted to one set of data are not necessarily robust, in the sense of generating accurate predictions for new sets of data. For more on the importance of distinguishing between fitting and prediction in economic modelling, see Harless and Camerer (1994), and Binmore and Shaked (2010).

No model of risky choice can predict people's behaviour in every pair of gambles correctly; therefore it is crucial that researchers refrain from constructing pairs that fit their model when testing it against competing theories. To avoid such a possible bias, Brandstätter, Gigerenzer, and Hertwig (2006) tested the predictive power of the priority heuristic exclusively against sets of gambles designed by the authors of competing theories (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Lopes and Oden, 1999) as well as randomly generated gambles (Erev et al., 2002, see Brandstätter, Gigerenzer, and Hertwig, 2006). These test sets included two-outcome gambles, five-outcome gambles, and choices based on certainty equivalents. Across 260 pairs of gambles, the priority heuristic predicted 87% of majority choices correctly, while cumulative prospect theory predicted 77% (the second most predictive theory was the security-potential/aspiration theory of Lopes and Oden, with 79% of majority choices).

The limits of the predictive power of the priority heuristic were analysed using 450 pairs of gambles designed by Mellers, Weiss, and Birnbaum (1992). The priority heuristic was more predictive than the modifications of expected utility theory when the problems were difficult (i.e., the ratio of the expected values of the two gambles was ≤ 2) but not when problems were easy (ratio > 2) or dominated. For easy problems, however, none of the modifications of expected utility theory could outperform the simple theory of expected value (for a discussion of the evidence, see Birnbaum, 2008, and Brandstätter, Gigerenzer, and Hertwig, 2006). These studies suggest that non-linear transformations of probabilities or monetary values may be needed neither for easy problems nor for difficult ones. Difficult problems can be modeled by the priority heuristic and easy ones by expected value theory, each of which is based on non-transformed values and probabilities. This result clarifies that "overweighting of small probabilities and underweighting of large probabilities", which is often evoked to account for anomalies, is in fact not necessary.

Leland (2010) distinguishes three approaches towards descriptive theories of choice. What he calls the "road taken" is the representation of lotteries as prospects that leads to a preoccupation with explaining violations of independence and has led to a plethora of modifications of expected utility theories, such as prospect theory. A representation of lotteries in terms of Savage's action-by-state matrices instead of prospect, however, makes violations of independence transparent, infrequent, and not the main problem. In this approach, the "road less travelled," more substantial violations such as transitivity and preference for dominated alternatives become more central, as in regret theory (Loomes and Sudgen, 1987). Common to both approaches, nevertheless, is that choices are interpreted as revealing properties of preferences. In the third approach, the "road not taken," choices do not reveal properties of the preferences but instead properties of the decision processes that individuals use to satisfy their preferences. The priority heuristic is a formal model of this third approach, as are the similarity models by Rubinstein (1988), and Leland (1994, 2002).

5.2 Measurement Theory

Measurement theory is predominantly concerned with the question of how certain abstract quantities, such as length, weight, and size, can be associated with numbers. In particular, we would like to attribute numbers to abstract quantities in a systematic fashion: Certain properties, or regularities, of the abstract quantity should be preserved. For instance, the abstract notion of "length" has the property of being additive: If I put one rod next to another rod of the same length (both pointing in the same direction), then the resulting rod will be twice as long as the two individual ones. A representation is, therefore, a transformation which preserves properties. In the present decision theoretic context, the abstract entity to be measured is preference, and the properties to be preserved under the transformation are expressed in the axioms on the preference relation.

In mathematics, transformations such as these are called *homomorphisms*. A homomorphism in general is a structure-preserving map between two algebraic structures; this is expressed in the following quote from *The Foundations of Measurement* (Krantz et al., 1971):

> *...if $\langle A, R_1, ..., R_m \rangle$ is an empirical relational structure and $\langle R, S1, ..., Sm \rangle$ is a numerical relational structure, a real valued function ϕ on A is a homomorphism if it takes each R_i into S_i, $i = 1, ..., m$.*

In our particular case of preference representations, the map ϕ will be between a structure $\langle \mathcal{A}, \succeq \rangle$, where $\mathcal{A}$ is the set of acts, and $\succeq$ is the preference relation defined on $\mathcal{A}$, and a structure $\langle \mathbb{R}, \geq \rangle$, where $\mathbb{R}$ is the real line, and $\geq$ denotes "greater or equal". Particular axioms imposed on $\succeq$ will therefore yield particular kinds of numerical representations for $\geq$.

There are, then, two types of axioms: *Necessary* and *structural* axioms. Loosely speaking, necessary axioms ensure that an appropriate homomorphism ϕ exists. There may be, however, several functions (or a class of functions) which yield the representation, only some of which may be interesting to the case considered. Then, structural axioms constrain the class of functions to those that are of interest; they are used to obtain the uniqueness properties of the representation. Structural axioms tend to lend themselves less easily to a normative reading, since they are used predominantly for technical reasons. For instance, Savage's axiom P6 (Small Event Continuity) implies that events are continuously divisible, making his theory inadequate for cases of countable state spaces. Moreover, Savage's axioms can be categorised as follows: P1 is an ordering axiom, P2 – P4 are independence conditions (which play the role of making the utility and probability components separable and linear), P5 is a non-triviality condition, P6 an Archimedean condition (ensuring that the utility function is real valued), and P7 a dominance axiom (making the representation applicable to non-simple measurable acts). (Fishburn, 1981).

There are several degrees to which a representation can be unique. In particular, the scales used can be nominal, ordinal, interval or ratio scales. Nominal scales use one-to-one transformations. Ordinal scales are unique up to monotonic increasing transformations, i.e. if the functions ϕ and ϕ' both map $\langle \mathcal{A}, \succeq \rangle$ into $\langle \mathbb{R}, \geq \rangle$, then $\phi' = f(\phi)$. Interval scales are unique up to affine (positive monotonic) transformations, such that $\phi' = \alpha\phi + \beta$, where α and β are constants and $\alpha > 0$. Ratio scales are unique up to multiplicative transformation, i.e. $\phi' = \alpha\phi$, with $\alpha > 0$ (Heilmann, 2010). In both the vNM and Savage representation theorems, the probability distribution is unique, and utility is measured on an interval scale.

5.3 Axiomatisation of Two-Attribute Lexicographic Heuristics

5.3.1 Preliminaries

In this section it is assumed that the two gambles have equal minimum gains[1]. This means that in this section we ignore the first step of the priority heuristic where minimum gains are compared.

Let B and C be sets containing the attributes of the gambles. A gamble is a pair (b, c) with $b \in B$ and $c \in C$, where b denotes the probability of the maximum[2] outcome and c the value of the maximum outcome. Let $\succeq$ be a binary relation on $B \times C$, the preference relation over gambles. The relation $\succeq$ is not assumed to be transitive.

Assume that $\succeq$ is *independent* in the following sense: For all b_1, b_2 in B and for all c_1, c_2 in C,

$$(b_1, c_1) \succeq (b_2, c_1) \text{ iff } (b_1, c_2) \succeq (b_2, c_2) \tag{5.1}$$

[1]Some important empirical evidence, such as the possibility effect of Kahneman and Tversky (1979), refers to zero minimum outcomes; theoretically, Rubinstein (1988) also makes this assumption.

[2]The priority heuristic, as stated in Section 5.1, compares probabilities of minimum outcomes. Given the additivity of probabilities, for gambles with two outcomes the probability of the maximum outcome is the complement of the probability of minimum outcomes. For convenience, we consider the mathematically equivalent case where the probabilities of maximum outcomes are compared.

$$(b_1, c_1) \succeq (b_1, c_2) \text{ iff } (b_2, c_1) \succeq (b_2, c_2) \tag{5.2}$$

The property of independence is expressed in the priority heuristic in the sense that the heuristic does not use trade-offs between attributes. Statement (5.1) induces an unambiguous order on B, denoted $\succeq_B$, and statement (5.2) induces the unambiguous order on C, denoted $\succeq_C$.

Furthermore, we define strict preference, $\succ$, and indifference, $\sim$, in terms of $\succeq$ in the usual sense: For all b_1, b_2 in B, and for all c_1, c_2 in C,

$$(b_1, c_1) \succ (b_2, c_2) \text{ iff } (b_1, c_1) \succeq (b_2, c_2) \text{ and not } (b_2, c_2) \succeq (b_1, c_1) \tag{5.3}$$

$$(b_1, c_1) \sim (b_2, c_2) \text{ iff } (b_1, c_1) \succeq (b_2, c_2) \text{ and } (b_2, c_2) \succeq (b_1, c_1) \tag{5.4}$$

Note that neither $\succ$ nor $\sim$ can be assumed to be transitive, since the weak preference relation $\succeq$ is not assumed to be transitive. The strict preference and indifference relations on B and C, $\succ_B$, $\sim_B$, $\succ_C$ and $\sim_C$ are defined similarly.

Next we define another relation on B, denoted P_B, which expresses the lexicographic nature of the decision rule of the priority heuristic. The following definition of P_B expresses the fact that the first attribute searched, probabilities, dominates the second attribute searched, maximum outcomes: For all b_1, b_2 in B

$$b_1 P_B b_2 \text{ iff for every } c_1, c_2 \text{ in } C,\ (b_1, c_1) \succ (b_2, c_2) \tag{5.5}$$

Suppose P is a binary relation on B. Then we can define two other relations, $I(P)$ and $W(P)$ in terms of it. The interpretation we would like to give to these relations is that $I(P)$ is an indifference relation on B, and $W(P)$ is a weak preference relation on B which is defined in a non-standard way (Luce, 1956 and 1978).

$$b_1 I(P) b_2 \text{ iff not } b_1 P b_2 \text{ and not } b_2 P b_1 \tag{5.6}$$

$b_1 W(P) b_2$ iff either

$$\begin{array}{ll} (i) & b_1 P b_2 \text{ or} \\ (ii) & b_1 I(P) b_2 \text{ and there exists a } b_3 \text{ in } B \\ & \text{such that } b_1 I(P) b_3 \text{ and } b_3 P b_2, \text{ or} \\ (iii) & b_1 I(P) b_2 \text{ and there exists a } b_4 \text{ in } B \\ & \text{such that } b_1 P b_4 \text{ and } b_4 I(P) b_2 \end{array} \tag{5.7}$$

This definition expresses the intuition that one probability is weakly preferred to a second probability if (i) the first probability is strictly preferred to the second, or (ii) the first and second probabilities are indistinguishable, and there exists a third probability that is indistinguishable from the first and strictly preferred to the second, or (iii) the first and second probabilities are indistinguishable, and there exists a fourth probability such that the first probability is strongly preferred to the fourth, and the fourth probability is indistinguishable from the second.

This intuition is expressed by the stopping rule of the priority heuristic: a user of the heuristic may weakly prefer obtaining the maximum outcome with a probability of $b_1 = .23$ to obtaining it with a probability of $b_2 = .22$. This weak preference may arise not because s/he has a strong preference for .23 over .22, but rather because s/he cannot discriminate between the two probabilities of .23 and .22, and there exists a third probability, e.g. $b_3 = .33$, such that s/he has a strict preference for .33 over .22 and cannot distinguish between .33 and .23; this is an example of case (ii) just above.

Let us now turn to the definition of a semiorder, as presented by Luce (1956). A semiorder is characterised by the properties of having a transitive strict preference part, and an intransitive indifference part. These properties make semiorders particularly well suited to modelling the behaviour of people who may express indifference between two elements they can essentially not distinguish. Nevertheless, there may exist a threshold beyond which indifference switches to strict preference. A semiorder is defined as follows:

A binary relation P on B is a *semiorder* iff, for all b_1, b_2 in B

$$
\begin{aligned}
&(i) && \text{not } b_1 P b_1 && (5.8)\\
&(ii) && b_1 P b_2 \text{ and } b_3 P b_4 \text{ imply either } b_1 P b_4 \text{ or } b_3 P b_2 &&\\
&(iii) && b_1 P b_2 \text{ and } b_2 P b_3 \text{ imply either } b_1 P b_4 \text{ or } b_4 P b_3 &&
\end{aligned}
$$

The first part of the definition holds that the relation P on B is irreflexive. The second and third parts of the definition convey the intuitive idea that an indifference interval should never span a strict preference interval (see Luce, 1956).

We now turn to the concept of an *indifference interval.* This formalises the idea that even though one attribute value may be weakly preferred to another attribute value, they may not be sufficiently different to induce a strict preference for one of them. The two attribute values will then span an indifference interval, such that all elements of it are considered indifferent, and such that b_1 and b_2 delimit the interval from above and below. For example, in the context of the priority heuristic, the probabilities $b_1 = .33$ and $b_2 = .23$ would span an indifference interval. The concept is formalised as follows:

If P is a semiorder on B, and if $b_1 W(P_B) b_2$ and $b_1 I(P_B) b_2$, the set

$$B(b_1, b_2) = \{b_3 \mid b_1 W(P) b_3 \text{ and } b_3 W(P) b_2\} \tag{5.9}$$

is called an indifference interval.

Finally, we introduce the relation P_C on C which is designed to single out that part of $\succeq$ where the dominant component, B, does not discriminate. Let $B(b_3, b_4)$ be an indifference interval, and let b_1, b_2 be elements of it. Then for all indifference intervals $B(b_3, b_4)$, and for all c_1, c_2 in C

$$c_1 P_C c_2 \text{ iff for every } b_1, b_2 \text{ in } B(b_3, b_4),\ (b_1, c_1) \succeq (b_2, c_2) \tag{5.10}$$

5.3.2 Axioms

Consider a binary relation $\succeq$ on $B \times C$, with the derived concepts $\succeq_B$, $P_B, W(P_B), B(b_3, b_4)$ and P_C defined above.

[**Axiom 1**] $\succeq$ is reflexive, complete, and independent

[**Axiom 2**] P_B is a semiorder

[**Axiom 3**] $W(P_B)$ is identical to $\succeq_B$

[**Axiom 4**] P_C is a simple order

[**Axiom 5**] P_C is identical to $\succeq_C$

[**Axiom 6**] There exists a finite or countable subset of B,
$X = \{..., x_{-2}, x_{-1}, x_0, x_1, x_2, ...\}$ such that for all x_{i-1}, x_i, x_{i+1} in X
(i) $x_i \succ_B x_{i-1}$,
(ii) $B(x_{i-1}, x_{i+1})$ is an indifference interval,
(iii) for b_1 in B, there exists an x_{i-1}, x_i in X with $x_i \succeq_B b_1 \succeq_B x_{i-1}$

[**Axiom 7**] For every b_1 in B, there exists some b_2 in B such that $b_2 I(P_B) b_1$, and for any b_3 in B with $b_3 \succ_B b_2$, then $b_3 P_B b_1$

Axioms 1 to 5 are necessary, whilst Axioms 6 and 7 are structural. This makes Axioms 6 and 7 less suited for constructing empirical tests. Axiom 6 states that the indifference intervals on B span all of B (for the priority heuristic, the entire probability scale), and overlap one another. This axiom ensures that the scales over B and C agree. Axiom 7 requires that the set of elements indistinguishable from a given element be closed from above. Together, Axioms 6 and 7 ensure the existence of a supremum.

Axiom 1 is standard except for the assumption that the preference relation over gambles $\succeq$ is not necessarily transitive, a property that the priority heuristic does not always satisfy. The property of independence implies that each attribute in (b_1, c_1) affects the relation $\succeq$ independently of the other attribute.

Axiom 2 requires that the strictly dominating part of $\succeq$, P_B, is transitive; this follows from the conjunction of statements (i) and (iii) of the definition of a semiorder. The definition of a semiorder implies that indifference intervals cannot cover elements between which there exists a strict preference. Also, the indifference relation $I(P_B)$ will not be transitive, as is the case for the priority heuristic. Axiom 4 is used to impose an order on that part of $\succeq$ where the first component of the tuples (b_1, c_1) does not dominate. Thus if two elements b_3 and b_4 are indistinguishable, then only elements of C should determine choice, and these should be ordered according to a simple order. In particular, Axiom 4 ensures that the restriction of the set $B \times C$ to $B(b_3, b_4) \times C$ agrees with the order between elements of C, which is a simple order.

Axiom 3 forces the order $\succeq_B$ on B induced by independence to be identical to the weak order $W(P_B)$ on B, which was defined in terms of the relation P_B. This implies that both $\succeq_B$ and P_B will be representable using the same numerical scale. Axiom 5 forces the order $\succeq_C$ on C induced by independence to be identical to the simple order P_C on C, which was defined on the indifference intervals only. This implies that both $\succeq_C$ and P_C will be representable using the same numerical scale.

This axiom system is similar to the one used by Luce (1978) for axiomatizing a two-attribute lexicographic model. Luce's (1978) model produces trade-offs between attributes in its second step, whereas this is not the case for the priority heuristic, which considers the second attribute alone when the first attribute does not determine choice.

5.3.3 Representation Theorem

[**Theorem 1**] Suppose $\langle B \times C, \succeq \rangle$ satisfies Axioms 1 - 7. Then there exist real-valued functions ϕ_B and δ_B on B, and ϕ_C on C such that for all b_1, b_2 in B, and c_1, c_2 in C,

1. $\delta_B(b_1) = \sup\limits_{\substack{b_2 \\ b_2 I(P_B) b_1}} [\phi_B(b_2) - \phi_B(b_1)] > 0$

2. $b_1 P_B b_2$ iff $\phi_B(b_1) > \phi_B(b_2) + \delta_B(b_2)$

3. $b_1 W(P_B) b_2$ iff $\phi_B(b_1) > \phi_B(b_2)$

4. $c_1 P_C c_2$ iff $\phi_C(c_1) \geq \phi_C(c_2)$

5. $(b_1, c_1) \succeq (b_2, c_2)$ iff either
(i) $\phi_B(b_1) > \phi_B(b_2) + \delta_B(b_2)$, or
(ii)$-\delta_B(b_1) \leq \phi_B(b_1) - \phi_B(b_2) \leq \delta_B(b_2)$, and $\phi_C(c_1) \geq \phi_C(c_2)$

If $f(.)$ is a strictly increasing and continuous function, and $\alpha, \beta_C > 0$ are constants, then ϕ'_B, δ'_B and ϕ'_C form another representation such that:

$$\phi'_B = f(\phi_B) \quad \delta'_B = f(\phi_B + \delta_B) - f(\phi_B) \quad \phi'_C = \alpha\phi_C + \beta_C$$

If such a representation exists, then Axioms 1 - 5 must hold. For the proof, see Appendix.

5.3.4 Comments

Jointly, Axiom 2 and 4 imply that empirically people find it hard to distinguish between probabilities which are close (Axiom 2), but they can distinguish very well between maximum outcomes (Axiom 4). This prediction about people having different abilities distinguishing outcomes and probabilities is a strong prediction and, to the best of our knowledge, a new one that should be tested empirically. This prediction is indirectly supported by research indicating that (i) people spend more time on outcomes than on probabilities suggesting that outcomes are more important than probabilities (Schkade and Johnson, 1989), (ii) in the extreme, people neglect probabilities altogether, and instead base their choices on the immediate feelings elicited by the gravity or benefit of future events (Loewenstein et al., 2001), (iii) highly emotional outcomes tend to override the impact of probabilities (Sunstein, 2003), (iv) anxiety is largely influenced by the intensity of the shock, not by its probability of occurrence (Deane,1969), and heuristics have been reported that rely on outcomes while ignoring probabilities, but not vice versa (Brandstätter, Gigerenzer & Hertwig, 2006, Table 3).

Theorem 1 axiomatises a class of heuristics of which the priority heuristic is a special case. By setting ϕ_B and ϕ_C to the identity functions and δ_B to .1, the representation expresses the priority heuristic for the case of equal minimum outcomes. Note that whilst setting the function ϕ_B to the identity function, the representation theorem yields a lexicographic structure with linear transformations of the probabilities, the theorem can also yield structures that use non-linear transformations of probabilities.

5.4 Axiomatisation of Three-Attribute Lexicographic Heuristics

5.4.1 Preliminaries

This section extends the framework of section 5.2 to the case of three attributes: minimum outcomes, probabilities of maximum outcomes, and maximum outcomes. We make the simplifying assumption that the maximum outcome, across all choices between gambles that the user of the priority heuristic is considering, is constant, and thus, the first step of the priority heuristic has a constant aspiration level. The general case of varying maximum outcomes remains an open problem that may require a different axiom system[3]. Under our assumption, the approach of Section 5.2 applies smoothly, and thus we go over the main concepts briefly.

Let a gamble be a triple (a, b, c) with $a \in A$, $b \in B$, and $c \in C$, where the set A includes minimum outcomes, B denotes the set of probabilities of maximum outcomes, and C denotes the set of maximum outcomes. Let $\succeq$ be a binary relation on $A \times B \times C$, the preference relation over gambles. The relation $\succeq$ is not assumed to be transitive.

Attributes affect the relation $\succeq$ independently from each other: For all $a_1, a_2 \in A$, for all b_1, b_2 in B, and for all c_1, c_2 in C:

$$(a_1, b_1, c_1) \succeq (a_2, b_1, c_1) \text{ iff } (a_1, b_2, c_2) \succeq (a_2, b_2, c_2) \tag{5.11}$$

$$(a_1, b_1, c_1) \succeq (a_1, b_2, c_1) \text{ iff } (a_2, b_1, c_2) \succeq (a_2, b_2, c_2) \tag{5.12}$$

$$(a_1, b_1, c_1) \succeq (a_1, b_1, c_2) \text{ iff } (a_2, b_2, c_1) \succeq (a_2, b_2, c_2) \tag{5.13}$$

The relations $\succeq_A$ on A, $\succeq_B$ on B, and $\succeq_C$ on C are derived from the independence

[3]In the first step of the priority heuristic, the aspiration level is given by 10% of the maximum outcome across both gambles. On this assumption, the aspiration level is a function of maximum outcomes, thereby violating the independence condition (11) above.

conditions.

The strict preference relation $\succ$ and the indifference relation $\sim$ on $A \times B \times C$ are defined as previously (see Section 5.2.1), and so are the relations $\succ_A$, $\sim_A$ on A, $\succ_B$, $\sim_B$ on B, and $\succ_C$, $\sim_C$ on C. The definitions of the indifference relation $I(P)$ and the weak preference relation $W(P)$ are the same as in Section 5.2.1. Similarly, the concepts of a semiorder and of an indifference interval are defined as previously.

We now define the relations P_A on A, P_B on B, and P_C on C. Thereby, the P_A relation singles out that part of $\succeq$ where the first attribute, minimum values, dominates the other two attributes, probabilities, and maximum outcomes. The P_B relation, by contrast, singles out that part of $\succeq$ where the first attribute, minimum values, does not discriminate, and where the second attribute, probabilities, dominates the third attribute, maximum values. Finally, the relation P_C is defined such that it characterises that part of $\succeq$ where neither the first, nor the second attribute discriminates. We assume that both P_A and P_B are semiorders, and that P_C is a simple order. Indifference intervals induced by the semiorder P_A on A will be called $A(a_3, a_4)$, and indifference intervals induced by the semiorder P_B on B will be called $B(b_3, b_4)$. The resulting structure will thus have nested indifference intervals, capturing the lexicographic nature of the decision rule of the priority heuristic. Consider the following definitions of P_A, P_B, and P_C:

For all a_1, a_2 in A,

$$\begin{aligned} &a_1 P_A a_2 \text{ iff for every } b_1, b_2 \text{ in } B \\ &\text{and for every } c_1, c_2 \text{ in } C,\ (a_1, b_1, c_1) \succ (a_2, b_2, c_2) \end{aligned} \tag{5.14}$$

For all indifference intervals $A(a_3, a_4)$ and for all b_1, b_2 in B,

$$\begin{aligned} &b_1 P_B b_2 \text{ iff for every } a_1, a_2 \text{ in } A(a_3, a_4) \\ &\text{and for every } c_1, c_2 \text{ in } C,\ (a_1, b_1, c_1) \succ (a_2, b_2, c_2) \end{aligned} \tag{5.15}$$

For all indifference intervals $A(a_3, a_4), B(b_3, b_4)$ and for all c_1, c_2 in C,

$$c_1 P_C c_2 \text{ iff for every } a_1, a_2 \text{ in } A(a_3, a_4) \text{ and for every } b_1, b_2 \text{ in } B(b_3, b_4),\ (a_1, b_1, c_1) \succ (a_2, b_2, c_2) \tag{5.16}$$

5.4.2 Axioms

Consider a binary relation $\succeq$ on $A \times B \times C$, and the derived concepts $\succeq_A$, P_A, $W(P_A)$, $A(a_3, a_4)$, $\succeq_B$, P_B, $W(P_B)$, $B(b_3, b_4)$, $\succeq_C$, and P_C defined above.

[**Axiom 1**] $\succeq$ is reflexive, complete, and independent

[**Axiom 2**] P_A is a semiorder

[**Axiom 3**] $W(P_A)$ is identical to $\succeq_A$

[**Axiom 4**] P_B is a semiorder

[**Axiom 5**] $W(P_B)$ is identical to $\succeq_B$

[**Axiom 6**] P_C is a simple order

[**Axiom 7**] P_C is identical to $\succeq_C$

[**Axiom 8**] There exists a finite or countable subset of A, $Q = \{..., q_{-2}, q_{-1}, q_0, q_1, q_2, ...\}$ such that for all q_{i-1}, q_i, q_{i+1} in Q
(i) $q_i \succ_A q_{i-1}$,
(ii) $A(q_{i-1}, q_{i+1})$ is an indifference interval,
(iii) for a_1 in A, there exists an q_{i-1}, q_i in Q with $q_i \succeq_A a_1 \succeq_A q_{i-1}$

[**Axiom 9**] There exists a finite or countable subset of B, $X = \{..., x_{-2}, x_{-1}, x_0, x_1, x_2, ...\}$ such that for all x_{i-1}, x_i, x_{i+1} in X
(i) $x_i \succ_B x_{i-1}$
(ii) $B(x_{i-1}, x_{i+1})$ is an indifference interval
(iii) for b_1 in B, there exists an x_{i-1}, x_i in X with $x_i \succeq_B b_1 \succeq_B x_{i-1}$

[**Axiom 10**] For every a_1 in A, there exists some a_2 in A such that $a_2 I(P_A) a_1$, and for any a_3 in A with $a_3 \succ_A a_2$, then $a_3 P_A a_1$

[**Axiom 11**] For every b_1 in B, there exists some b_2 in B such that $b_2 I(P_B) b_1$, and for any b_3 in B with $b_3 \succ_B b_2$, then $b_3 P_B b_1$

5.4.3 Representation Theorem

[Theorem 2] Suppose $\langle A \times B \times C, \succeq \rangle$ satisfies Axioms 1 - 11. Then there exist real-valued functions ϕ_A and δ_A on A, ϕ_B and δ_B on B, and ϕ_C on C such that for all a_1, a_2 in A, b_1, b_2 in B, and c_1, c_2 in C,

1. $\delta_A(a_1) = \sup_{\substack{a_2 \\ a_2 I(P_A) a_1}} [\phi_A(a_2) - \phi_A(a_1)] > 0$

2. $\delta_B(b_1) = \sup_{\substack{b_2 \\ b_2 I(P_B) b_1}} [\phi_B(b_2) - \phi_B(b_1)] > 0$

3. $a_1 P_A a_2$ iff $\phi_A(a_1) > \phi_A(a_2) + \delta_A(a_2)$

4. $a_1 W(P_A) a_2$ iff $\phi_A(a_1) > \phi_A(a_2)$

5. $b_1 P_B b_2$ iff $\phi_B(b_1) > \phi_B(b_2) + \delta_B(b_2)$

6. $b_1 W(P_B) b_2$ iff $\phi_B(b_1) > \phi_B(b_2)$

7. $c_1 P_C c_2$ iff $\phi_C(c_1) \geq \phi_C(c_2)$

8. $(a_1, b_1, c_1) \succeq (a_2, b_2, c_2)$ iff either

(i) $\phi_A(a_1) > \phi_A(a_2) + \delta_A(a_2)$, or
(ii) $-\delta_A(a_1) \leq \phi_A(a_1) - \phi_A(a_2) \leq \delta_A(a_2)$, and $\phi_B(b_1) \geq \phi_B(b_2)$, or
(iii) $-\delta_A(a_1) \leq \phi_A(a_1) - \phi_A(a_2) \leq \delta_A(a_2)$, and $-\delta_B(b_1) \leq \phi_B(b_1) - \phi_B(b_2) \leq \delta_B(b_2)$, and $\phi_C(c_1) \geq \phi_C(c_2)$

If $f(.), g(.)$ are strictly increasing and continuous functions, and $\alpha, \beta_C > 0$ are constants, then ϕ'_A, δ'_A, ϕ'_B, δ'_B, and ϕ'_C form another representation such that:

$\phi'_A = f(\phi_A) \quad \delta'_A = f(\phi_A + \delta_A) - f(\phi_A)$

$\phi'_B = g(\phi_B) \quad \delta'_B = g(\phi_b + \delta_B) - g(\phi_B)$

$\phi'_C = \alpha\phi_C + \beta_C$

If such a representation exists, then Axioms $1 - 7$ must hold. The proof of Theorem 2 is a straightforward extension of the proof of Theorem 1 (see Appendix).

5.4.4 Comments

The interpretation of the axioms mirrors that given in Section 5.2.2. Note that Axiom 2 implies that people find it hard to distinguish between close minimum values. This is a strong prediction of the framework and should be tested empirically. We are not aware of any research that has addressed this hypothesis.

Theorem 2 axiomatises a class of heuristics of which the priority heuristic is a special case. By setting ϕ_A, ϕ_B, and ϕ_C to identity functions, δ_A to some constant, and δ_B to .1, the representation expresses the priority heuristic for the case of a constant maximum outcome.

5.5 Towards a Theory of Bounded Rationality

The term "bounded rationality" has been used for at least three different research programs: Optimisation under constraints (e.g. Sargent, 1993), deviations from optimisation (e.g. Kahneman, 2003), and for the study of decision processes in situations where optimisation may be out of reach (Gigerenzer and Selten, 2001; Simon, 1955; Selten, 2001). Note that these three uses are not the same. The first two emphasise rationality and irrationality, respectively, but share optimisation as a reference point. The third program models the process of decision rather than optimisation or deviations from optimisation. As mentioned before, in this program, choices reveal decision processes (Leland, 2010). The priority heuristic is such a formal model of the decision process. The three building blocks – rules for search, stopping, and decision – are also part of other heuristics in what is termed the "adaptive toolbox" of humans (Gigerenzer and Selten, 2001). To date, the study of bounded rationality has accumulated converging evidence that heuristics can model decision making in both experts and laypeople, and that heuristics can often make more accurate predictions than can complex forecasting models, including linear regression, neural networks, Bayesian models, and classification trees (Katsikopoulos, 2011). Yet as Selten (2001, p. 14) noted, a comprehensive, coherent theory of bounded rationality is not yet available.

This chapter is a step in the direction of providing a theory of bounded rationality, in particular, by providing greater conceptual clarity through the use of an axiomatic representation. In general, axioms give exact behavioural characterisations which can be tested empirically. Also, by using axioms on (unobservable) preference relations and thereby yielding a representation result which models decisions consistent with heuristics, this chapter provides a link between existing axiomatic theories of decision making and bounded rationality.

The contribution made here can be seen as an exercise consistent with the empiricist school of thought: Starting from observable phenomena, by abstraction a theory is derived – the priority heuristic –, and from the theory, we obtain mathematical concepts – the axiomatisation. This contrasts with approaches in the tradition of expected utility theory, where mathematical principles are used on the basis of an *a priori* notion of rationality, rather than on the basis of evidence. However, our approach is, in fact, consistent with the origins of probability and decision theory: Decision theory was first studied by Blaise Pascal and Pierre de Fermat as an attempt to understand gambling behaviour. The priority heuristic is a theory which, as explained above, predicts just these choices between gambles well, and is therefore a good starting point for the derivation of axiomatic characterisations of bounded rationality.

A.1 Appendix

Proof of Theorem 1

Sufficiency

Statement 1 and 2: Define δ_B by Statement 1. Axioms 6 and 7 insure the existence of the supremum, and by Axioms 6(ii) and 6(iii), $\delta_B > 0$. By the definition of δ_B and Axiom 7, Statement 2 follows.

Statement 3: By Axioms 2 and 6, and Theorem 16.7 of Suppes et al. (1989), there exists a real-valued function ϕ_B on B such that $b_1 W(P_B) b_2$ iff $\phi_B(b_1) > \phi_B(b_2)$, and with the asserted uniqueness properties.

Statement 4: By Axiom 4 and Theorem 2.1 of Krantz et al. (1971), there exists a real-valued function ϕ_C on C such that $c_1 P_C c_2$ iff $\phi_C(c_1) \geq \phi_C(c_2)$, and with the asserted uniqueness properties.

Statement 5: Statement 4 says that ϕ_C preserves P_C. By the definition of P_C, it is identical to $\succeq$ when $\succeq$ is applied to $B(b_3, b_4) \times C$ and restricted to C. So, ϕ_C also preserves the order $\succeq$ when it is applied to $B(b_3, b_4) \times C$ and restricted to C. By Axiom 6 (ii), there are successive indifference intervals on B with nontrivial regions of overlap. Forcing the local scales to agree yields a global scale on $B \times C$. The restriction of this scale to C, ϕ_C preserves P_C as well. Statement 5 follows from this, together with the other four statements and the whole construction.

Necessity of Axioms 1 to 5

Axiom 1: The reflexivity and completeness of $\succeq$ follow immediately from Statement 5. To show independence of the first attribute from the second, consider a c_1 in C and assume $(b_1, c_1) \succeq (b_2, c_1)$. By Statement 5, this means $\phi_B(b_1) > \phi_B(b_2) + \delta_B(b_2)$, which in turn means that $(b_1, c_2) \succeq (b_2, c_2)$ for any c_2 in C. To show independence of the second attribute from the first, consider a b_1 in B and assume $(b_1, c_1) \succeq (b_1, c_2)$. By Statement 5, this means that $\phi_C(c_1) > \phi_C(c_2)$, which in turn means that $(b_2, c_1) \succeq (b_2, c_2)$ for any b_2 in B.

Axiom 2: **Part (i)** of the definition of a semiorder follows immediately from Statement 2.

For **Part (ii)** of the definition, we assume $b_1 P_B b_2$, $b_3 P_B b_4$ and show that if also not $b_1 P_B b_4$, then $b_3 P_B b_2$. By Statement 2, $b_1 P_B b_2$ implies $\phi_B(b_2) + \delta_B(b_2) < \phi_B(b_1)$, and not $b_1 P_B b_4$ implies $\phi_B(b_1) \leq \phi_B(b_4) + \delta_B(b_4)$. Thus, also $\phi_B(b_2) + \delta_B(b_2) < \phi_B(b_4) + \delta_B(b_4)$. This, together with $\phi_B(b_4) + \delta_B(b_4) < \phi_B(b_3)$ (which holds from $b_3 P_B b_4$ and Statement 2), means that $\phi_B(b_2) + \delta_B(b_2) < \phi_B(b_3)$, or, by Statement 2, $b_3 P_B b_2$.

For **Part (iii)** of the definition of a semiorder, we assume $b_1 P_B b_2$ and $b_2 P_B b_3$, and considering a b_4 in B, we show that either $b_4 P_B b_3$ or $b_1 P_B b_4$. Specifically, we show that, if **(a)** $\phi_B(b_4) \geq \phi_B(b_2)$, then $b_4 P_B b_3$, and if **(b)** $\phi_B(b_4) < \phi_B(b_2)$, then $b_1 P_B b_4$.

For **(a)**, $b_2 P_B b_3$ implies, by Statement 2, that $\phi_B(b_2) > \phi_B(b_3) + \delta_B(b_3)$. Together with $\phi_B(b_4) \geq \phi_B(b_2)$, this means $\phi_B(b_4) > \phi_B(b_3) + \delta_B(b_3)$, or, by Statement 2, $b_4 P_B b_3$.

For **(b)**, we first show that $b_1 P_B b_4$ holds if additionally $\phi_B(b_4) + \delta_B(b_4) \leq \phi_B(b_2) + \delta_B(b_2)$. This, together with $\phi_B(b_2) + \delta_B(b_2) < \phi_B(b_1)$ (by $b_1 P_B b_2$ and Statement 2), means that $\phi_B(b_4) + \delta_B(b_4) < \phi_B(b_1)$, or, by Statement 2, $b_1 P_B b_4$ as required.

To complete the argument, we show by contradiction that $\phi_B(b_4) + \delta_B(b_4) \leq \phi_B(b_2) + \delta_B(b_2)$. Suppose $\phi_B(b_4) + \delta_B(b_4) > \phi_B(b_2) + \delta_B(b_2)$. Then it is possible to find a b_5 in B such that: $\phi_B(b_4) + \delta_B(b_4) = \phi_B(b_5) > \phi_B(b_2) + \delta_B(b_2)$. By Statement 2, $\phi_B(b_5) > \phi_B(b_2) + \delta_B(b_2)$ implies $b_5 P_B b_2$.

By Statement 2, $\phi_B(b_4) + \delta_B(b_4) = \phi_B(b_5)$ implies that not $b_5 P_B b_4$. Also, by Statement 1,$\phi_B(b_4) + \delta_B(b_4) = \phi_B(b_5)$ implies that $\phi_B(b_4) < \phi_B(b_5) < \phi_B(b_5) + \delta_B(b_5)$. By Statement 2, this implies that not $b_4 P_B b_5$. Together, not $b_5 P_B b_4$ and not $b_4 P_b b_5$ imply that $b_5 I(P_B) b_4$.

By the assumption of **(b)**, $\phi_B(b_4) < \phi_B(b_2)$ and by Statement 1, $\phi_B(b_4) < \phi_B(b_2) + \delta_B(b_2)$. By Statement 2 this implies that not $b_4 P_B b_2$. Furthermore, from $\phi_B(b_4) + \delta_B(b_4) > \phi_B(b_2)$, which we assumed for contradiction, it follows that not $b_2 P_B b_4$. From not $b_4 P_B b_2$ and not $b_2 P_B b_4$ it follows that $b_4 I(P_B) b_2$.

Having established $b_5 I(P_B) b_4$, $b_4 I(P_B) b_2$ and $b_5 P_B b_2$, by the definition of weak preference $b_4 W(P_B) b_2$.

By Statement 3, $b_4 W(P_B) b_2$ implies $\phi_B(b_4) > \phi_B(b_2)$ which is inconsistent with the assumption of **(b)**, $\phi_B(b_4) < \phi_B(b_2)$. Whence, $\phi_B(b_4) + \delta_B(b_4) \leq \phi_B(b_2) + \delta_B(b_2)$ as required.

Axiom 3: By Statement 5, ϕ_B preserves the order $\succeq_B$ and by Statement 3, ϕ_B preserves the order $W(P_B)$, so $\succeq_B$ and $W(P_B)$ are identical.

Axiom 4: By Statement 4 and Theorem 2.1 of Krantz et al. (1971), Axiom 4 follows.

Axiom 5: By Statement 5, ϕ_C preserves the order $\succeq_C$ and by Statement 4, ϕ_C also preserves P_C, so $\succeq_C$ and P_C are identical.

Chapter 6

Conclusion

6.1 Introduction

Taking Savage's subjective expected utility theory as a starting point, this thesis has argued for the distinction between different types of uncertainty: ambiguity, option uncertainty and state space uncertainty. We have argued that it is essential to understand the nature of uncertainty – and, in particular, the idea that not all uncertainties are alike – to be able to model decision making in a large and uncertain world in a precise way. Real world decisions are rarely clear-cut cases of small world decision making, where the agent's uncertainty can be reduced to uncertainty over what the true state is. Once one grants this claim, a number of "anomalies" in decision making, which may otherwise seem irrational, become comprehensible.

This chapter will return to the fundamental arguments on which this thesis is based. In particular, this chapter gives an argument for pluralism in decision theory, perhaps the most contentious claim of this thesis. Our pluralist view of decision theory implies that the rationality conditions imposed under uncertainty are different from those imposed within a typical small world. We conclude with an overview of the applications of decision making under the types of uncertainty characterised in this thesis.

6.2 Idealisation and abstraction

Savage's theory of decision making in small worlds abstracts from all particulars a decision situation might have, condensing decision making to its very essence: all decisions are based on beliefs and desires. Consistency of one's decision with one's personal beliefs, as characterised in the probability function over the state space, and desires, as expressed in the agent's utility function over consequences, is then the criterion for optimality. In a recent paper, Hosni (forthcoming) phrases this as follows:

> *"Standard Bayesianism can be fruitfully seen as the solution to the following problem: [...] How should a maximally idealised agent behave when facing a maximally abstract choice problem?"*

An *idealised* agent then has no cognitive limitations, it is this assumption which motivates the rationality assumption on the part of the agent. An agent is said to be idealised when limitations in time, information, and computational capacities play no role in the decision making process of the agent. The idealisation assumption on the agent gives rise to the normative content of Bayesian decision theory, as an idealised agent will not make any mistakes, and can therefore be seen as making optimal decisions that more limited agents should strive to attain. In an *abstract* decision problem, the specific situation the agent is faced with is reduced to only those features which are decision-relevant. Savage's theory can be seen as an abstract theory in the sense that a small world model is designed to capture every relevant aspect of the decision maker's problem. An abstracted decision problem contrasts with the real world decision problem, which contains details which are irrelevant to the decision problem.

This thesis has maintained both the idealisation and abstraction assumption above. In particular, we have argued that the uncertainty the agent perceives prevents the agent from modelling a given decision problem as a small world, not, as one might think, because the agent has cognitive limitations which prevent them from taking more optimal decisions, but rather because the agent faces situations of uncertainty that even an idealised agent cannot respond to more optimally. For instance, consider again John and Lisa's problem (see section 3.1), where John and Lisa consider buying an insurance policy against heart disease. John and Lisa obtain contradictory figures regarding the likelihood of developing such a disease. Even when we assume that John and Lisa have no cognitive constraints, it is not clear what the rational response to the problem is.

However, the nature of the idealisation of the agent changes as one considers decision problems under uncertainty greater than that compatible with Savage's theory. Savage's theory assumes that the agent will perceive any given decision problem as a small world, such that no uncertainty other than that over what state in a given state space is true matters to the decision. In this thesis, this latter assumption is relaxed, such that agents are permitted to perceive the uncertainty over the true consequences of their actions, or uncertainty over what probability distribution over the state space is true. Arguably, such an agent can be seen as less idealised, as they are presumed to be incapable of performing the reduction of uncertainty required to model a decision problem using a small world.

We have also maintained Savage's abstraction assumption, as the large world frameworks we constructed differed from the real world problems in the sense that all irrelevant details which a real world problem may contain were considered immaterial. The large world decision problems we constructed were assumed to be exactly the problem the agent faces; no relevant details were assumed to be elided, and no irrelevant details were included in the representation of the problem.

6.3 Pluralism of decision theory under uncertainty

Thus, whilst we grant that Savage's decision theory answers the question how a maximally idealised agent should behave when facing a maximally abstract choice problem for the most part successfully, two different readings of Savage's theory are possible. On a first reading, Savage's theory can appear to be a *general* theory, as prima facie, it seems possible to cast *any* particular decision problem in the mould of Savage's decision theory. On a second view, Savage's theory is a *specific* theory, and only some, maybe even few, decision problems can be analysed in the fashion of Savage's framework. This thesis is committed to this latter view, and it is on this position that our argument for a pluralistic approach to decision theory is based.

To rephrase these views in the terminology of this thesis, on the view that Savage's theory is general it is the case that subjective expected utility theory applies to *both* small and large worlds, whereas on the view that Savage's theory is specific, it is applicable *only* to small worlds. Binmore (2007) refers to adherents of the first view as "Bayesianites", and proponents of the latter view as "Bayesians", where he counts himself as a Bayesian, but not as a Bayesianite (Binmore, 2007, 2009). Quite conceivably, Savage himself would side also with the Bayesian, rather than Bayesianite, view, for he holds that "the "Look before you leap" principle is preposterous if carried to extremes" (Savage, 1954, p. 16). Moreover, the latter half of Savage's *Foundations of Statistics* is committed to characterising decision making under complete ignorance, which suggests that Savage adheres to the view that under extreme uncertainty, subjective expected utility theory is not applicable.

The kinds of large worlds this thesis has identified can be seen as intermediate between complete ignorance and cases of mild uncertainty typically modelled within small world matrices. Ambiguity, option uncertainty and state space uncertainty are each cases which deviate from the typical small world setting in the sense that there is a source of uncertainty which the small world model does not capture, but each case deviates from the small world model only via a minimal extension. Due to the strong similarity between small world models and the large world models suggested here, it may be tempting to ask "wouldn't the decision maker be *more* rational if they modelled the problem as a small world decision matrix?" We answer this question in the negative: under uncertainty, deviations from the behaviour which is rational in the small world setting are permissible. An agent who treats a large world problem as a small world model will act as if the greater uncertainty present in the large world were irrelevant. This strategy will not necessarily be successful, however, unless uncertainty really is irrelevant. Otherwise, the agent may be unpleasantly surprised by some factor that the small world decision matrix they constructed did not take account of (Binmore, 2009).

For instance, consider a decision maker who would like to decide between buying a Porsche and a Ferrari, and assume that since the decision maker has not driven a sports car before, they are unsure what value best represents their desire for the consequence of possessing a Porsche or Ferrari. When the agent treats this decision problem as a small world, attaching a unique utility value to the consequences of all acts, then on buying the Porsche, the agent may find that the greater noise of the car and the somewhat uncomfortable seats do not generate quite as high a utility value as the agent's small world model attributed to the consequences of buying the car. In short, when an agent treats a large world problem as a small world, the agent risks being unpleasantly surprised.

Savage's theory is *specific* in the sense that it applies only to small worlds, namely cases where it is possible to reduce uncertainty over what to do to uncertainty over what the true state is. As this thesis has aimed to show, such a representation cannot always be found, and in those cases where it is not possible to represent the decision problem as a small world, Savage's theory must be extended appropriately. Savage's theory can be extended to situations of uncertainty by asking the same question Savage's theory replies to, namely that of how an idealised agent should behave in a maximally abstract decision problem, but by deviating from Savage's theory by answering this question in a manner which takes the greater uncertainty of large worlds into account. Using this strategy one can apply variants of Savage's theory to worlds larger than those considered by Savage.

We have argued in this thesis that depending on the kind of uncertainty faced by the agent, different normative constraints must be placed on the preferences of the agent. For instance, Chapter 3 has argued that under ambiguity, Savage's axioms P2 and P4 are not compelling as requirements of rationality. Chapter 4 has argued that under option uncertainty, Ghirardato's axioms 8 and 9 can be seen as rationality postulates governing the attitude the agent has with respect to option uncertainty. The view that different types of uncertainty require different rationality postulates on the agent's preferences can be seen as a pluralistic account of decision making, which denies that a single set of axioms is valid for all possible decision problems.

A related stance to the concept of rationality has been taken by Gilboa, Postlewaite and Schmeidler (2009). The authors argue that there may not be a unique set of axioms or rules which can be seen as synonymous with the notion of rationality; according to Gilboa et al., rationality is not a binary notion. In particular, the authors argue that "the quest for a single set of rules that will universally define *the* rational choice is misguided". This thesis concurs with Gilboa et al. in the sense that we adhere to the view that a response which is rational in a situation of option uncertainty may appear irrational in the context of ambiguity, and a decision made under ignorance may be far from optimal, but, given the constraints of the situation, rational nevertheless. The notion of rationality which is appropriate must be *relative* to the decision problem faced by the agent, and in particular relative to the severity of uncertainty the agent must grapple with.

However, the pluralism this thesis advocates is not an unconstrained one; Chapter 2 has shown that in some cases, reductions of one type of uncertainty to another are admissible, as no further insight is gained by distinguishing certain types of uncertainty from others. Reduction can and should be conducted to the extent that the reduced representation fully captures the decision problem faced by the agent. Pluralism is required only where the pluralistic account yields new, and more convincing, theoretical insights than the more general, reduced theory can achieve. Conversely, we have not claimed that the classification of uncertainty this thesis has advocated is exhaustive. It may well be that there exist other decision-relevant kinds of uncertainty we have failed to distinguish.

The pluralist view of decision theory advocated here implies that it may be the case that two different representations for a given decision problem exist; for instance, some decision problems may either be cast into a model with option uncertainty, or as a problem featuring ambiguity. This raises the question which, if any, is the "right" representation for the decision problem, and how the agent should evaluate the possible different frameworks against each other. In answering this question, it is useful to return to Savage's example of a man buying a car and pondering the question whether or not to buy it with a radio installed, which we discussed in section 1.4. As we explained in Chapter 1, Shafer's (1986) response to Savage's example is that ultimately it is up to the decision-making agent to decide which representation is best suited to making an optimal decision; there is no unique framework which is objectively the best representation. A similar argument must hold true for an agent who is debating whether to analyse a given decision using a representation featuring option uncertainty versus a representation containing ambiguity. This is consistent with the subjective nature of the optimality of a decision problem: In Savage's framework, a s decision is optimal if taken consistently with the agent's personal beliefs and desires. There exists no extraneous device which would make the optimality of the agent's decision verifiable.

Let us investigate this argument in greater detail. As discussed in Chapter 2, a decision problem containing option uncertainty may be converted into a decision problem containing ambiguity via a refinement of the state space. A decision problem containing option uncertainty would be fully reduced to a problem with ambiguity if it is possible to refine the state space sufficiently to obtain unique consequences at every state. Yet, as we argued in Chapter 2, such a reduction may not be possible in all cases, such that there remains residual uncertainty with respect a particular consequence at a particular state. We may ask then, to what extent the decision making agent should aim to reduce option uncertainty to ambiguity. In answering this question, it is useful to remind ourselves of the fact that the reduction of option uncertainty to ambiguity will not eliminate uncertainty, but rather convert one kind of uncertainty to another. Given that uncertainty cannot be eliminated via refinement, which level of refinement is best suited to the analysis of the decision problem depends on the subjective stance of the agent towards the decision problem. A decision problem should be modelled using a decision matrix containing option uncertainty whenever the agent perceives option uncertainty as relevant to their decision problem; similarly, a decision problem should be modelled using a decision matrix containing ambiguity whenever the agent perceives the decision problem as ambiguous.

6.4 The role of heuristics under uncertainty

This thesis has argued that whilst under mild uncertainty (see section 2.3) Savage's framework is valid from a normative point of view, in large world scenarios Savage's theory is limited. In large and uncertain worlds, the limitations of the agent's information must be taken into account, and decision rules inconsistent with Savage's theory may be rational. Whilst Savage's framework is convincing as a normative model for decisions under mild uncertainty, we may ask also what models are descriptively adequate. Chapter 5 of this thesis investigates the normative implications of such a descriptively successful model, namely the priority heuristic. By axiomatising the priority heuristic, this thesis allows for detailed comparisons between, for instance, Savage's theory and the priority heuristic. Although we do not wish to advocate the priority heuristic as a normatively valid model for decision making under mild uncertainty, descriptive theory is an important domain of research in its own right, particularly with a view to prediction.

While the priority heuristic is strong descriptive model for choice under risk, it would be interesting for further research to study a variant of the priority heuristic adequate to situations of uncertainty. In particular, as this thesis showed normative claims must be weaker under uncertainty than under risk. The concepts of rationality and optimality must of necessity be relative to a given (small world) model, and consequently relative to the information the agent has. However, under extreme cases of uncertainty, e.g. the absence of all probabilistic information, heuristic decision making may be appropriate. Under severe uncertainty, bounded rationality and rationality may coincide.

One reason why heuristics are a successful decision-making strategy in the face of uncertainty is that under uncertainty, the decision maker's task changes from one of choice (that of choosing the best action relative to a given model) to a task of inference (that of predicting how the world will evolve). Savage's framework relies on the idea that the agent is exclusively faced with a choice, but not an inference task. Given a small world model, Savage's framework requires the agent to weigh consequences by their respective probabilities, and add over the possible states of the world; a decision which maximises subjective expected utility is then called rational. Under uncertainty, however, the agent must make good decisions despite the fact that they are not given a particular small world environment; Savage's weighing and adding strategy may be out of reach. Agents faced with changing environments will then be confronted with a task of making *robust* decisions, i.e. decisions which sacrifice optimality with respect to a particular decision environment for success over a broad range of the decision environments. It is the fact that under uncertainty, optimality may be out of reach which makes heuristic decision rules adequate to situations of uncertainty.

To date, there exists no heuristic which would be suitable to modelling choice under uncertainty; this is a task future research may address. However, heuristic decision rules appropriate for situations of uncertainty may seem more rational than one might expect, potentially yielding close analogies between normative models and models of bounded rationality.

6.5 Applications

Uncertainty is an endemic feature of decision making, and taking account of uncertainty implies that we can model with exactitude decision problems which do not fit the mould of the small world. Once one models decision making in large worlds precisely, many empirical phenomena which may otherwise seem irrational become comprehensible. Let us consider how the concepts of ambiguity, option uncertainty and state space uncertainty can be applied to various domains of theoretical and empirical inquiry. These domains of applications may serve as a basis for further research, as the implications of the presence of the types of uncertainty identified in this thesis have not been studied in great detail to date.

Ambiguity

Chapter 2 of this thesis has identified ambiguity with uncertainty over the correct probability distribution over the state space. This kind of uncertainty may affect decision making in many particular instances, as we rarely hold sufficient information to assign unique prior probabilities to all states. However, to date much of economic theory is based on Bayesian decision theory. It is therefore interesting to consider how relaxing the first tenet of Bayesianism helps in explaining experimental evidence.

A review of the possible applications of the ambiguity literature to economic theory is contained in Mukerji and Tallon (2004), who identify three domains in which economic modelling may gain from modelling ambiguity and ambiguity attitude: financial markets, contract theory and game theory. In each case, uncertainty over the correct probability distribution over the state space may affect economic decision making.

With respect to *financial markets*, Dow and Werlang (1992) applied Schmeidler's (1989) Choquet expected utility model to portfolio choice of agents, and show that when an agent is ambiguity averse, there may exist a nondegenerate price interval at which the agent will strictly prefer a zero position in a risky asset to either buying or selling it short. Such an interval would be reduced to a unique point in the case of an expected utility decision maker, who would switch between buying and short selling at that point.

Epstein and Wang (1994) extend Gilboa and Schmeidler's (1989) multiple prior model to a dynamic setting, and show that in an economy where prices of assets are determined at an equilibrium, indeterminacy in the equilibrium prices can arise. This implies that a large volatility in prices may be consistent with equilibrium. Epstein and Miao (2003) use this finding as an explanation of the *home bias* in asset demand: agents buy more assets from their own country than from foreign countries. This can be explained by ambiguity aversion, provided that agents perceive home assets as less ambiguous in their payoffs than foreign assets.

The effect of ambiguity aversion on *optimal risk-sharing* arrangements in contracts is studied by Chateauneuf, Dana and Tallon (2000). The authors show that in a general equilibrium setting, the Pareto-optimal outcome will obtain when agent's preferences satisfy the axioms of the CEU model, and when all agents hold the same beliefs. However, when agents do not hold the same beliefs, the Pareto-optimal outcome may not obtain.

Mukerji (1998) investigates the effects of the presence of ambiguity on *incentive contracts*, the implications of which hinge on contingent events. In particular, when agents are ambiguity-averse, the best possible contracts may be incomplete and inefficient.

Finally, the concept of ambiguity has been applied to *non-cooperative game theory*. Lo (1996) gives a definition of strategic equilibrium in normal form games when agents hold MEU preferences. Dow and Werlang (1994), Klibanoff (1996) and Marinacci (2000) respectively define equilibrium concepts with ambiguity aversion which differ from Lo's as they do not restrict equilibrium beliefs to only those which are best responses; they therefore allow other priors than only those which are best responses as equilibrium beliefs. Thereby, the set of rational equilibrium strategies is larger than that envisage by Lo.

These applications of the concept of ambiguity show that many empirical observations, such as, for instance, the home bias in asset demand, become comprehensible once we grant that ambiguity may affect decision making. Our stance that ambiguity may be objectively given substantiates the view that these empirical phenomena do not arise out of the irrationality of agents, but rather constitute rational reactions to the presence of ambiguity. It may be interesting, then, to study the implications of the presence of objective ambiguity on financial decision making, contract theory and game theory. In each of these cases, the assumption that ambiguity is an objective feature of the decision problem seems justified. The framework suggested in Chapter 3 provides one way of modelling ambiguity objectively, and can be used to study ambiguity in a variety of economic contexts.

Another avenue for future research is to examine the connection between ambiguity and social norms. In real world settings where agents do not have access to probabilistic information, or where probabilistic information is sparse, social norms may dictate particular responses to ambiguity. For instance, threats from unknown diseases, such as BSE (Bovine Spongiform Encephalopathy) are perceived by the public as particularly severe, and ambiguity averse responses are common (Anand, 2002). Similarly, Gigerenzer (2006) shows that in the aftermath of the September 11th, 2001 terrorist attacks, fatalities due to road traffic accidents peaked as a result of the fact that agents who would otherwise travel by airplane chose to travel by car instead. Thereby, terrorist attacks are one instance of low probability, high damage events, so-called *dread risks*. Both in the case of BSE and in the case of terrorist attacks, we can assume that risks are poorly understood by the public, and are hence perceived as ambiguous. There is then a social norm to respond to this ambiguity in an an extremely ambiguity-averse manner.

Option uncertainty

In chapter 2, we identified option uncertainty with uncertainty over the true outcome of one's action at any given state, and we have argued that option uncertainty is separate from ambiguity. Option uncertainty may affect decision making in many real world situations, and the theoretical study of option uncertainty may yield interesting theoretical insights.

Chapter 4 has shown that Eliaz and Ortoleva (2011) have conducted an experiment the results of which can be interpreted as revealing option uncertainty aversion. It may be interesting for further research to study option uncertainty aversion in greater detail, and to investigate the relation between option uncertainty aversion and ambiguity aversion. One question which further research might address is whether ambiguity averse agents are also option uncertainty averse, and if so, to what degree. Eliaz and Ortoleva's variation of Ellsberg's paradox featuring what we call option uncertainty is a good starting point for further investigation.

Moreover, Chapter 4 has shown that option uncertainty aversion can explain *status quo bias*. Perhaps it would be possible to study the relation between option uncertainty aversion and status quo bias experimentally, for instance, by comparing the decisions an agent makes when no status quo is singled out with those when the agent is endowed with a certain status quo gamble. Using a similar set-up to Eliaz and Ortoleva's, one could test whether agents who are more averse to option uncertainty reveal a greater bias toward the status quo.

In Chapters 2 and 4 we have also argued that one possible interpretation of option uncertainty is ethical uncertainty, namely uncertainty regarding what values best reflect the agent's desire for consequences. This interpretation may be useful whenever decisions which have an ethical aspect need to be made. For instance, this might be the case in military decision making; in section 2.1, we gave the example of the Head of State of Isreal deciding on whether to launch an attack on Iran. Uncertainty over the value of the consequences of launching an attack may have an impact of the agent's decision making process. A further domain where ethical uncertainty may be particularly important is medical decision making. It may be interesting to apply the concept of option uncertainty to these two domains, and to study ethical uncertainty empirically.

Finally, it is possible to envisage applications of option uncertainty to game theory, since game theory can be seen as an extension of individual choice theory. For instance, consider the case where in a two-player normal form game, the payoffs agents receive are contained within an interval. Then players would need to consider in their choice of strategy not only their own payoffs for any given strategy and the strategy of the opponent, but also how option certainty averse they are themselves, and how option uncertainty averse they think their opponent is.

State space uncertainty

In Chapter 2, we have identified state space uncertainty as the case where unforeseen contingencies may occur, such that the state space can no longer be assumed to be exhaustive. One natural application of state space uncertainty is contract theory, where unforeseen contingencies may affect the contracting parties. Kreps' (1992) model of unforeseen contingencies can be interpreted as a model of incomplete contracts. In particular, in Kreps' model the agent entertains a state space which contains some elements which the agent does not understand. The agent is then willing to contract only over those states he does understand, but not over the states he doesn't understand. A contract contingent on such a state space is then incomplete. Kreps gives a representation theorem for the agent's preferences over so-defined contracts.

Dekel, Lipman and Rustichini (1998) argue that Grossman and Hart (1986), Hart and Moore (1988) and Hart's (1995) model of incomplete contracts can be interpreted as a model of unforeseen contingencies in Kreps' spirit. In particular, Grossman, Hart and Moore argue that in many cases contracts are incomplete, since it is impossible to specify, at the time a contract is signed, all the terms and conditions for all possible contingencies. Hence, contracts will often be incomplete. Grossman, Hart and Moore show that when a contract is used to regulate trade between two parties who must each make relationship-specific investments, then first-best results will not generally obtain. In particular, the second-best outcome will then involve under-investment.

Kraus and Sagi (2006) apply the concept of unforeseen contingencies to asset pricing in financial markets. The authors interpret unforeseen contingencies as exogenous events which agents fail to foresee, and which affect the welfare of the agents - an unforeseen contingency will then result in a utility shock. Kraus and Sagi argue that an agent who makes decisions under unforeseen contingencies behaves just like an agent who experiences private taste shocks. This interpretation of unforeseen contingencies yields an interesting analogy with ethical uncertainty: both under ethical uncertainty and under unforeseen contingencies the agent's utility function may not be stable. In Kraus and Sagi's model, agents are assumed to be consciously unaware, in the sense that they know that unforeseen contingencies may impinge on their optimisation process. The authors show that securities can be traded only on demand- and price-contingent events. Furthermore, the market will be incomplete, and the agent's preferences will not satisfy expected utility theory.

The topic of unforeseen contingencies has not received much attention in the literature, presumably because it is difficult to find a compelling answer to the question how a rational agent should behave when unforeseen contingencies may affect their optimisation problem. However, particularly in the case of optimal contracting and financial markets the topic of unforeseen contingencies seems highly relevant, as in each case, unforeseen contingencies occur on a regular basis. It is therefore important to devise models which allow for unforeseen contingencies; this is a task further research may address.

6.6 Concluding remarks

This thesis has examined types of uncertainty which are incompatible with Savage's decision theory for small worlds. We have argued that these types of uncertainty are what we may call *large world* decision situations, namely cases where a small world representation facilitating the application of Savage's subjective expected utility theory is not feasible. These types of uncertainty require separate treatment to problems which are representable using small world decision matrices, since a reduction of the uncertainty would imply eliding details of the large world matrix which are relevant to the agent's decision problem. We have argued that in large worlds, the requirements of rationality placed on the agent's preferences differ from those applicable in small worlds.

Uncertainty is not a binary concept, but rather comes in shades of grey. An agent's uncertainty may not only vary in severity, but also in type. Chapter 2 of this thesis has provided a framework which allows the classification of different types of uncertainty, and which can be used to characterise the impact of the different types of uncertainty on agents' decisions with greater precision. This taxonomy opens up numerous avenues for future work; in particular, option uncertainty and ethical uncertainty are novel concepts which may have theoretical and empirical applications in a wide range of fields.

We hope to have shown in this thesis that minimal extensions to Savage's theory for small worlds can yield interesting new insights, for many of the most challenging decisions we have to make within the course of our lives are precisely those that are beset with uncertainty greater than that compatible with a Savage small world. It is in those cases that it is particularly important to approach decision problems in a rational manner. Whilst the contribution this thesis has made may be but a small step towards understanding decision making in large worlds, perhaps it can be seen as one piece of the puzzle that is decision making under uncertainty.

Bibliography

Al-Najjar, N. I. and Weinstein, J. (2009). The ambiguity aversion literature: A critical assessment. *Economics and Philosophy*, **25**, 249 – 284.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine. *Econometrica*, **21**, 503 – 546.

Anand, P. (1993). *Foundations of Rational Choice under Risk*. Oxford, UK: Clarendon Press.

Anand, P. (2002). Decision making when science is ambiguous. *Science*, **295**, 1839.

Anscombe, F. J. and Aumann, R. J. (1963). A definition of subjective probability. *Annals of Mathematical Statistics*, **34**, 199 – 205.

Bewley, T. F. (1986 / 2002). Knightian decision theory. Part I. *Decisions in Economics and Finance*, **25**, 79 – 110.

Binmore, K. (2007). Making decisions in large worlds. *ESRC Centre for Economic Learning and Social Evolution, University College of London.*

Binmore, K. (2009). *Rational Decisions*. Princeton, NJ: Princeton University Press.

Binmore, K. and Shaked, A. (2010). Experimental economics: Where next? *Journal of Economic Behavior and Organization*, **73**, 87 – 100.

Birnbaum, M. (2008). Postscript: Rejoinder to Brandstätter et al. *Psychological Review*, **115**, 260 – 262.

Blavatskyy, P. (2010). Back to the St. Petersburg paradox? *Management Science*, **51**, 677 – 678.

Bostrom, N. and Ord, T. (2006). The reversal test: Eliminating status quo bias in applied ethics. *Ethics*, **116**, 656 – 679.

Bradley, R. (forthcoming). *Decision Theory With a Human Face*. Cambridge, UK: Cambridge University Press.

Bradley, R. and Drechsler, M. (2012). Types of uncertainty. *Forthcoming*.

Brandstätter, E., Gigerenzer, G., and Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, **113**, 409 – 432.

Broome, J. (1991). *Weighing Goods: Equality, Uncertainty and Time*. Cambridge, MA: Basil Blackwell Ltd.

Camerer, C. F. and Weber, M. (1992). Recent developments in modelling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, **5**, 325 – 370.

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

de Finetti, B. (1937). La prévision: Ses lois logiques, ses sources sources subjectives. In H. Kyburg and H. Smokier, editors, *Annales de l'Institut Henri Poincaré*, pages 93 – 158. Studies in Subjective Probability. New York: Wiley.

de Finetti, B. (1977). Probabilities of probabilities: A real problem or a misunderstanding? In A. Aykac and C. Brumat, editors, *New Directions in the Application of Bayesian Methods*, pages 1 – 10. Amsterdam: North- Holland.

Deane, G. E. (1969). Cardiac activity during experimentally induced anxiety. *Psychophysiology*, **6**, 17 – 30.

Dekel, E., Lipman, B. L., and Rustichini, A. (1998). Recent developments in modelling unforeseen contingencies. *European Economic Review*, **42**, 523 – 542.

Dekel, E., Lipman, B. L., and Rustichini, A. (2001). A unique subjective state space for unforeseen contingencies. *Econometrica*, **69**, 891 – 934.

Dow, J. and da Costa Werlang, S. R. (1992). Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica*, **60**, 197 – 204.

Eliaz, K. and Ortoleva, P. (2011). A variation on Ellsberg. *Unpublished manuscript*.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, **75**, 643 – 669.

Epstein, L. G. (1999). A definition of uncertainty aversion. *Review of Economic Studies*, **66**, 579 – 608.

Epstein, L. G. and Miao, J. (2003). A two-person dynamic equilibrium under ambiguity. *Journal of Economic Dynamics and Control*, **27**, 1253 – 1288.

Epstein, L. G. and Seo, K. (2009). Subjective states: A more robust model. *Games and Economic Behavior*, **67**, 408 – 427.

Epstein, L. G. and Wang, T. (1994). Intertemporal asset pricing under Knightian uncertainty. *Econometrica*, **62**, 283 – 322.

Epstein, L. G. and Zhang, J. (2001). Subjective probabilities on subjectively unambiguous events. *Econometrica*, **69**, 265 – 306.

Fishburn, P. C. (1970). *Utility Theory for Decision Making*. New York: Wiley.

Fishburn, P. C. (1981). Subjective expected utility: A review of normative theories. *Theory and Decision*, **13**, 139 – 199.

Ghirardato, P. (2001). Coping with ignorance: Unforeseen contingencies and non-additive uncertainty. *Economic Theory*, **17**, 247 – 276.

Ghirardato, P. (2002). Revisiting Savage in a conditional world. *Economic Theory*, **20**, 83 – 92.

Ghirardato, P. (2004). Defining ambiguity and ambiguity attitude. In I. Gilboa, editor, *Uncertainty in Economic Theory*, pages 36 – 45. London, UK: Routledge.

Ghirardato, P. and Marinacci, M. (2002). Ambiguity made precise: A comparative foundation. *Journal of Economic Theory*, **102**, 251 – 289.

Ghirardato, P., Maccheroni, F., and Marinacci, M. (2004). Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory*, **118**, 133 – 173.

Gigerenzer, G. (2006). Out of the frying pan into the fire: Behavioral reactions to terrorist attacks. *Risk Analysis*, **26**, 347 – 351.

Gigerenzer, G. and Selten, R. (2001). Rethinking rationality. In G. Gigerenzer and R. Selten, editors, *Bounded Rationality: The Adaptive Toolbox*, pages 1 –12. Cambridge, MA: MIT Press.

Gilboa, I. (2004). In I. Gilboa, editor, *Uncertainty in Economic Theory*, pages 3 – 19. London, UK: Routledge.

Gilboa, I. and Marinacci, M. (2012). Ambiguity and the Bayesian paradigm. *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress of the Econometric Society*.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, **18**, 141 – 153.

Gilboa, I. and Schmeidler, D. (1995). Case-based decision theory. *Quarterly Journal of Economics*, **110**, 605 – 639.

Gilboa, I., Postlewaite, A., and Schmeidler, D. (2009). Is it always rational to satisfy Savage's axioms? *Economics and Philosophy*, **25**, 285 – 296.

Grossman, S. J. and Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, **94**, 691 – 719.

Harless, D. W. and Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, **62**, 1251 – 1289.

Hart, O. D. (1995). *Firms, Contracts, and Financial Structure*. Oxford, UK: Clarendon Press.

Hart, O. D. and Moore, J. (1988). Incomplete contracts and renegotiation. *Econometrica*, **56**, 755 – 786.

Heilmann, C. (2010). Rationality and time. *PhD Thesis, London School of Economics and Political Science*.

Hosni, H. (2012). Towards a Bayesian theory of second order uncertainty: Lessons from non-standard logics. *Unpublished manuscript*.

Hurwicz, L. (1951). Optimality criteria for decision making under ignorance. *Cowles Statistics Paper*, **370**.

Jaffray, J.-Y. (1989). Linear utility theory for belief functions. *Operations Research Letters*, **8**, 107 – 112.

Johnson, E. J. and Goldstein, D. G. (2003). Do defaults save lives? *Science*, **302**, 1338 – 1339.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, **58**, 697 – 720.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263 – 291.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives*, **5**, 193 – 206.

Katsikopoulos, K. V. (2011). Psychological heuristics for making inferences: Definition, performance, and the emerging theory and practice. *Decision Analysis*, **8**, 10 – 29.

Katsikopoulos, K. V. and Gigerenzer, G. (2008). One-reason decision-making: Modelling violations of expected utility theory. *Journal of Risk and Uncertainty*, **37**, 35 – 56.

Kempf, A. and Ruenzi, S. (2006). Status quo bias and the number of alternatives - An empirical illustration from the mutual fund industry. *Journal of Behavioral Finance*, **7**, 204 – 213.

Klibanoff, P. (1996). Uncertainty, decision and normal form games. *Mimeo, Northwestern University*.

Klibanoff, P., Marinacci, M., and Mukerji, S. (2005). A smooth model of decision making under ambiguity. *Econometrica*, **73**, 1849 – 1892.

Knetsch, J. L. and Sinden, J. (1984). Willingness to pay and compensation demanded: Experimental evidence of an unexpected disparity in measures of value. *Quarterly Journal of Economics*, **99**, 507 – 521.

Knight, F. H. (1921). *Risk, Uncertainty, and Profit*. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.

Krantz, D. M., Luce, R. D., Suppes, P., and Tversky, A. (1971). *The Foundations of Measurement*, volume 1. New York: Academic Press.

Kraus, A. and Sagi, J. S. (2006). Asset pricing with unforeseen contingencies. *Journal of Financial Economics*, **82**, 417 – 453.

Kreps, D. M. (1992). Static choice in the presence of unforeseen contingencies. In P. Dasgupta, D. Gale, O. Hart, and E. S. Maskin, editors, *Economic Analysis of Markets and Games*, pages 258 – 281. Cambridge, MA: MIT Press.

Leland, J. W. (1994). Generalized similarity judgments: An alternative explanation for choice anomalies. *Journal of Risk and Uncertainty*, **9**, 151 – 172.

Leland, J. W. (2002). Similarity judgments and anomalies in intertemporal choice. *Economic Inquiry*, **40**, 574 – 581.

Leland, J. W. (2010). The hunt for a descriptive theory of choice under risk - A view from the road not taken. *The Journal of Socio-Economics*, **39**, 568 – 577.

Lewis, D. (1981). Causal decision theory. *Australasian Journal of Philosophy*, **59**, 5 – 30.

Lewis, D. (1986). A subjectivist's guide to objective chance. In *Philosophical Papers, Vol. 2*. Oxford, UK: Oxford University Press.

Lo, K. C. (1996). Equilibrium in beliefs under uncertainty. *Journal of Economic Theory*, **71**, 443 – 484.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., and Welch, N. (2001). Risk as feeling. *Psychological Bulletin*, **127**, 267 – 286.

Loomes, G. and Sugden, R. (1987). Testing for regret and disappointment in choice under uncertainty. *The Economic Journal*, **97**, 118 – 129.

Loomes, G., Orr, S., and Sugden, R. (2009). Taste uncertainty and status quo effects in consumer choice. *Journal of Risk and Uncertainty*, **39**, 113 – 135.

Lopes, L. L. and Oden, G. C. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and SP/A theory. *Journal of Mathematical Psychology*, **43**, 286 – 313.

Luce, R. D. (1956). Semi-Orders and a Theory of Utility Discrimination. *Econometrica*, **24**, 178 – 191.

Luce, R. D. (1978). Lexicographic trade-off structures. *Theory and Decision*, **9**, 187 – 193.

Luce, R. D. and Raiffa, H. (1957 / 1989). *Games and Decisions. Introduction and Critical Survey*. New York: Dover Publications.

MacCrimmon, K. R. (1968). Descriptive and normative implications of the decision-theory postulate. In K. H. Borch and J. Mossin, editors, *Risk and uncertainty*. New York: St. Martin's Press.

Machina, M. J. (1982). "Expected utility" analysis without the independence axiom. *Econometrica*, **50**, 277 – 323.

Machina, M. J. and Schmeidler, D. (1992). A more robust definition of subjective probability. *Econometrica*, **60**, 745 – 780.

Marinacci, M. (2000). Ambiguous games. *Games and Economic Behavior*, **31**, 191 – 219.

Masatlioglu, Y. and Ok, E. A. (2005). Rational choice with status quo bias. *Journal of Economic Theory*, **121**, 1 – 29.

Mellers, B., Weiss, R., and Birnbaum, M. H. (1992). Violations of dominance in pricing judgments. *Journal of Risk and Uncertainty*, **5**, 73 – 90.

Modica, S., Rustichini, A., and Tallon, J.-M. (1998). Unawareness and bankruptcy: A general equilibrium model. *Economic Theory*, **12**, 259 – 292.

Mukerji, S. (1997). Understanding the non-additive probability decision model. *Economic Theory*, **9**, 23 – 46.

Mukerji, S. and Tallon, J.-M. (2004). An overview of economic applications of David Schmeidler's models of decision making under uncertainty. In I. Gilboa, editor, *Uncertainty in Economic Theory*, pages 283 – 302. London, UK: Routledge.

Nehring, K. (2009). Coping rationally with ambiguity: Robustness versus ambiguity aversion. *Economics and Philosophy*, **25**, 303 – 334.

Neilson, W. and Stowe, J. (2002). A further examination of cumulative prospect theory parameterizations. *Journal of Risk and Uncertainty*, **24**, 31 – 46.

Ortoleva, P. (2010). Status quo bias, multiple priors and uncertainty aversion. *Games and Economic Behavior*, **69**, 411 – 424.

Patel, J., Zeckhauser, R., and Hendricks, D. (1991). The rationality struggle: Illustrations from financial markets. *American Economic Review*, **81**, 232 – 236.

Ramsey, F. P. (1926). Truth and probability. In *Foundations of Mathematics and Other Logical Essays*, pages 156 – 198. London, UK: Routledge.

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **34**, 1446 – 1465.

Ritov, I. and Baron, J. (1992). Status quo and omission biases. *Journal of Risk and Uncertainty*, **5**, 49 – 61.

Rubaltelli, E., Rubichi, S., Savadori, L., Tedeschi, M., and Ferretti, R. (2005). Numerical information format and investment decisions: Implications for the disposition effect and the status quo bias. *Journal of Behavioral Finance*, **6**, 19–26.

Rubinstein, A. (1988). Similarity and decision-making under risk (is there a utility theory resolution to the Allais paradox?). *Journal of Economic Theory*, **46**, 145 – 153.

Samuleson, P. A. (1952). Probability, utility and the independence axiom. *Econometrica*, **20**, 670 – 678.

Samuleson, W. and Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, **1**, 7 –49.

Sargent, T. J. (1993). *Bounded rationality in macroeconomics*. Oxford, UK: Oxford University Press.

Sarin, R. and Wakker, P. P. (2004). A simple axiomatization of nonadditive expected utility. In I. Gilboa, editor, *Uncertainty in Economic Theory*, pages 136 – 154. London, UK: Routledge.

Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.

Schkade, D. A. and Johnson, E. J. (1989). Cognitive processes in preference reversals. *Organizational Behavior and Human Decision Processes*, **44**, 203 – 231.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, **57**, 571 – 587.

Selten, R. (2001). What is bounded rationality? In G. Gigerenzer and R. Selten, editors, *Bounded Rationality: The Adaptive Toolbox*, pages 13 – 27. Cambridge, MA: MIT Press.

Shafer, G. (1986). Savage revisited. *Statistical Science*, **1**, 463 – 501.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, **69**, 99 – 118.

Simon, H. A. (1991). *Models of My Life*. New York: Basic Books.

Skiadas, C. (1997). Subjective probability under additive aggregation of conditional preferences. *Journal of Economic Theory*, **76**, 242 – 271.

Slovic, P. and Tversky, A. (1974). Who accepts Savage's axioms? *Behavioral Science*, **14**, 368 – 373.

Stalnaker, R. (1981). A theory of conditionals. In W. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs: Conditionals, Belief, Decision, Chance, and Time*, pages 41 – 56. Dordrecht, NL: Reidel.

Steele, K. (2006). What can we rationally value? *Mimeo, University of Sydney*.

Tetlock, P. E. and Boettger, R. (1994). Accountability amplifies status quo bias when change creates victims. *Journal of Behavioral Decision Making*, **7**, 1 –23.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization*, **1**, 39 – 60.

Tversky, A. and Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, **211**, 453 – 458.

Tversky, A. and Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics*, **106**, 1039 – 1061.

von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Wakker, P. P. and Tversky, A. (1993). An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty*, **7**, 147 – 176.

Walker, O. and Dietz, S. (2011). A representation result for choice under conscious unawareness. *Working Paper, Grantham Research Institute, London, UK.*

Wu, G. and Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, **42**(12), 1676 – 1690.

Zhang, J. (2002). Subjective ambiguity, expected utility and Choquet expected utility. *Economic Theory*, **20**, 159 – 181.