PRECAUTIONARY DECISION-MAKING

AN EXAMINATION OF BAYESIAN DECISION NORMS IN THE DYNAMIC CHOICE CONTEXT

Katie Steele

A thesis submitted for the degree of Doctor of Philosophy

Department of Philosophy, University of Sydney, 2007

PREFACE

Declaration of originality. I hereby certify that the content of this thesis, unless otherwise indicated, is my own original work, and has not been previously submitted for another degree.

Published papers. This thesis contains material from papers that I have authored which have been published elsewhere. I thank the respective publishers for permitting me to use this material in my thesis. In particular, the introduction contains content from "The precautionary principle: A new approach to public decision-making?" (Steele, 2006), which is published in the journal *Law, Probability and Risk*. Additionally, large portions of Chapter 5 are taken from my paper "Distinguishing indeterminate belief from 'risk-averse' preferences" (Steele, 2007), which is published in the journal *Synthese* (in the special section *Knowledge, Rationality and Action*).

ABSTRACT

In this thesis, I investigate whether the idea of "precautionary" reasoning presents a challenge to the dominant model of rational choice—subjective expected utility (SEU) theory. The major themes of this work, explored in Part I and Part II respectively, reflect the two somewhat distinct ways in which "precaution" can be interpreted. The first is the issue of what constitutes rational expectations about the future, in particular, one's future beliefs and desires, and how such expectations should impact on decision-making (Part I). This requires an examination of Bayesian norms for belief and preference change. The other aspect of "precaution" concerns agents' attitudes towards risk. I consider whether prominent examples of "risk-sensitive" choice behaviour should be considered desirable, or even rational. The question is whether such "risk sensitivity" is compatible with SEU theory, and if not, whether there are plausible alternative approaches to the representation and handling of uncertainty (Part II). I demonstrate that the sequential-choice framework is critical to examining both these major issues. Not only does the sequential model serve to expose the temporal structure of a decision problem, it also provides a rich context for justifying/assessing the key Bayesian epistemic and pragmatic decision norms.

ACKNOWLEDGEMENTS

I acknowledge the ARC Centre for Legume Research with which I was formally affiliated while at UQ, and which partially supported my PhD research.

Otherwise, it is a pleasure to have this space to thank a number of people who have helped me in this project. First, I owe considerable thanks to my principal supervisor, Mark Colyvan, who has guided me throughout this whole PhD process, with wonderful enthusiasm for discussing philosophy and much thought for my professional development more generally. It is perhaps Mark's particular combination of ambition, egalitarianism and story-telling ability that makes him so nice to work with! He is also a fantastic editor, and has continually helped me improve my writing.

I am also very grateful for the support of my associate supervisors—Helen Regan and William Grey. Particularly in the early stages, William was a great source of calming advice, and has given thoughtful feedback on my work. Helen Regan's generosity as a mentor even extended to putting me up for several weeks in her home in San Diego! This was a wonderful period for exploring new academic (and coastal) territory, and for having lively chats with Helen about environmental decision-related problems.

There are a few other people that I would like to single out: Teddy Seidenfeld has read a number of my papers, and always gives very incisive yet supportive feedback. Alan Hájek is another person I am fortunate to have had many exchanges with; he makes probability/decision theory seem like the activity of poets, and just plain fun. I also particularly thank my fellow PhD traveller Adam La Caze, from whom I have learnt much, and who is always a sympathetic audience for my half-baked ideas.

A number of others have been willing collaborators, reading-group buddies, fellow coffee-drinkers, or have generally buoyed my enthusiasm for research somewhere

along the way. I particularly want to mention: Horacio Arlo-Costa, Carol Booth, Hugh Breakey, Rachael Briggs, Mark Burgman, Lucy Carter, Damian Cox, Phil Dowe, Kenny Easwaran, Fiona Filardo, David Gray, Hilary Greaves, Peter Gresshoff, Paul Griffiths, Jason Grossman, Jim Hawthorne, Simon Hutteger, Marguerite La Caze, Julian Lamont, Joan Leach, Stefan Linquist, Christian List, Martyn Lloyd, Alex London, Aidan Lyon, Mark Machina, Gary Malinas, Hatha McDivitt, Fabien Medvecky, Kristie Miller, Alex Moffett, Emily Nicholson, Brian Skyrms, Carl Wagner and John Wilkins.

In addition, my mother, my father and my friend Cristy proofread parts of my near-completed thesis. Of course, I could not do without the friends and family who always support me, whether it was to be this project or another. And finally, I thank Colin.

TABLE OF CONTENTS

II RISK-SENSITIVITY

INTRODUCTION

The art of choice. We are continuously confronted with decisions. Consider a situation where I want to organise dinner with a friend. I need to decide whether I should book ahead or not; I hesitate, because there is something to be said for having an open plan, but on the other hand, it may be more relaxing to know that a table is ready and waiting. Upon arrival at the restaurant, we set to navigating the menu, as well as scanning the bottles of wine. After eating, there are generally drinks and desserts available; I might reason that herbal tea is the more sensible choice at the close of the night, but then again, coffee and cake taste better. Already this is starting to sound a bit taxing, and yet it probably represents only a small portion of a day's decision-making work!

Fortunately, the majority of decisions that we have to make are similar to those above in that the consequences, whatever one finally decides upon, are not very significant, or else not significantly different. Even at a domestic level, there are occasions, however, when we need to make decisions on relatively important issues. These seem to deserve more careful thought. Let's say I am planning an around-the-world trip, and, in my case, it is no small budgeting matter. I need to decide which route to take, given that the tickets are variously priced and allow me to visit different places. Surely in this instance it is worth comparing the different options conscientiously. Take another example: Mary may be deciding whether to keep running her old car, or else buy a new "green" car while the old one is still in reasonable condition. It is not immediately obvious which of these options will be better in terms of personal cost, or even in terms of environmental cost (the new car will be cleaner to run, but there is the production process to consider as well). Since Mary cares a lot about these issues, she is wise to spend some time reflecting on the decision problem.

Determining how much time and energy to spend gathering more relevant information and fine-tuning one's decision-making in any given scenario is a difficult decision problem in itself. Clearly we do not want to spend half a day agonizing over which restaurant to meet for dinner, because that would, in a sense, defeat the purpose of a fun night out. In many cases, it is surely better to make a quick choice rather than reflect too long on the range of alternative actions and their likely consequences; there is much to be said for well-developed habits, or else for acting spontaneously. On the other hand, in situations where there is ample time and a lot at stake, it is worthwhile being conscientious when choosing between courses of action. In any case, different time and resource constraints do not substantially change the decision process. Whether it is a trivial or a weighty decision matter, the basic problem is the same: in general terms, the person (or agent) is entertaining a number of courses of action, and wants to determine which is the best one to perform, given their respective consequences. Note that one of these proposed actions might just be to defer the decision until more information is attained.

Even when the options available have determinate outcomes, i.e. the consequences of each action are assured, decision-making is not always straightforward. It may still be difficult to specify the consequences, and to rank them. The airline ticket decision outlined above makes for a good example. There will be a fact of the matter about the final price and the routing rules for each of the tickets on offer, yet it takes some calculations to match the ticket to my intended journey. And even when the details of each ticket are clear, it can be difficult to decide which is best; for instance, I might be undecided on whether it is better to pay $200 extra for a stopover in San Francisco. In other circumstances, when the agent in question is uncertain about the consequences of an option, there is a further element to the decision-making difficulties. When purchasing my airline ticket, for example, I may be given the option of taking out expensive travel insurance. This requires considering both the import and the likelihood of the various possibilities regarding the safety of my property and my personal health during the trip. Or consider Mary's decision about whether to buy the new "green" car. Perhaps her decision is partly affected by whether she thinks oil prices will rise significantly over the next 5 years, in which case it will be very

expensive to keep running the old car. So Mary's final choice will depend on how she balances the possible consequences associated with each option.

Theory of decision. We would expect a theory of decision, then, to provide us with guidance in any kind of decision scenario, whether the consequences of actions are known for sure, or whether they are uncertain. Alternatively, we would want to know whether the choices that an agent makes are justified, given their opinions about the likelihood and value of the consequences associated with each option. Note that I am using normative language here, and indeed, this thesis is concerned with how one *should* choose—the normative dimension of choice—rather than how us ordinary sloppy agents do in fact make decisions in everyday life. The distinction is not entirely clear-cut. Normative decision models are employed in the empirical sciences (e.g. in cognitive science or economics) to approximate and make predictions about how human agents actually reason,[1] and empirical findings do impact on the normative study of choice. (I briefly discuss the significance of empirical findings to normative decision models in Chapter 4 of this thesis.)

What we seek then are general or universal principles for good choice. We might reasonably expect any such principles to be quite straightforward, even if they are difficult to put into practice. When the consequences of actions are known for certain, it would seem very obvious what the agent should do—they ought to rank the options in terms of the goodness of their consequences, and simply choose the best! Moreover, many have an intuitive idea about what is the appropriate way to choose under conditions of uncertainty. Consider a lottery scenario. Let's say each ticket costs $1 and literally has a "one in a million" chance of winning, the prize money being $500,000. A wise agent who is thinking purely in monetary terms might judge that this lottery is not worth participating in. Why? Well, it seems reasonable to appeal to the *expected value* of the lottery, which is just the sum, over all outcomes, of

[1] Machina (1989) gives a good account of why normative decision models are important in economic modelling. The general idea is that many agents acting together tend to approximate ideal agents. Moreover, any irrational agents in an economic model would be immediately exploited by others, and the market does not exhibit these sorts of obvious and large-scale "money sinks".

the probability of each outcome multiplied by its value. There are two outcomes in this lottery example—the participant either takes home the prize (minus the price of the ticket), or else they simply lose the price of the ticket. The lottery thus has an expected value of $0.000001 \times \$499{,}999 + 0.999999 \times -\$1 = -\$0.50$. In other words, the lottery yields an expected loss, and so it is surely not worth playing. In fact, the standard decision model is not much more complicated than what I have just outlined here. Decisions depend on the interaction of probability (or belief) and value (or utility) of potential outcomes. Where consequences are uncertain, we are advised to rank actions on the basis of expected value considerations. Where the consequences of an act are certain, we simply have the special case in which the relevant probability is equal to 1. If we want to assess challenges to the standard model (as I will do in this thesis) then we need to depict the standard model in more precise terms and carefully consider its justification, but that does not change the fact that the core expected value principle for decision-making under uncertainty is very straightforward, and intuitive to many.

Unfortunately, if the consequences of actions are uncertain, a good decision need not always turn out well for the agent in question. Sometimes the least preferred possible outcome for an action eventuates, even if it was estimated to have very low probability. By the same token, bad decisions can sometimes be very fortuitous for the agent. Consider the lottery outlined above. Someone has to win the prize, and that particular person may well be extremely pleased with him/herself for buying the $1 ticket. But this does not change the fact that the winner, just like all the other lottery participants, made a bad decision (if we are speaking purely in monetary terms). If presented with the opportunity to buy another lottery ticket, our former winner would be wise to turn it down. The point here is that, while good decision-making and good outcomes are related, there is an important distinction between the two. And we are wise to aspire to good decision-making, rather than hope for good outcomes.

The call for precautionary decision-making. The question I investigate in this thesis is whether the idea of "precautionary" reasoning presents a challenge to the standard expected value decision model. I will describe the standard decision model in more

detail in the next section. Let us first consider what precaution has come to stand for in the context of practical decision-making. Recent acknowledgement of failures in public policy and management, particularly as concerns the natural environment, has served to focus attention on the handling of "risk" in decision-making. Given our apparent blindness towards looming environmental and social crises, the so-called "Precautionary Principle" has risen in prominence, both in the context of international agreements and conventions, and also at more local levels. It is referred to explicitly in a number of UN documents, such as the 1992 Rio Declaration, and bodies such as UNESCO and the EU have released position statements on the principle.[2] At its core, the Precautionary Principle holds that public policy should include measures to avoid or diminish morally unacceptable harms to human health and environment, even if the harms in question are merely possible, as opposed to being certain outcomes of an action (see Steele forthcoming). This is clearly a rather broad recommendation; it allows much scope for interpreting just what our social-environmental obligations are supposed to be, and importantly, what aspects of existing decision procedures should be reformed.[3] Indeed, the various formulations of the Precautionary Principle in both academic and policy papers interpret its requirements differently, and emphasise different faults in existing decision procedures.

In the academic Precautionary Principle literature, in particular, there is a large focus on the representation and handling of uncertainty in decision-making.[4] A question that is ever present in this work, at least implicitly, is whether or not the standard economic decision model provides an adequate treatment of uncertainty.[5] Whether

---

[2] The COMEST (2005) report on the Precautionary Principle outlines its various recommendations for public decision-making. The report also catalogs other references to the Precautionary Principle in international policy contexts (including the 1992 Rio Declaration and a 2000 statement from the EU).

[3] This ambiguity is not necessarily a bad thing, because if the Precautionary Principle is to serve as a higher-order legal principle, then it should be broadly applicable, and a consequence of this is that there should be some flexibility with respect to its application in particular cases.

[4] For discussions of Precautionary Principle recommendations regarding risk/uncertainty, see, for instance, Cranor (2001), Keeney and von Winterfeldt (2001), Manson (2002), Sandin *et al*. (2002) and Resnik (2003).

[5] Again, I am referring to the standard decision model, or subjective expected utility (SEU) theory, which I will outline in the next section.

these are intentional challenges to standard decision theory or not, a number of variants to the standard decision model have been put forward. The general concerns seem to be that uncertainty (and potentially different types of uncertainty) be properly represented in a decision model, and given sufficient weight in decision calculations. One of the least credible suggestions as far as a modified decision rule goes is something that has been disparagingly referred to by commentators as the "catastrophe principle". This principle holds that any action that has some chance of leading to a catastrophic, morally reprehensible outcome should be forbidden. The obvious problem here is that nearly every action has *some* chance of disastrous outcomes. (For criticism of this principle, see, for instance, Manson 2002.) Others, e.g. Resnik (2003), appeal to the use of special decision rules for cases in which we are extremely ignorant about the possible outcomes of an action, i.e. cases in which we are unsure about what is the chance of harm. Resnik appeals to a modified "maxi-min" rule; "maxi-min" recommends ranking actions on the basis of their worst-case possible consequence/outcome, as opposed to the expected value of their consequences. Yet others have simply stipulated what preventative measures should be applied to particular sorts of actions when the potential harms involved are of a specific type and have a certain level of predictability (see Manson 2002).

The standard decision model. While the idea of "precaution" has gained increasing currency in public decision-making contexts, some of the lessons drawn from past mistakes have been misguided. Many formulations of the Precautionary Principle (including some of the above-mentioned suggestions) would benefit from more explicit reference to formal decision theory. There is a tendency to overlook just how powerful and flexible the standard decision model is. Before proceeding much further, it will be useful to introduce some terminology, and present in more detail what I have been calling the "standard" decision framework; it is otherwise referred to as subjective expected utility (SEU) theory.[6] I have already given a rough outline of the

[6] This terminology—subjective expected utility (SEU) theory—is used by many, including Fishburn (1981), Skyrms (1986) and Joyce (1999). Joyce (1999, p. 9) notes that SEU theory is the most widely defended normative decision model. The reason the theory is referred to as *subjective* expected utility theory should become clear in what follows. In brief, the probability terms in the model represent an agent's subjective attitudes, rather than some objective feature of the world. Note that Savage (1954) uses similar terminology—he refers to

standard rule for evaluating actions that have either certain or uncertain consequences. Recall the lottery example, and the general principle that we should look to the expected value of the set of consequences associated with each option. We estimated the value of the lottery to be $0.000001 \times \$499{,}999 + 0.999999 \times -\$1 = -\$0.50$. SEU theory retains this expected value rationale, but the theory is presented in more general terms; it incorporates standards for rational belief, and does not require value to be measured in any particular currency, whether money or some other good.

Let me give an outline of SEU theory. The situation is one in which an agent must choose between a set of predefined acts $A = \{A_1, A_2, \ldots A_n\}$. For instance, we might be modelling Mary's car predicament. Let's say A_1 represents the act of buying the new "green" car, while A_2 stands for keeping on with the old car. Each act is associated with a set of possible outcomes $O = \{O_{i1}, O_{i2}, \ldots O_{im}\}$. If Mary keeps the old car, one possible scenario is that oil prices will soar in the near future, with the consequence/outcome that Mary will be running an extremely expensive car that is also very polluting. Indeed, it is customary to associate the particular outcomes for each act with possible states of the world, or possible ways that the world could turn out. In Mary's case, an obvious choice of states is the partition $\{S_1, S_2\}$ where S_1 is the state in which oil prices soar, and S_2 is the state in which oil prices decrease or stabilise. States should be mutually exclusive, and the set of states should exhaust the ways the agent believes the world might turn out. Typically, a decision problem is represented in the form of a matrix or table (as per Figure I-1 below), where the acts are the rows, the states are the columns, and each act yields a particular outcome for each possible state. (This is referred to as the normal-form or static representation of a decision problem.)

---

his expected utility theory as a *personalist* theory of decision.

Figure I-1

	S_1	S_2	…		S_j	…		S_m
A_1	O_{11}	O_{12}	…		O_{1j}	…		O_{1m}
A_2	O_{21}	O_{22}	…		O_{2j}	…		O_{2m}
…								
A_n	O_{n1}	O_{n2}	…		O_{nj}	…		O_{nm}

The agent is considered to have a probabilistic belief function over the set of states. In other words, there should be a probability function $\Pr(S_i)$ representing the agent's uncertainties about the states—how likely the agent thinks it is that each state will turn out to be the true state of the world.[7] Since the states are mutually exclusive and exhaustive, it should be the case that

$$\sum_{i=1}^{m} \Pr(S_i) = 1.$$

In addition, the agent is expected to have a value/utility function $\mathrm{U}(O_{ki})$ over the set of possible outcomes; value is not measured in terms of money or any other good, but is rather given an abstract numerical measure that is referred to as *utility*. The utility function should satisfy the von Neumann-Morgenstern axioms, which I will outline shortly. SEU theory holds that an act A_k is (strictly) preferred, or ranked above another act A_j (I denote strict preference by the symbol "$\succ$"), if and only if the expected utility (EU) of A_k is greater than the expected utility of A_j. In formal terms:

$$A_k \succ A_j \;\Leftrightarrow\; \mathrm{EU}(A_k) > \mathrm{EU}(A_j) \;\Leftrightarrow\; \sum_{i=1}^{m} \mathrm{U}(O_{ki}) \times \Pr(S_i) > \sum_{i=1}^{m} \mathrm{U}(O_{ji}) \times \Pr(S_i)$$

In addition, an act A_k is considered indifferent to another act A_j (I denote indifference by the symbol "$\approx$"), if and only if the expected utility (EU) of A_k is

[7] Some versions of SEU theory, e.g. Savage's (1954) theory, assume that the likelihood of the states is independent of the acts. In other words, the states should be specified in such a way that performing any act under consideration has no affect on the probability of the states. Other versions of SEU theory, e.g. Jeffrey's (1983) theory, do not include this constraint; in such case, each act may be associated with a different probability function over the states.

equivalent to the expected utility of A_j:

$$A_k \approx A_j \;\Leftrightarrow\; \mathrm{EU}(A_k) = \mathrm{EU}(A_j) \;\Leftrightarrow\; \sum_{i=1}^{m} U(O_{ki}) \times \Pr(S_i) = \sum_{i=1}^{m} \mathrm{U}(O_{ji}) \times \Pr(S_i)$$

It is worth pausing a moment to consider what acts, states and outcomes actually amount to. It seems clear enough that an act is something that we may choose to perform in order to realise certain goods, whereas states and outcomes are things that we have basic belief or desire attitudes towards. More precisely, and in keeping with the above account, we can identify states as objects of "credal" probability judgment (or belief), outcomes as objects of intrinsic or basic desire, and acts as objects of instrumental desire.[8] We could well consider these as three distinct sorts of objects, but Jeffrey (1965) has provided a very useful generalisation of the decision framework: he suggests that acts, states and outcomes should all be considered propositions that effectively describe possible "prospects", or sets of "possible worlds". While the details of Jeffrey's decision theory may be controversial,[9] the simplicity that comes with depicting the terms in a decision model as propositions is hard to dismiss, and has been well accepted by decision theorists. Joyce (1999, pp. 67–69) gives the following more formal account of such a general decision framework:

> We can start by thinking of the partitions ***A*** (the set of acts), ***S*** (the set of states), and ***O*** (the set of outcomes) as embedded within a larger set of propositions Ω that has the structure of a Boolean σ-algebra. Ω can be defined as the smallest collection of propositions that includes the partitions ***A***, ***S***, and ***O***; is closed under negation, so that $\sim X \in \Omega$ whenever $X \in \Omega$; and is closed under countable disjunction, so that $(X_1 \vee X_2 \vee X_3 \vee \ldots) \in \Omega$ whenever $X_1, X_2, X_3, \ldots \in \Omega$.

---

[8] Both Joyce (1999, p. 68) and Levi (1997, p. 74) give a similar account to the above of the roles that acts, states and outcomes play in a decision model (or in Savage's 1954 decision model, at least).

[9] For starters, Jeffrey's theory in fact denies some of the aforementioned distinctions, and permits acts, states and outcomes to be objects of both belief and desire judgments. And Jeffrey's theory does not attend to the causal structure of a decision problem.

Throughout this thesis (unless otherwise indicated), I will assume this general decision framework. Moreover, at times I explicitly refer to "Jeffrey-style" possible worlds. These represent particular ways that the agent thinks the world could turn out, or in other words, they are potential world histories, described in all their relevant detail.

The ordinal preference ranking of acts that results from evaluating them according to their expected utilities has a number of important properties. Indeed it is these properties that give weight to the claim that SEU theory is the appropriate model of rational choice. In fact, the story is generally told the other way around: we first consider what are logical constraints on ordinal preference, and then we consider how to numerically represent an agent whose preferences satisfy these constraints. Theorems to this effect—those showing that any agent with rational preferences can be represented as subscribing to a particular decision calculus—are referred to as *representation theorems*. In Appendix 1, I outline the main postulates that comprise Savage's (1954) representation theorem for SEU theory. As I alluded to above, Jeffrey (1965, 1983) offers a slightly different version of the theorem that is, in a sense, more general, because it treats all terms in the decision model—acts, states and outcomes—as propositions.[10] While the differences between these two theorems (as well as other variants of SEU theory)[11] are significant, they are not the focus of this thesis. Indeed, I tend to speak of the SEU representation theorems in general, and I make use of both Savage's and Jeffrey's core frameworks. I single out Savage's theorem in Appendix 1 simply because it is slightly easier to outline, and because it is perhaps more widely known. Here I will describe in informal terms an even more straightforward (and somewhat abbreviated) version of the expected utility theorem—von Neumann and Morgenstern's (1944) theorem. My presentation of the theorem closely follows that of Resnik (1987, p. 88–92). This will give an idea of the

[10] Jeffrey's (1983) SEU representation theorem is based on some mathematical work of Ethan Bolker.

[11] There are other variants of SEU theory. For instance, Joyce (1999) has developed an alternative SEU theorem that could be considered a modification of Jeffrey's, and that is intended to take into account the causal structure of a decision problem.

axiomatic relationship between preference axioms and a numerical utility function, but it is not as comprehensive as the other representation theorems because it does not constrain an agent's beliefs over the possible states of the world; von Neumann and Morgenstern's theorem does not tell us how to evaluate acts with (subjectively) uncertain outcomes, and these are the sort of acts that feature in regular decision-making scenarios.

Von Neumann and Morgenstern's expected utility theorem

Suppose the expression $L(a, x, y)$ stands for the lottery that gives you a chance equal to a at the prize x and a chance equal to $1 - a$ at the prize y. (The prizes x and y may be simple outcomes, or they may themselves be lotteries involving simple outcomes.)

The expression $x \succ y$ means that the agent (strictly) prefers x to y, and the expression $x \approx y$ means that the agent is indifferent between x and y.

Axiom 1 (ordering):

[Anti-symmetry] O1: If $x \succ y$, then not $y \succ x$.

O2: If $x \succ y$, then not $x \approx y$.

O3: If $x \approx y$, then not $x \succ y$ and also not $y \succ x$.

[Completeness] O4: $x \succ y$ or $y \succ x$ or $x \approx y$, for any relevant outcomes x and y.

[Transitivity] O5: If $x \succ y$ and $y \succ z$, then $x \succ z$.

O6: If $x \succ y$ and $x \approx z$, then $z \succ y$.

O7: If $x \succ y$ and $y \approx z$, then $x \succ z$.

O8: If $x \approx y$ and $y \approx z$, then $x \approx z$.

Axiom 2 (continuity condition):

Given three alternatives x, y, z with y ranked between x and z, agents must be indifferent between y and some lottery yielding x and z as prizes.

> For any lotteries x, y, and z, if $x \succ y$ and $y \succ z$, then there is some real number a such that $0 \leq a \leq 1$ and $y \approx L(a, x, z)$.

Axiom 3 (reduction-of-compound-lotteries condition):

Agents must evaluate compound lotteries in agreement with the probability calculus.

> For any lotteries x and y and any numbers a, b, c, d (again between 0 and 1, inclusively),
>
> if $d = ab + (1 - a)c$, then $L(a, L(b, x, y), L(c, x, y)) \approx L(d, x, y)$.

Axiom 4 (better-prizes condition):

Given two otherwise identical lotteries, agents will prefer the one giving the better "first" prize—if everything else is equal.

> For any lotteries x, y and z, and any number a such that $0 \leq a \leq 1$, $x \succ y$ if and only if $L(a, z, x) \succ L(a, z, y)$ and $L(a, x, z) \succ L(a, y, z)$.

Axiom 5 (better-chances condition):

Given two otherwise identical lotteries, agents will prefer the one giving the best chance at the "first" prize.

> For any lotteries x and y and any numbers a and b (both between 0 and 1, inclusively), if $x \succ y$, then $a > b$ just in case $L(a, x, y) \succ L(b, x, y)$.

Expected Utility Theorem:

If an agent's ordinal preferences over prizes (including simple outcomes and all lotteries compounded from these simple outcomes) obey the above 5 axioms, then we can construct an interval utility function U with the following properties:

1) $U(x) > U(y)$ if and only if $x \succ y$
2) $U(x) = U(y)$ if and only if $x \approx y$
3) $U[L(a, x, y)] = aU(x) + (1 - a)U(y)$
4) Any U' also satisfying (1)–(3) is a positive linear transformation of U.

Formulating the decision problem. While I do not deny that there are legitimate challenges to the SEU decision framework (indeed, such challenges will be the main preoccupation of this thesis), I think it is important, from the outset, not to lose sight of the power of the standard model. Indeed, I argue (Steele forthcoming) that the attention the Precautionary Principle has directed towards reforming the handling of uncertainty in decision-making is somewhat out of line with the sort of public policy problems that the principle is concerned with. As discussed above, the SEU model is very minimal; it tells the decision maker how to choose amongst a set of acts, given their particular belief and value functions over the possible outcomes. In other words, the model tells an agent how to act, or alternatively, whether they have well-structured preferences (i.e. preferences that obey the axioms outlined above), whatever idiosyncratic beliefs and values they bring to the table. Of course, we might think some sets of beliefs and values are better justified than others, but that is a whole other matter. It is not the fault of SEU theory if an agent attributes higher value to an ugly and inequitable world, say, than a world with plentiful biodiversity and decent living standards for all.

Clearly, then, a very substantial part of any actual decision-making process is the formulation of the decision problem—identifying the available actions and determining appropriate beliefs and values. It is reasonable to think that the problem set-up should be the first thing to come under investigation when public or personal decisions appear to go wrong.[12] Indeed, an obvious starting point is to consider whether outcomes are attributed appropriate values. For instance, before we start worrying about the uncertainty associated with whether a particular land management option will lead to a species extinction, it is important that the persistence of the species has some value in the decision model, otherwise this biodiversity measure cannot have any effect on the final choice. The basic point here is that standard decision theory does not overlook uncertainties, provided the things we are uncertain about are given significance, or have an appropriate influence on the value of

[12] Recall my earlier point that it is important to be careful in making accusations of faulty decision-making. An agent may simply be unlucky when their chosen action results in a bad outcome. The action may have had maximal expected utility, but sometimes the low-probability, undesirable outcomes eventuate.

outcomes. Furthermore, SEU theory says nothing to the effect that the value of outcomes must be measured entirely in monetary or market terms, even if the outcomes of public decisions are regularly construed in this way.[13]

It is worth mentioning that the Precautionary Principle literature does include a number of useful suggestions about the formulation of a decision problem, and the sort of consequences that should be given weight. There is a recognised link between the principle and the ethical ideals associated with "sustainable development" (see, for instance, Stein 2000, O'Riordan and Jordan 1995, and the 2005 COMEST report). A central aspect of this ethical outlook is concern for the plight of future generations, which means suitably taking into account the long-term as well as the short-term consequences of actions.[14] More concrete proposals to this effect include the notion of "polluter pays"—the idea that producers be legally responsible for the environmental and social effects of their activities. In general, the Precautionary Principle has encouraged discussion about what are appropriate legal burdens of proof for cases where public goods are at risk.[15] For restricted activities, the task is to determine the amount of evidence required (or the degree of certainty that must be established) that the consequences of an action are sufficiently safe, for it to be permissible. In other contexts, the question is what amount of evidence that an activity has harmful effects provides sufficient grounds for particular legal remedies. Such questions about evidential standards can be investigated through the standard SEU model. We want to select the appropriate level-of-evidence, or the evidential threshold, in such a way that the risk of permitting an unsafe action balances appropriately with the risk of

---

[13] While SEU theory is very permissive with respect to an agent's value function, there is a case to be made for the theory being biased towards consequentialist ethical theories, perhaps even towards utilitarianism in particular. Colyvan *et al.* (to appear) discuss this issue, and find that the case rests on whether we are interested in a model that does justice to ethical motivations, or whether it is sufficient for the model to yield the right results, according to the various ethical theories. In any case, this is a much more subtle issue than the question of whether the standard decision model can measure value in anything other than monetary or market terms.

[14] There is a large literature on what are suitable ways to account for the interests of future generations. For an overview, see, for instance, Wolf (2003).

[15] For discussions of burdens of proof, see, for instance, Manson (2002), Tickner (2003), the COMEST (2005) report on the Precautionary Principle and the EC (2000) "White Paper on Environmental Liability".

prohibiting a safe action. One way to go about this is to select the level of evidence such that the expected utility of the legal measure in question (say, prohibiting an activity) is greater than the expected utility of not pursuing the measure.

Precaution and the logic of decision. What I have been drawing attention to is that the standard decision model—SEU theory—can play an important role in developing "precautionary" public decision-making practices. I think it is worth dwelling on this point for a moment. Many decision processes could be substantially improved through wiser use of the standard decision model. This means attending to the set-up of the decision problem—the available acts, the range of consequences or outcomes, and the significance of these consequences for the continuing well-being of the community. Attending to these details will have much more immediate impact on contemporary decision outcomes than any subtle revisions of the decision calculus. But having drawn attention to their more immediate significance, I will now also put the tough practical questions of problem formulation to one side. My interests here are theoretical. While I think public decision-making is in many ways well served by SEU theory, I investigate how the idea of precaution nonetheless presents a challenge to the standard decision model. The major themes of this work, explored in Part I and Part II respectively, reflect the two somewhat distinct ways in which "precaution" can be interpreted. The first is the issue of how an agent's future attitudes figure in a decision model (Part I). The second concerns the rationality of "risk-sensitive" choice, which raises questions about the representation and handling of uncertainty in a decision model, and whether SEU theory should have the monopoly on rational choice (Part II). Of course, we should never expect the decision calculus to tell us how much weight to give to our potential interests in 20 years time, or how much we should respect our great grandchildren; these are substantial matters of value. But we do want our decision model to properly take account of whatever our values happen to be in such matters. Likewise, formal decision theory alone is not going to provide us with any substantial standards for belief, but it should allow us to represent any consistent epistemic or risk-oriented attitude towards a given prospect or proposition.

While theoretical challenges to SEU theory may be subtle issues for applied decision-

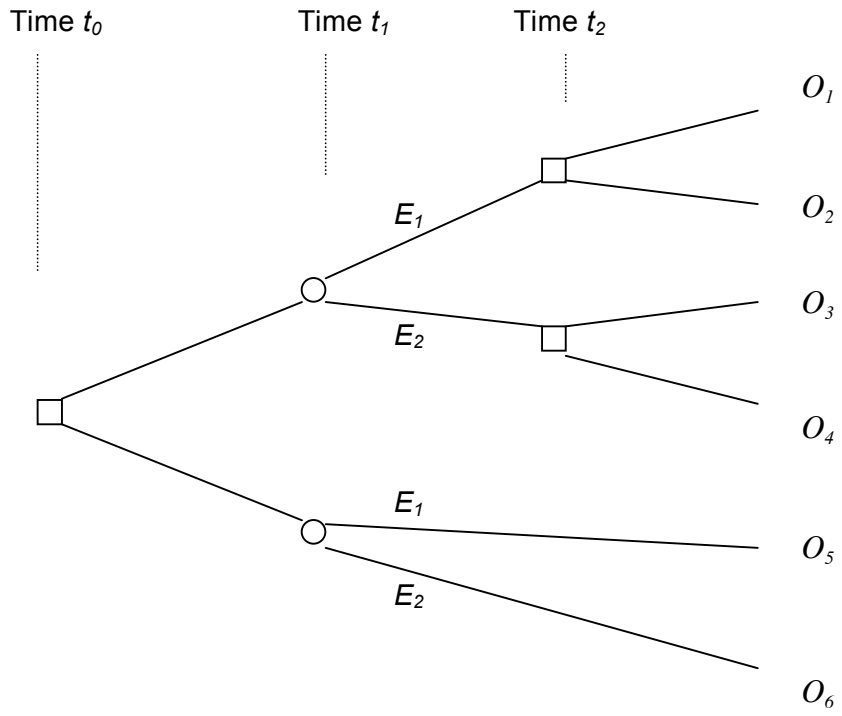
making, given the crude sorts of problems that occur in practice (such as a failure to attend to long-term consequences in the first place), debates about the logic of decision nonetheless have ongoing importance. Consider an agent who prefers a sure $100 to a lottery that offers a 10% chance of winning $1000. Our agent may prefer the sure $100 precisely because it is guaranteed, whereas the lottery has uncertain outcomes. It is important to understand what this kind of risk-sensitivity amounts to, and we do not want to criticise risk-sensitive agents if it cannot be shown that such behaviour is irrational. Furthermore, if we want to treat future interests in decision-making seriously, then we need to explore how the temporal dimension is represented and handled in the decision model.

Both the issues just outlined are, I think, best investigated in the context of a "sequential" rather than a "static" decision model. I will detail the main differences between the two types of decision model in the next section. In brief, sequential- or dynamic-decision models draw attention to the series of choices that an agent may expect to face. Such models make explicit that the current choice is part of a decision *strategy*, rather than being concerned with isolated immediate actions. While this may sound reasonably straightforward, there is in fact considerable controversy surrounding the characteristics of rational sequential choice, and how sequential- and static-choice models should be reconciled. Indeed, I outline my position with respect to these contentious preliminary issues—how we should formulate and "solve" sequential decision problems—in Chapter 1 of this thesis. In later parts of the thesis, I show that, as well as assisting with an agent's on-the-ground decision-making, so to speak, the sequential-choice model has come to play an important justificatory role with respect to the key decision-making norms pertaining to "risk" and "the future". The model provides a rich context for assessing decision rules (both SEU theory and its competitors). It also sheds light on how an agent should relate to their future beliefs and preferences. In fact, I hope to show that there are some important links between the sequential-choice arguments that can be put forward in relation to these seemingly distinct issues.

The basics of the sequential/static distinction. While there are different views about

the role and finer details of the dynamic/sequential decision model, it has some obvious special characteristics. On paper, at least, dynamic and static decision models look very different. Typically, the static model has tabular (or normal) form, as depicted in Figure I-1, and the sequential/dynamic decision model has a tree (or extensive) form. An example of a sequential or dynamic decision model is depicted in Figure I-2 below. The initial node is the position from which the agent must make their choice. The rest of the tree, extending from the initial branches, shows how the agent expects the world to unfold with time. The ever-splitting branches of the tree distinguish different possible scenarios or sets of worlds that may turn out to be actual, depending on how some series of chancy events and future choices are resolved. Indeed, we generally call a decision model 'dynamic' if it depicts decisions being made after the resolution of some uncertainty (Machina 1989, p. 1632). For example, consider again Mary's dilemma about whether to trade in her old car. Perhaps she realises that in 6 months time, she will be on holidays again and so will have time to reconsider whether she wants to buy the new "green" car. Thus if Mary persists with the old car, she will be able to reassess her options in 6 months time. Figure I-2 can be considered a sequential representation of Mary's decision scenario. The choice at the initial node is whether to buy the new car (down) or persist with the old (up). One of two events forming a partition $\{E_1, E_2\}$ will occur at time t_1. Or, more accurately, at time t_1 Mary expects to learn that one of $\{E_1, E_2\}$ is true. (The standard convention is that choice nodes are represented by boxes, and chance nodes by circles.) Let's say E_1 corresponds to the state where the petrol price is greater than or equal to \$1.40/Litre. The other possibility, E_2, represents the petrol price being less than \$1.40/Litre. At time t_2, if it was the case that she initially chose "up" (keeping the old car), Mary will face another choice as to whether she wants to buy the new car. The final outcomes must take into account potential further rises in petrol price, but I will not go into these details here; this sketch is enough to convey the basic idea.

Figure I-2: Sequential decision model



As mentioned, the sequential model makes it clear that an agent must choose a strategy, as opposed to a single instantaneous act. A strategy is just a sequence of choices that the agent plans to make. Let me leave Mary's plight aside now, and just concentrate on the abstract representation of the decision problem in Figure I-2. For instance, one strategy is to initially choose "up", then "up" again if E_1 is learnt (leading to outcome O_1), or else "down" if E_2 is learnt (which will lead to outcome O_4). Another strategy is simply to choose "down" from the outset. This will lead to outcome O_5 if E_1 is learnt, or else outcome O_6 if E_2 is learnt. As to be expected, we can translate sequential or extensive-form decision models into normal-form models. But just what is an appropriate translation from extensive to static form is a matter of some controversy; it is the first issue that I address in Chapter 1 of this thesis. As a first approximation, however, we might trace out all the possible choice combinations and list each as a possible strategy or act. For the problem in Figure I-2, this would result in a static model as shown in Figure I-3.

Figure I-3: Possible static model of problem in Figure I-2

ACT	E_1	E_2
At t_0: up. Then up if E_1, else up if E_2.	O_1	O_3
At t_0: up. Then down if E_1, else up if E_2.	O_2	O_3
At t_0: up. Then up if E_1, else down if E_2.	O_1	O_4
At t_0: up. Then down if E_1, else down if E_2.	O_2	O_4
At t_0: down.	O_5	O_6

I will argue in Part I, Chapter 1, that it is not necessarily the case that all of the combinations of choices at choice nodes are in fact possible, given the agent's current beliefs about their future attitudes. This is to say that we should not just assume that all of the identifiable decision strategies in a sequential-choice scenario are available options for the agent, as we did in formulating the table in Figure I-3. Nevertheless, the figures above present a good initial sketch of the relationship between sequential- and static-form decision models.

SEU theory as foundations of Bayesianism. I have stated that my aim in this thesis is to investigate the subjective expected utility (SEU) model, with the aim of better understanding "precautionary" decision-making. This entails examining the dynamics of decisions—what sorts of predictions about future beliefs/preferences are appropriate, and how these future attitudes should be taken into account. The other key issue is whether SEU theory provides the only rational approach to risk/uncertainty. Understanding the concept of "precaution" is important for practical decision-making. It should be clear that my investigation of the logic of decision is at least partly motivated by broader debates about precautionary decision-making in the public realm. Importantly, there is another thread of questioning that runs through this thesis. Not only does SEU theory have direct practical relevance, but the model also

underpins a broader theoretical approach towards probabilistic reasoning and scientific inference. This is something that I will continually draw attention to when considering challenges to the standard decision model.

To begin with, I uphold a particular theoretical outlook towards decision models. My approach can be called "Bayesian" in the sense that I treat decision models as being ultimately concerned with the *subjective* attitudes of an agent, be this agent a single person or some sort of unified group. This is to say that the probabilities, as well as the utilities in the decision model, are given a subjective interpretation. It is not so obvious that this should be the case. Some simple decision examples might tempt us to think that what we really want to model are objective probabilities. For instance, most would agree that an agent should value a bet that pays $100 if a fair coin lands heads, and nothing otherwise, at 0.5 multiplied by the utility of receiving $100 plus 0.5 multiplied by the utility of receiving nothing. Importantly, the 0.5 here seems to be an objective probability, independent of the opinions of the agent. But even in these straightforward sorts of decision problems, a Bayesian holds that the probability of the coin landing heads represents the agent's belief about the occurrence of this outcome, as opposed to representing some other feature of the world, such as the physics of the coin. Of course, I am not suggesting that an agent's beliefs should have no objective basis; it is of course wise to try to match one's beliefs with the way the world really is. This might involve attempting to align one's subjective probabilities with the objective chances (if there are any of the latter). The point is just that, within the context of this thesis at least, the probabilities and utilities in a decision model ultimately represent an agent's subjective point of view.

The subjective/objective distinction becomes very important in the discussion of risk/uncertainty in Part II. In Chapter 5, I will be concerned (at least in part) with the degree to which the probabilities in a decision model are supported by evidence. For my purposes, it is important that any way of accommodating such variation in evidential support within the decision model respects the basic principle that what we are trying to model is an agent's subjective state of belief. In short, I focus on modifications of SEU theory that are in the spirit of the "Bayesian" approach. I should

point out, however, that I will not attempt to defend these Bayesian sympathies in any detail here. There may be other ways to understand differences in evidential support, or, more generally, other ways to conceive of the standard decision model (e.g. using objective probabilities and utilities), but, in this thesis, I do not consider such possibilities. In any case, decision-making is an activity for intentional agents, so it is very natural to interpret the probabilities and utilities in subjective terms. And indeed, the above-mentioned representation theorems (particularly those of Savage and Jeffrey) make it explicit that the probability and utility functions concern the subjective attitudes of an agent. Moreover, the Bayesian approach has found much favour amongst contemporary decision-theorists.[16]

Not only are the probabilities in the SEU model taken to be subjective, many think that SEU theory provides the *best defence* for modelling subjective belief with probabilities. This basic principle is surely the crux of the "probabilist" or "Bayesian" school of thought.[17] The general idea is that belief is a nuanced kind of subjective state that comes in many shades or degrees, as opposed to being a mere binary (yes/no) matter. And then there is the basic coherency requirement that partial beliefs conform to the probability axioms. (I discuss this coherency claim in some detail in Chapter 2.) Beyond that, the core Bayesian doctrine involves a limited number of other rules, most importantly the rule of conditionalisation (based on Bayes' rule) for updating partial beliefs upon receipt of new evidence, which I also discuss in Chapter 2.[18] Although simple, the Bayesian epistemological approach has many fruits. Indeed, many of the fruits are a direct result of the simplicity. For instance, just the fact that the Bayesian model gives a concise account of how new evidence ideally affects an

---

[16] In addition to Savage (1954) and Jeffrey (1983), most of the decision theorists cited in this thesis can be understood as working within the "Bayesian" framework, in the broad sense of the term. Such people include (and note that this is very far from an exhaustive list of "Bayesian" decision theorists): Skyrms (e.g. 1986), Levi (e.g. 1980), Lewis (e.g. 1981), Joyce (e.g. 1999), Armendt (e.g. 1986) and Seidenfeld (e.g. 1988).

[17] I acknowledge that the term "Bayesian", in particular, is used in a variety of ways, but I think it is safe to say that Bayesians of all stripes hold that partial belief should conform to the probability calculus.

[18] Lewis's (1980) Principal Principle might be considered another constraint on partial belief (dictating how partial belief should relate to objective chance), but this is a much more controversial rule.

agent's set of beliefs can be considered a major achievement. This basic evidential framework allows for various quantitative comparisons of the amount that a given piece of evidence confirms competing hypotheses (see, for instance, Howson and Urbach 1989, Earman 1992, Fitelson forthcoming). Indeed, here we have the underpinnings of Bayesian statistics, which might be considered a family of approaches to scientific experimental design and inference that uphold the probabilistic approach to theory appraisal (Howson and Urbach 1989, for instance, introduce the central tenets of Bayesian statistics). Moreover, because the Bayesian approach is typically justified in pragmatic terms, it makes clear what the connection is between belief and action.

When it comes to challenges to SEU theory, it could be said that a lot hangs in the balance. Let me give a very brief account of how SEU theory is thought to underpin probabilism or Bayesianism. The central claim is that it is only through understanding rational ordinal preference that we can determine what are the properties of rational belief. This is one thing that is supposedly achieved by the SEU representation theorems mentioned earlier (particularly Savage's or Jeffrey's). Arguably, the theorems show that only probabilistic belief functions are rational, because only this type of belief function can be reconciled with ordinal preferences that are rational by the lights of SEU theory. A related pragmatic justification for probabilism/Bayesianism is the synchronic Dutch book argument (DBA). In essence, a Dutch book argument shows that that under specific (and supposedly very reasonable) betting conditions, an agent who does not satisfy the nominated epistemic norm may suffer sure loss, or an avoidable bad outcome, and is thereby shown to be irrational. In addition to the synchronic DBA, there is a "diachronic" DBA—a Dutch book argument that makes use of the sequential-choice framework—for the belief-updating rule of conditionalisation. I address both of these Dutch book arguments in Chapter 2. What I draw attention to now (and will discuss in more detail in Chapter 2) is that all of these pragmatic justifications of Bayesianism rely (to varying extent) on expected-utility principles. The significance of this is that any attack on SEU theory makes for an attack on these pragmatic defences of Bayesian epistemic norms. I do not mean to say that this observation should curtail precaution-based criticisms of SEU theory, just that it is worth bearing in mind both the practical and the theoretical

import of any such challenges.

Overview of chapters. I will now give a brief overview of the chapters in this thesis. It should be clear that, in the process of exploring the dynamics and risk/uncertainty features of the SEU model, I am continually aware of the theoretical role that SEU theory plays in the broader Bayesian story. Recall, the basic structure of the thesis: Part I (Chapters 1–3) chiefly concerns how we should account for our future attitudes in decision-making, while Part II (Chapters 4–6) investigates the key axioms of SEU theory that govern its approach to risk/uncertainty. It is hoped that the following brief chapter descriptions will serve as a reference point that the reader can return to.

1: RECONCILING STANDARD 'ONE-SHOT-ONLY' DECISION WITH SEQUENTIAL CHOICE

This chapter sets the foundations for what is to follow—I argue for a particular position with respect to the sequential-choice model. To begin with, I claim that sequential- and static-choice models should not give an agent conflicting advice. In other words, the sequential model just serves to unpack the dynamics that are implicit (or should be implicit) in a standard "one-shot-only" decision model. It remains to navigate the conflicting approaches to sequential choice that have arisen in the literature. I argue for the sophisticated sequential-choice approach, because this is the most plausible account of how both the past and the future should affect an agent's current choice of strategy.

2: THE PRAGMATICS OF BELIEF

It is generally thought that Bayesian belief-updating norms constrain how a rational agent should assess sequential decision strategies. Conversely, SEU theory and the sequential-choice model play a role in interpreting and justifying Bayesian epistemic norms. Here I investigate this two-way relationship between pragmatic and epistemic rational principles. I consider both the synchronic and the diachronic "Dutch book arguments" for, respectively, "probabilism", or partial beliefs conforming to the probability calculus, and the rule of conditionalisation for updating beliefs. In particular, I draw attention to the "weak points" of these epistemic norms—firstly,

they both rest on some controversial pragmatic assumptions, and secondly, they do not provide as much direction to aspiring rational decision makers as one might initially think.

3: PLANNED CHANGES IN DESIRE

I consider how lessons about belief change, and the rule of conditionalisation in particular, should be brought to bear on the issue of preference or desire change. In many ways, it seems that desire can be, and indeed is, treated analogously to belief in the standard decision model. But clearly the two kinds of propositional attitude have some important differences. Here I argue that, unlike belief, there is a further way (beyond conditionalisation) in which an agent can be understood to *plan* a preference change—the agent might favour a genuine change in taste, and so choose a strategy that makes the relevant change in their utility function more likely. I explore this possibility through reference to "higher-order" preferences.

4: ALLAIS'S PROBLEM AND THE INDEPENDENCE AXIOM

I introduce the notion of the risk/regret associated with particular acts. I appeal to Allais's problem because it is a carefully formulated to isolate the risk/regret issue. My particular focus here is whether the risk-sensitive "Allais-choices" can be reconciled with SEU theory, or whether they indeed contradict the independence axiom. The question hinges on what, if anything, constrains the content of, as opposed to the relationship between, our preferences. While there are a number of ways in which risk/regret might feature in the description of outcomes, I argue that any such risk/regret "property" must answer to some suitable consistency requirements if it is to count as a genuine property of outcomes. Otherwise there is the danger of SEU theory being irrefutable or lacking content.

5: ELLSBERG'S PROBLEM AND THE ORDERING AXIOM

With reference to Ellsberg's well-known decision problem, I argue that there are good normative reasons for representing indeterminate belief with *sets* of probability distributions. At the least, however, this amounts to forgoing the SEU requirement of

a complete preference ordering. Moreover, while such a model can make the "Ellsberg-choices" rationally permissible, without some further element to the story, it does not explain how an agent may come to have *unique preferences* for the less "risky" options in the two Ellsberg decision problems. Levi (1986) holds that the extra element amounts to innocuous secondary "risk" considerations that are used to break ties when more than one option is rationally permissible. While I think a lexical choice rule of this kind is very plausible, I argue that this sort of risk-sensitivity involves a greater break with SEU theory than mere violation of the ordering axiom.

6: ASSESSING DECISION RULES IN THE SEQUENTIAL-CHOICE CONTEXT

This final chapter concerns the justification of decision rules. My investigations in previous chapters show that there is good motivation for relaxing both the ordering and independence constraints on rational choice. Here I consider whether the sequential-choice context reveals there to be any subtle problems with such modifications of SEU theory. I initially analyse Hammond's so-called "consequentialist" (1977, 1988c) argument for upholding SEU theory in its entirety. While I do not agree with Hammond's rationale, it provides inspiration for an alternative sequential-choice argument that is reminiscent of the diachronic Dutch book argument for conditionalisation. I then consider Seidenfeld's (1988) claim that specific types of decision rule are subtly inconsistent in the sequential-choice context. The question is how these various sequential-choice arguments stack up when it comes to the key decision theories discussed in this thesis.

I PLANNING FOR THE FUTURE

1 RECONCILING STANDARD 'ONE-SHOT-ONLY' DECISION WITH SEQUENTIAL CHOICE

1.1 Introduction

I have indicated in the introduction that what we call the "sequential" representation of a decision problem is going to be very prominent in this thesis; the dynamic context brings to light additional decision-making considerations, and this in turn leads to a richer understanding of subjective expected utility (SEU) theory and its rivals. In the introduction, I also gave a sketch of the differences between the sequential and static representations of a decision problem. A superficial distinction is that the former is packaged as a tree or extensive-form model, while the latter has tabular or normal form. More substantially, the sequential-choice model depicts the full sequence of choices that an agent expects to face, as well as the uncertainties that are expected to be resolved at various future times. These dynamics are only implicit in the static model—it depicts isolated acts and the outcomes that will potentially result from these. It might be thought that the relationship between these two sorts of decision model should be rather straightforward. But this is a matter of some controversy, as I will make clear in this chapter. Several different sequential-choice approaches have been articulated, the most prominent being the resolute and sophisticated approaches. Disagreement surrounding the sequential-choice model is in fact not surprising, because it seems that issues to do with the temporal dimension of choice have largely been swept under the carpet in traditional decision theory debates. The relatively recent interest in sequential-choice models has thus served to bring a number of outstanding issues to the forefront. In what follows, I articulate, and argue for, my particular position on how we should approach the sequential-choice model, including its relationship with the standard "one-shot-only" representation of a

decision problem. This will set the foundations for my analyses of Bayesian decision norms in later chapters.

1.2 Standard decision theory and the temporal dimension

Standard decision theory does not make a big deal of the temporal dimension of choice. Indeed, there is room for confusion in both Savage's (1954) and Jeffrey's (1983) theories with respect to how we should conceive of the temporally-extended consequences of acts. For instance, if we consider Lewis's (1981) possible worlds rendition of Jeffrey's decision-theoretic framework, we have a model where acts, states and outcomes all amount to propositions, which can be cashed out as sets of possible worlds. These worlds apparently describe a whole history, as opposed to, say, representing possible states of affairs at some snapshot in time just after a proposed action is performed. While the model stipulates that possible worlds are temporally extended, the details of these world histories are, however, glossed over. In other words, the decision maker is directed to consider the impact of their action now and in the future, but beyond that not much detail is provided as to what sort of future trajectories are indeed possible.

The ambiguity about possible futures in the standard model has, I think, led to decision theorists talking past one another on a number of issues. It is my hope in this chapter to cut through such confusion. First, we need to be clear about the kind of problem our decision theory is supposed to answer, or, in other words, what kind of agent we are trying to model. Both Savage and Jeffrey are of course in the business of normative decision theory. They are concerned with the ideally rational agent, at least from the pragmatic perspective. But I think implicit in this framework is the fact that standard decision theory is only interested in modelling ideal practical rationality *at a time*. The task is to choose the best act, given an assessment of the outcomes or possible worlds that each act may yield, where these possible worlds can include any manner of future behaviour. The model does not require a continuing ideally-rational

agent, nor even the expectation that the agent will continue to be so.

At least, this openness about the agent's future behaviour, I think, is the best way to interpret the standard decision model. Joyce (1999, pp. 57–61) notes that it may not be how Savage himself envisaged the relationship between present and future choices. According to Savage (1954, p. 17), "the person decides 'now' once (and) for all; there is nothing for him to wait for, because his one decision provides for all contingencies". Now it is unclear how comprehensive Savage's so-called contingencies are. Do they, for instance, call into question the agent's own future rationality, or permit any change or uncertainty with respect to the agent's future beliefs and preferences? It seems not, as Savage (1954, pp. 16) elsewhere says

> What in the ordinary way of thinking might be regarded as a chain of decisions, one leading to the other in time, is in the formal description proposed here regarded as a single decision…it is proposed that the choice of a policy or plan be regarded as a single decision.

But Savage is not very careful about his use of the term "act". He generally does not even acknowledge possible external interferences to an agent's plans, let alone changes in the agent's own belief/desire attitudes. For example, Savage (1954, p. 15) describes a situation in which an agent is about to add a final suspect-looking egg to an omelette. His two acts are "break into bowl, and in case of disaster have toast", and "break into bowl, and in case of disaster take family to a neighbouring restaurant for breakfast". The latter "act", in particular, is not something that the agent can simply choose to do, because it depends on the agent's family cooperating with the proposal, on getting to the restaurant safely, on the restaurant being open and willing to serve breakfast, etc. While there are good reasons for making such oversimplifications when sketching an example decision problem,[19] I nonetheless suggest that we do not pin our interpretation of the standard decision model, including its treatment of future

[19] Indeed, I am being a bit unfair on Savage here. We simply must use rough descriptions of acts and outcomes when giving example decision problems, because the precise details would be overwhelming.

contingencies, to Savage's own comments about acts and strategies.

Joyce (1999, pp. 57–61), I think, gives a more careful account than Savage of what an "act" is, and how Savage's (as well as Jeffrey's) decision model should accommodate future choices/actions of the agent. Strictly speaking, actions refer to only that which the agent currently has direct control over.[20] For example, Joyce considers the supposed act "walk to work in the rain". While this is the way we generally think of acts, the agent does not in fact have direct control over whether they "walk to work in the rain". Just like the example above about taking one's family to a neighbouring restaurant for breakfast, walking to work in the rain is a temporally extended rather than an instantaneous act, and so it depends on the cooperation of future time slices of the agent (which may not be assured). Furthermore, during the period in which the act takes place, a number of external events may frustrate the agent's plans, even if they do maintain the resolve to brave the wet weather on their way to work. For instance, the neighbour might accidentally run the agent down with a bicycle when they step out onto the footpath. The point here is that it is useful for an agent to make temporally extended plans, but the carrying out of any plan is a matter of probability rather than certainty.

Like Joyce, I think it is a mistake to think of a sequence of choices as an "act", because this is hardly something that the agent has control over *at a particular time*. While the agent should, of course, think carefully about future contingencies, and the choices they expect to face, he/she only ever has the power to enact the next step along any proposed path of action. Of course, it can be useful shorthand to refer to an extended plan as a single act, because, after all, *resolving* to follow a strategy makes it highly likely, in most cases, that the strategy will be followed. Furthermore, there are arguably restrictions on a rational agent's *plans* regarding their future

[20] Here I am talking about a decision model that is depicted in as fine detail as possible. Such a model is referred to as a "grand world" representation of the decision problem. Acts, states and consequences are individuated with respect to *all* relevant differences. (A "small" world model has, for instance, states that include a number of possible worlds with differing utility for the agent.) See Joyce (1999, pp. 70–77) for a more thorough comparison of "grand" and "small" worlds.

beliefs/preferences, something that I will go on to discuss in some detail in Chapters 2 and 3. I think it is reasonable to begin with the assumption, however, that an agent might predict, in at least some cases, that their beliefs or preferences will change in ways that are not entirely planned, and that are somewhat mysterious from their current point of view. We are to some extent estranged from our future selves. To put it a bit differently, there is surely some possibility that my future self will have quite different likes and dislikes, or else different sorts of beliefs about the world, from the ones that I currently hold. But here we return to the question that I do not think is sufficiently well answered in either Savage's or Jeffrey's respective theories—how should an agent take into account, in their current decision-making, whatever predictions they happen to make about their own future beliefs and preferences?

1.3 What should we expect from a sequential-choice model?

Just as there is some ambiguity in the static model when it comes to accommodating future choice behaviour, there is a similar ambiguity associated with sequential choice. On first appearances, it might seem that dynamic or sequential-choice models assume that the agent is ideally rational at all stages in the sequential decision problem. After all, a sequential model makes explicit all the decisions that an agent must face, and it is reasonable to think that a normative model will stipulate how *all* these choices should be made. Indeed, McClennen (1990, p. 202) makes the comment that stories such as Ulysses and the sirens (where Ulysses expects to take on bizarre preferences, or suffer an inevitable weakness of will when he passes the island of the sirens) should be considered examples for an imperfect theory of rationality that deals with the foibles of ordinary persons. McClennen states that his theory of dynamic choice concerns, rather, the ideally rational agent, both now and in the future. But this introduces an interesting puzzle: in some cases (e.g. Ulysses' predicament) the standard one-shot-only decision model will recommend a different action from McClennen's dynamic choice model. The latter counts on the rationality of Ulysses' later choices (by the lights of his current beliefs and preferences) whereas the former

makes no such assumptions.[21] At least, this is the case if we subscribe to Joyce's interpretation of acts and future contingencies with respect to the static model. I have argued that this is the more robust reading of the standard decision model, and Savage's loose comments about acts should not be taken too literally.

Let us consider for a moment what exactly it means for an agent to be ideally rational, both at the current and at all later times. I am referring to an agent who at any given time has SEU preferences, and whose beliefs and preferences are expected to change only in accordance with rational principles. Just what these principles of rational belief/preference change are supposed to be is not something that I want to address just now. I discuss this question in some detail in Chapters 2 and 3. But, as mentioned in the previous section, whatever the details of belief and preference updating rules may be, I think it is safe to say that there will be at least some circumstances in which an agent expects to be somewhat disconnected from their future self. And I don't think this should count against the agent's *current* rationality. If the agent predicts some unusual change in their outlook, then so be it; surely we should take such predictions into account in decision-making, rather than pretend that ideal conditions will prevail.

Given my comments above, I prefer to begin with a dynamic choice model that is entirely general—it should be able to represent in expanded form the future predictions underlying a one-shot-only choice, whatever expectations the agent has about their future self. In extreme cases, an agent might think that, despite their best intentions, their future self will suddenly have perverse preferences, such as a desire to be sold into slavery, or will be under the delusion that they are a super hero and able to fly, or will cease even to have preferences that conform to SEU theory. We want our sequential-choice model to be capable of representing all such possibilities. The idea I am pushing is that, while the dynamic framework provides an expanded

---

[21] McClennen is well aware of the fact that a decision model, whether sequential or otherwise, should deliver agents like Ulysses the right advice. He thus claims that the agent must, in the end, choose a strategy that is psychologically feasible as well as objectively feasible or ideally rational. I will discuss McClennen's sequential choice approach in more detail in the next section.

tree-form representation of a decision problem, it should not introduce extra idealisations that constrain the set of possible worlds arising from a particular action. Of course, a generalised model will still allow us to examine the special cases—those cases where the agent has particular expectations about their future self. Indeed, the special cases play a large role in later chapters, when I consider sequential-choice arguments for particular decision norms. The point is that it is better to start with the general model and then build in assumptions that allow us to investigate specific cases, rather than formulate a dynamic choice model so idealised that in many cases it does not in fact give the agent the right advice about what action/strategy they should pursue in ordinary decision-making scenarios.

So I am making two claims here regarding the right way to envisage the relationship between static and sequential decision models. The first is that the sequential model should be understood simply as an expanded representation of a decision problem. It simply makes explicit the temporal structure of the decision problem that is implicit (or should be implicit) in a static or normal-form representation. In other words, the dynamic framework is just a different type of *representation*, rather than a different type of *problem* from the kind that Savage and Jeffrey address. The second claim is that both static and sequential models should accommodate whatever predictions an agent may have reason to make about their future choice functions. This is not to deny the fact that there are plausible belief/preference updating rules that a rational agent should subscribe to. It is just that an agent may not always have complete control over their prospective future selves. The fully generalised decision theory should not mandate a particular progression of beliefs and preferences along each path of a decision tree, and it should not assume that future choices will always be rational by the agent's current lights.

Note that there has been some previous debate in the sequential-choice literature about whether static (normal-form) and sequential (extensive-form) decision solutions should coincide. Some (but not all) sequential-choice approaches are thought to yield identical solutions to their static counterparts, or in other words, they are purported to

obey what McClennen (1990, p. 161) refers to as "reduction to normal form".[22] In fact, McClennen favours a sequential-choice approach that obeys this condition. Levi (1991) and Seidenfeld (1994), on the other hand, argue that "reduction to normal form" is not an important condition for sequential-choice models to satisfy. Their claim is that there can be legitimate differences between the solutions to the normal-form and extensive-form models of a decision problem. My own position differs from both of these stances. Like McClennen, I think it is important for normal- and extensive-form decision solutions to coincide. But I do not agree with McClennen's approach to sequential choice, and I think he introduces unnecessary idealisations into the basic sequential-choice model. As will become apparent, I support the sequential-choice model advocated by Levi and Seidenfeld, but I do not think we should accept a difference between the normal- and extensive-form decision solutions.[23] If a static model does not recommend the same action as its sequential counterpart, then it is simply wrong. In my opinion, normal-form models should be pulled into line with the sequential details of a decision problem, in the spirit of Joyce's comments about how to properly understand acts and future contingencies.

1.4 Comparing sequential-choice models

Three major approaches to dynamic choice models have appeared in the literature.

---

[22] McClennen (1990, p. 161) holds that sequential-choice approaches can be distinguished according to where they stand with respect to the following three conditions: "dynamic consistency", "separability" and "reduction to normal form". The first stipulates that an agent should carry out their chosen plan to completion, if all beliefs/preferences change as predicted. The second stipulates that an agent's choice at a particular node should depend only on their beliefs/preferences at that node and not on the greater structure of the sequential decision tree. "Reduction to normal form" is the property that we are interested in at present; it stipulates that the extensive-form and normal-form decision solutions should coincide. I do not refer to these properties explicitly in my discussion of the various sequential-choice approaches in the next section, but I do cover the same issues.

[23] Admittedly, it is useful to talk about the distinction between normal- and extensive-form decision solutions if the context is one in which we are trying to assess Bayesian decision norms (issues that are taken up in later chapters of this thesis). This is indeed the context in which Levi and Seidenfeld refer to the condition of "reduction to normal form". Here we are not interested in questions of justification, however; we are interested in modelling the fully general decision problem, as opposed to some class of idealised decision problems.

These are the naïve or myopic approach, the resolute approach and the sophisticated approach. I will argue that only the sophisticated approach can properly unpack the dynamics that underlie the fully general decision problem. It is the one sequential-choice method that can handle the full range of predictions that an agent might hold with respect to their future beliefs and preferences. In my opinion, this makes sophisticated choice the only approach that can be reconciled with the static representation of a decision problem. Both resolute and naïve choice make dynamic-decision models depict *different kinds of decision problems* from the familiar kind in which all that is expected of an agent is that they be rational *now*. As stated, this is problematic because surely rationality should recommend only one course of action for an agent at a particular time, rather than one according to the familiar static approach and another according to the dynamic approach. In what follows I will outline the sophisticated method first and then explain how the other two approaches differ (to their detriment).

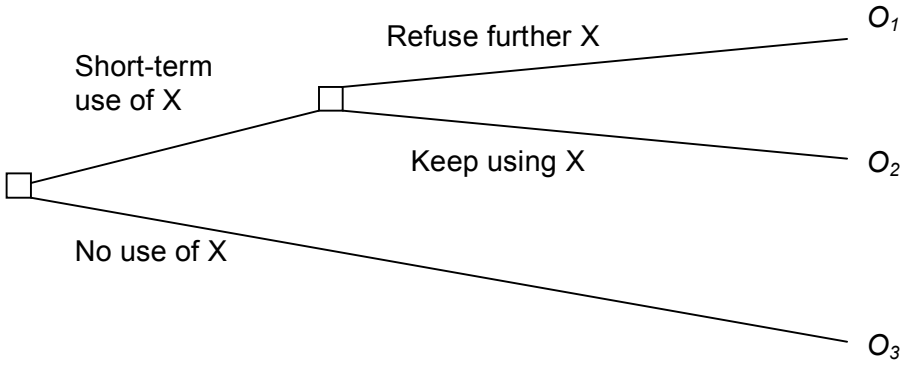
Proponents of the sophisticated approach to dynamic choice include Seidenfeld (1988a), Levi (1991), Maher (1992) and in passing, Earman (1992). Let us assume that the decision tree shows, at each choice node, all possible options that will be available to the agent at that node.[24] The hallmark of the sophisticated approach is its emphasis on backwards planning; the sophisticated chooser does not assume that all paths through the decision tree, or in other words, all possible combinations of choices at the various choice nodes, will be possible. The agent considers, rather, what they will be inclined to choose at later choice nodes when they get to the temporal position in question. Indeed, the agent starts with the final choice nodes in the tree, and considers what will be chosen at these nodes, given their predicted preferences at each of these positions. These choices are effectively locked in, so when the agent next considers the second last choice nodes, they make their assessments in light of how the final choices will be resolved. Branches of the tree that involve choices at the final nodes that simply will not eventuate are dead options

[24] This is actually an important point to note about the extensive-form (or sequential) representation of a decision problem. For instance, if the decision tree showed only those options that the agent would be inclined to choose at each choice node, then there would be no difference between the extensive and normal form models. The models differ only when all available options are depicted at choice nodes.

and should not influence choice. Once the second-last choices have been determined, the agent moves to the third-last choice nodes, and so on back to the initial choice node. The result is that only certain paths through the decision tree are deemed possible, and only these strategies are worth evaluating and deciding between, at the initial choice node.

The "potential addict" decision problem described by Hammond (1976, p. 161) is useful for illustrating sophisticated choice. An agent must decide now whether to start taking some pleasurable drug X or whether to abstain. Drug X is completely harmless in small doses. The trouble is, X is not only addictive, but also very harmful after sustained use. Inevitably, once the agent starts taking drug X, they will be faced with a further decision as to whether to continue taking it. The expected course of events is illustrated in Figure 1-1. The final outcomes O_1, O_2 and O_3 represent "short-term use of X", "long-term use of X", and "no use of X", respectively.

Figure 1-1



From the perspective of the initial choice node, the agent prefers O_1 to O_3 and O_3 to O_2. But the agent is a sophisticated chooser and so they consider first what their choice function will recommend at the second node. The agent knows, sad but true, that at this node, after some initial use of drug X, they will prefer O_2 to O_1, contrary to their current preferences. So the second node effectively leads to option O_2. The sophisticated agent then wisely chooses down (option O_3) at the initial choice node,

because they realise that what they face is really a choice between outcomes O_3 and O_2, and their current preferences adjudicate in favour of O_3.

Before moving on to naïve choice (which corresponds to sophisticated choice only under certain conditions), I want to briefly consider cases where the agent is unsure what their future preference function will look like. Perhaps the agent assigns probabilities (representing degrees of belief) to various alternative preference functions that they might hold at some choice node in the future. In this sort of case, each possible preference function will dictate a particular choice at the node in question, so the agent will end up assigning probabilities to the choices they might make at this point. While slightly more complicated, this kind of scenario is well handled by the sophisticated approach. The agent again simply plans in a backwards fashion, beginning from the final choice nodes in order to determine probabilities of particular paths of action, which then inform choices at earlier nodes (and these choices might also be probabilistic if the agent is unsure of their preferences here too). To give a simple example, the agent facing the "potential addict" problem (depicted in Figure 1-1) might have credence 0.6 in holding a choice function at the second node ranking outcome O_2 over O_1, and credence 0.4 in holding a choice function at that time ranking outcome O_1 over O_2. Thus, the expected value of choosing "up" at the first node will be 0.6 × (utility of O_2) + 0.4 × (utility of O_1). Whether or not the agent should choose "up" or "down" in this problem then depends on how the sum just mentioned compares with the utility of outcome O_3.

Unlike the sophisticated approach, the naïve or myopic approach to dynamic choice runs into problems when it comes to cases like the "potential addict". The "naïve" agent assumes that any path through the decision tree is possible, and so sets off in pursuit of whichever path they have calculated to be optimal. For instance, when faced with the problem in Figure 1-1, the naïve chooser opts for the strategy leading to outcome O_1, (short-term use of drug X), because this is their most preferred outcome. If the agent had taken into account their preferences at the second node, however, they would have realised that it is futile to set off in pursuit of O_1, because once they reach the second node, they expect to change strategies and seek *b* instead.

The agent's plans in fact commit her to long-term and harmful use of drug X, and it thus seems that the naïve/myopic chooser does indeed deserve to be called "naïve".

One might wonder why anyone would defend the naïve/myopic approach to dynamic choice. Indeed nobody does defend naïve choice in the context of problems like the "potential addict" or "Ulysses and the sirens". Those who do defend this method of assessing strategies must regard dynamic decision problems to be a different kind of problem altogether from the familiar static variety. Or else naïve choice defenders must think both static and dynamic decision problems involve strict constraints with respect to the preferences and beliefs of an agent's future self. Hammond (1976, 1988b, 1988c), for example, defends the naïve or myopic approach (although he doesn't actually refer to his method as such). Unlike my account of sophisticated planning, Hammond is not interested in how an agent should decide upon a strategy given any old beliefs about what their future preference/belief functions will look like. Hammond is interested, rather, in whether an agent's choice functions at various times complement each other. His criteria of rationality apply to the agent both at the time of decision and in the future. The temporally extended agent is rational just in case naïve planning makes sense. In other words, the agent's expected future beliefs/preferences should be such that any path through the decision tree that is deemed optimal in terms of current beliefs/preferences will be viable.

According to Hammond, the temporally-extended Ulysses is simply not rational. The potential addict is not rational either because it is not coherent to expect to have preferences at the second node that differ in this way from preferences at the first node. In fact, Hammond can be interpreted as recommending that an agent should always expect to have stable preferences and beliefs over time, the only changes being those that involve conditionalisation on new evidence. (Even in this case we can say that the agent's beliefs/preferences do not change, because their updating commitments were always present in their initial conditional probabilities.) Under these idealised expectations about the future, Hammond's naïve planning corresponds with sophisticated planning. The question is: must an agent have such idealised expectations regarding their future self? As mentioned, I address this issue in the

detail that it deserves in Chapters 2 and 3. For now, let me just say that there are surely occasions when an agent has good reason to think that they will undergo somewhat inexplicable changes in belief/preference. For instance, Ulysses may well have very good frequency data showing that the majority of men just like himself have in the past been completely seduced by the sirens' sweet song whilst sailing past, despite their best intentions. Likewise, the potential addict may have witnessed that the preferences of many other drug experimenters became distorted by addiction. I think we are wise not to ignore cases in which a perfectly able agent predicts that their beliefs/preferences will change in ways at odds with conditionalisation if they pursue a particular path. Given that the naïve approach cannot handle such situations, I do not think it is the correct way to conceive of sequential choice.

The worries I have about naïve choice turn out to be worries for the "resolute" approach to sequential choice as well. The resolute approach is championed by both Machina (1989) and McClennen (1990). Although they defend it in slightly different ways, the two have a common aim—both Machina and McClennen are interested in relaxing the independence axiom of SEU theory,[25] and they see resolute choice as supportive of this kind of decision-making rule. Machina argues that the structure of a sequential decision tree is very important to the choices that an agent will make at a given time. In other words, the chancy events or option sets that an agent has already faced at various times in the past are relevant to their current decision. The significance of this claim is best illustrated by a well-known decision problem given by Allais (1953). Note that Allais's problem and the independence axiom are the focus of Chapter 4. Here I give only a brief introduction to the problem in order to

---

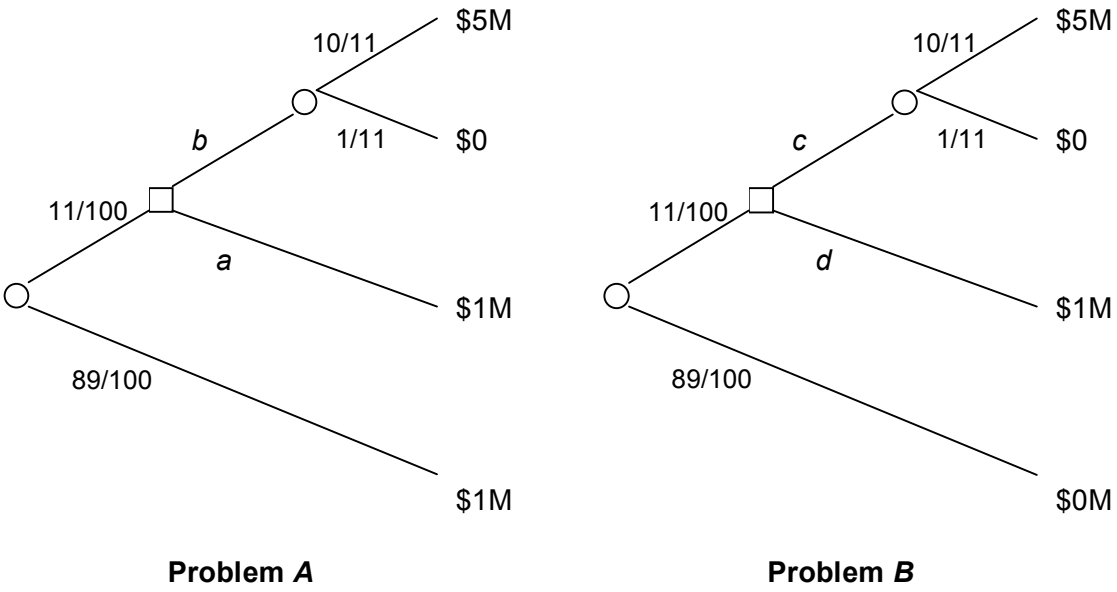
[25] An informal statement of the independence axiom is given by Joyce (1999, p. 86): "a rational agent's preference between (acts) A and A^* should not depend on what happens in circumstances where the two yield identical outcomes." Joyce (1999, p. 85) also formally presents Savage's version of independence as follows:

> Suppose that A and A^* produce the same outcomes in the event that E is false, so that $A_{\sim E} = A^*_{\sim E}$. Then, for any act $B \in \mathbf{A}$ (where $\mathbf{A}$ is the set of all acts, including constant acts), one must have
>
> $A > A^*$ if and only if $A_E \,\&\, B_{\sim E} > A^*_E \,\&\, B_{\sim E}$
>
> $A \geq A^*$ if and only if $A_E \,\&\, B_{\sim E} \geq A^*_E \,\&\, B_{\sim E}$

Note that I restate the independence axiom in Chapter 4, when I discuss it in more detail.

illustrate the rationale for resolute choice. Figure 1-2 depicts a possible sequential version of Allais's famous problem.[26] (There are alternative ways of depicting the problem in sequential form that do not necessarily highlight the point that Machina/McClennen wish to make.)

Figure 1-2



As Allais's story goes, there are two choice situations, and the agent is asked to consider what they would do in each of the two cases. Machina thinks that it is reasonable for the agent to choose *a* ("down" at the choice node) in problem *A* and *c* ("up" at the choice node) in problem *B*, despite the fact that the decision trees in each case are identical *if we consider only the possibilities that remain open to the agent at the choice node*. What Machina, in particular, argues is that the choice nodes in *A* and *B* are not in fact identical, given that they are part of larger decision trees containing different past possibilities. These past possibilities (or risks already borne) can legitimately influence choice, so that it is not irrational to choose differently at the two choice nodes.

[26] This sequential version of the problem was formulated by Raiffa (1961) to argue for the inviolability of the independence axiom.

Sensitivity to past possibilities supposedly allows an agent to be resolute. What Machina has in mind with the Allais example is that the agent in question subscribes to some kind of independence-violating theory such as cumulative prospect theory.[27] When it comes to problem *A*, for instance, such a decision theory may well recommend at the outset that the agent choose *a* ("down" at the choice node). When the agent gets to this position, however, the same choice function might recommend that the agent choose *b* ("up") *if they were to treat the choice node as if it was a new, stand-alone decision problem.* Unlike the naïve chooser, however, Machina's agent does not approach the problem in this way and simply change strategy when they arrive at the choice node. The agent resolutely takes into account their past strategy, and makes a choice that is sensitive to what has come before. In fact, according to Machina, at each point in a decision problem the resolute agent honours an originally chosen strategy, even if it conflicts with what their preferences would recommend were the past considered irrelevant.

McClennen also defends resolute choice in relation to decision theories that relax independence. In the early part of his (1990) book, McClennen (p. 157) offers a similar explanation of resolute choice to the one given above:

> …the agent can be interpreted as resolving to act in accordance with a particular plan and then subsequently intentionally choosing to act on that resolve, that is, subsequently choosing with a view to implementing the plan originally adopted.

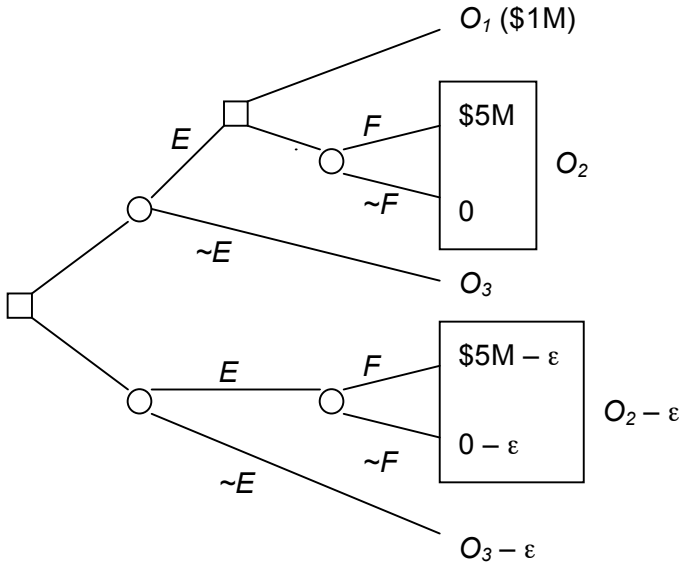
Later in his book, however, McClennen depicts resolute choice as a kind of agreement between an agent's present and future self, and claims that such an

---

[27] Machina (1989, p. 1631) lists a number of decision theories that relax independence, and cumulative prospect theory is one amongst these that has particularly nice properties. I have more to say about independence-violating decision theories in Chapter 4.

agreement should only be expected to be honoured in cases where both present and future self benefit. The Allais problem *A* is not one of these cases, because the agent at the choice node is made worse off by sticking to the original plan; at this point the agent (according to my story about their choice function) would really prefer the lottery that gives a 10/11 chance of $5 million to a sure $1 million. McClennen's revised version of the resolute approach does not prescribe that an agent always honour past resolutions. The agent is expected to be resolute only to the extent that it furthers their preferences at all times. The kind of example that McClennen has in mind is given in Figure 1-3 below.[28]

Figure 1-3



As the story goes, the agent's preferences violate independence in the following way: she prefers O_1 to O_2, but she prefers the lottery that gives O_2 if E and O_3 if $\sim E$ to the lottery that gives O_1 if E and O_3 if $\sim E$. (I will refer to the former lottery as L_2 and the latter as L_1). Some examination reveals that such a sophisticated

[28] The example in Figure 1-3 was in fact formulated by Rabinowicz (1995, p. 599). He uses it to illustrate a potential problem for sophisticated choice that was identified by McClennen (1990, pp. 190–195). The fact that sophisticated choice seems problematic for an independence-violating agent in cases like this in part motivates McClennen's revised version of resolute choice.

agent ends up choosing "down" in the problem depicted here in Figure 1-3, which effectively amounts to the lottery (L_2 - ε) (for some positive ε). This seems mistaken, given that this strategy is, in a sense, dominated by the strategy that amounts to L_2. According to McClennen, this is precisely the kind of situation in which resolute choice is the way to proceed. At all times the agent prefers L_2 to (L_2 - ε), and from the initial vantage point, at least, L_2 (O_2 if E and O_3 if $\sim E$) is preferred to L_1 (O_1 if E and O_3 if $\sim E$). McClennen concludes that it is in the agent's best interests to pursue the L_2 lottery and stick to this plan, despite the fact that O_1 will look better than O_2 at the second choice node, should the agent reach this position.

I think the resolute approach to sequential choice fails for a couple of reasons. The first reason, however, applies primarily to Machina's account of resolute choice (and to McClennen's earlier account, which is similar to Machina's). My concern is that, like Hammond's naïve approach, the resolute method that is being endorsed here makes certain idealising assumptions about the sequential decision problem. In particular, it is assumed that at future choice nodes, the agent will be both capable of and inclined to cooperate with their past self by honouring the strategy that was originally chosen. When it comes to cases like Ulysses or the potential addict, this assumption seems overly optimistic, to say the least. The resolute Ulysses is supposed to simply continue homewards to Ithaca, despite the fact that he is completely overcome by the song of the sirens at the time when he sails past them. Likewise, resolute choice flies in the face of medical statistics when it recommends that the potential addict use the pleasurable drug X several times only and then refuse any further such experiences. Defenders of resolute choice, such as McClennen (1990, p. 202), respond that these sorts of cases involve a weakness of will that should not be experienced by an ideally rational agent, and so they are not the right sorts of examples for a theory of ideal rationality to contend with. McClennen suggests that we are welcome to modify the ideal case in order to capture human weakness of will, by assessing whether strategies are psychologically feasible, after an initial assessment of rational feasibility. This is messy, but so is everyday human reasoning when compared to the ideal case, or so the argument might go.

As I made clear with respect to the naïve approach, I think that the sequential representation of a decision problem should be just that—an alternative (expanded) representation of the same sort of decision problem that Savage or Jeffrey were trying to model. And it is my opinion that whether one employs the static or sequential representation, idealising assumptions should apply only to the agent at the time they are deliberating, and not to future time-slices of the agent. Admittedly, even the sophisticated approach assumes that an agent always chooses in accord with their preferences at the time. This in itself might be considered a kind of idealisation, as in some cases the agent may not be psychologically able to respect their own preferences. Perhaps McClennen looks upon Ulysses' plight in this way; upon reaching the island of the sirens, Ulysses' weakness of will prevents him from choosing in accord with his preferences, rather than it being the case that his preferences actually change.[29] In the case of the former, both the resolute and the sophisticated approaches must handle Ulysses' problem in the same way, if they have any hope of giving him the right advice—they must regard sailing straight past the sirens as a psychological impossibility, and thus not an available option. But the possibility remains that Ulysses may not in fact predict a psychological barrier of this kind; he might predict, rather, an actual preference change. We do not want our decision model to be silent in these scenarios. So if the resolute approach cannot help Ulysses or the potential addict when their preferences are expected to actually change, then we should reject such a model of sequential choice.

McClennen's second version of resolute choice does not require an agent to blindly stick to a strategy, whether or not they are capable of doing so. Moreover, the agent expects to act resolutely only in those cases where it is both possible and beneficial with respect to their preferences at all times. Figure 1-3 above presents an example of such a decision scenario. Even these limited expressions of

[29] Jeffrey (1974) explores this issue as to whether weakness of will prevents some apparent options from being actual options (and preferences remain constant), or whether weakness of will is just a case of unplanned preference change. He concludes that there is simply a tension between these possibilities.

resoluteness, however, are rather mysterious if one takes the view (as I do) that an agent can only ever act on their current preferences. The agent's preference function in Figure 1-3 is, by assumption, stable. This means that if the agent reaches the second choice node, they will prefer O_1 to O_2. Why then would the agent resolutely opt for O_2? I am not sure how to make sense of an agent having a given set of preferences when it comes to the prospects before them, but then choosing at odds with these preferences out of loyalty to some pre-determined plan. Of course, the agent may be the sort of person who gives considerable weight to honouring previous commitments, but such integrity concerns should be reflected in the agent's *current* preference function. I discuss this important role that past resolutions can play in the next section. The point remains that it is almost a matter of definition that at each point in the decision tree, an agent can only act on their current preferences (if we ignore the aforementioned possibility of failure to properly execute choices).

1.5 How sophisticated choice can benefit from "resolute" considerations

While I am critical of the resolute approach and its insistence that we should honour past choices of strategy, there are some compelling aspects of this account of dynamic choice that seem to ring true to many people. To put it bluntly, the past often does matter to decision makers. And the resolute, as opposed to the sophisticated approach, is perceived to capture this important aspect of sequential choice. For example, surely the fact that I today solemnly resolve to watch the sunrise each morning hereafter should be thought to have some bearing on the choices that I make early tomorrow morning when the alarm clock sounds. At that very early hour, my complete disinterest in the waking world may well end up winning the day, but importantly, I expect my resolution to have *some* impact, otherwise I wouldn't bother resolving to do anything at all. There are surely many examples of how past turns of events, as well as past choices, can legitimately affect how an agent assesses their current options. It seems reasonable to suppose that the agent facing Allais's example problem in Figure 1-2, for instance, may well be affected by the risks that they have

already borne when making their choices in each of the situations A and B.

These sorts of considerations seem to have led a number of theorists to express a kind of permissive attitude towards resolute choice. The idea here is that resolute strategizing should not be mandatory, especially considering it is not feasible in Ulysses' or the potential addict's case, but it is at least permissible that choice be affected by past aspects of the decision process. I have already mentioned McClennen's later version of resolute choice that recommends an agent stick to a planned strategy only when it "furthers the preferences of the agent at all times". Rabinowicz (1995) offers an even more permissive position with respect to the resolute approach; his is a hybrid theory of sequential choice that contains elements of both the sophisticated and resolute methods. In line with sophisticated choice, Rabinowicz recommends that an agent work out what they will choose at future choice nodes, so as not to make the mistake of setting off on a path that is by their own predictions non-viable. In accordance with the resolute approach, however, he allows that future choices may be affected by parts of the decision tree that are no longer live options. In effect, Rabinowicz rejects the claim that an agent must treat a choice node as if it were the beginning of a new, stand-alone decision problem. He thus agrees with the resolute approach insofar as it permits the past structure of a decision tree to influence present choice.

I agree with the basic claim here that the past can matter to a decision maker, but I do not think this necessitates a hybrid sophisticated/resolute approach to sequential choice. I stand by my claim that any appeal to resolute choice, whether this strategy is regarded as mandatory or merely permissible, does not respect a basic requirement that an agent's choices should depend only on their current preference function. The trick to accounting for the past, I claim, is rather a matter of describing outcomes in sufficient detail, and recognizing that the past affects an agent's *current preferences*. I will start with an obvious example and then move to the more difficult cases defenders of the resolute approach have in mind. We all recognise that whether or not an agent will now choose to have coffee or a pasta dish will likely depend on past happenings—whether the agent has already eaten their fill, for instance. We are not

here inclined to say, however, that the agent's choice should take into account their past preferences/choices. What is going on here is that the agent prefers coffee to the pasta dish under the condition that they are rather full already, and they prefer the pasta dish to coffee otherwise. Whether or not a particular condition is fulfilled at a particular point in the decision tree of course depends on past occurrences and choices, but once an agent has used this information to predict what their preferences over final outcomes will be at the choice node in question, there is no need to reference the past details of the decision problem.[30] What determines choice at any point are just the beliefs and the preferences, or the probability and the utility function, of the agent at that time.

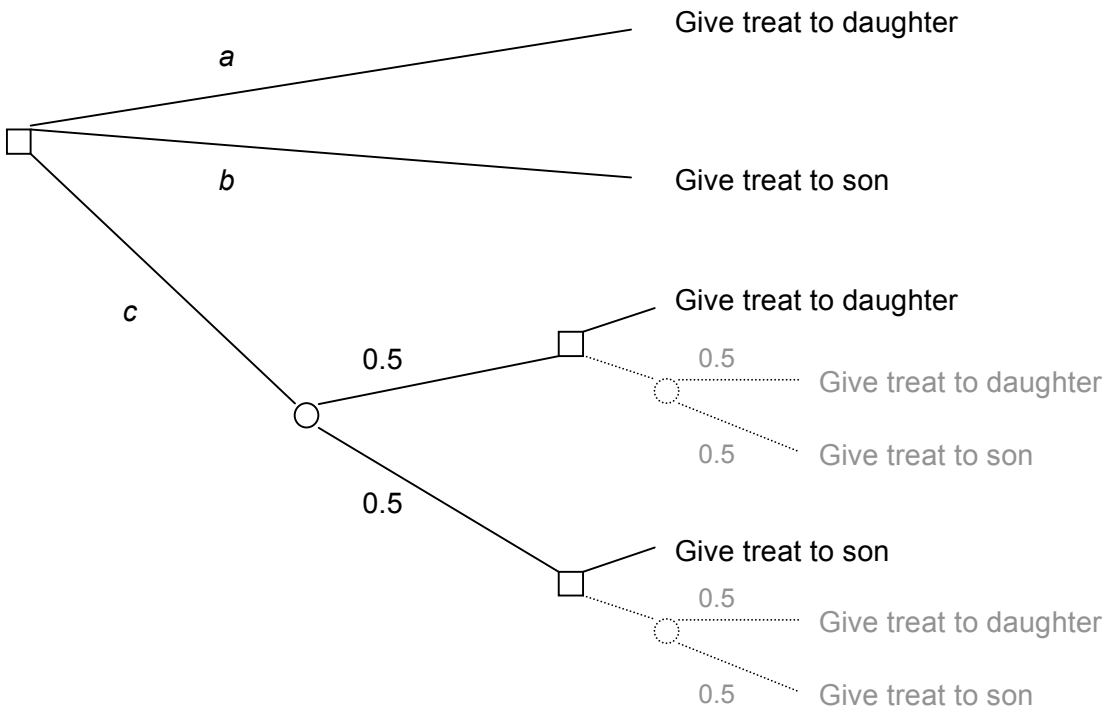
So the past matters, but only insofar as it affects current beliefs and preferences. This means that a sophisticated agent cannot ignore the details of a decision tree when they are determining their choices at later nodes. They must consider what events would have taken place along the path in question in order to predict what their belief and preference functions will look like at a particular node. The difference between sophisticated and resolute choice is that the former holds that an agent's preference function and the options left open to them at a particular time completely determine the choice at hand, whereas the latter allows that the agent's choice may be further influenced by the past structure of the decision tree. I maintain that not only is sophisticated choice the more plausible model for sequential decision-making, it can moreover account for all our reasonable intuitions about how the past affects choice. As a further example, consider the case where I resolve to wake up at sunrise: it is reasonable to think that my early-morning preference function, given that I make the resolution to wake up early, will likely tend more in favour of jumping out of bed than what it would have been had I not made the resolution at all, if I place any importance at all on sticking to my word.

Let me now address an example that Machina (1989, pp. 1643–44) uses to defend the

[30] We can say that the agent's preference for pasta dish/coffee changes, depending on whether the condition that they are full is satisfied. This is equivalent to distinguishing between the outcomes "pasta dish on a full stomach", "pasta dish on an empty stomach", etc.

resolute approach's attention to the past structure of a decision tree.[31] The scenario is one where a mother has a single treat that she can give to either her daughter or her son. She is indifferent between giving the treat directly to one child or the other. The mother strictly prefers, however, a 50/50 mixture of these acts to either of the pure acts (which involve giving the treat directly to one of her children). According to the depiction of the problem in Figure 1-4 (i.e. the way the outcomes are described), the mother's preferences show a violation of the independence axiom of SEU theory. (The mother prefers option *c* to either *a* or *b*, and according to the independence axiom, a mixture of two options should never be preferred to both these options.)

Figure 1-4



Machina claims that the violation of independence is well motivated in this case. Moreover, the example seems to present a good case for resolute choice. If the mother pursues option *c*, then she will be happy to give the treat to her daughter should the

[31] McClennen (1990, p. 208) runs essentially the same argument as Machina using a similar example.

odds go in her favour, and otherwise happy to give the treat to her son. Machina argues that the mother would not, after the first coin toss, opt for a further coin toss (represented in Figure 1-4 in grey), even though she prefers the mixed outcome to either of the pure outcomes. *This is because the past structure of the decision tree has a bearing on the mother's choices.* The idea is that even though the mother's preferences remain stable, the choice she makes at the original node and what she would be inclined to choose after the coin has already been tossed are not identical.

While Machina's example is intended to motivate both violations of independence and the mother's preference for tossing the coin only once, I don't think this combination of choices can be supported, because I don't think the resolute approach holds up. If the mother's preferences really violate independence as specified, and also remain stable with time, then if she were given the chance to toss the coin again she would take this option, rather than award the treat directly to the first winner. My position here simply reflects my chief criticism of the resolute approach. The argument is an analytic one—our very understanding of the terms "choice" and "preference" hinges on the direct relationship between the two; an agent's choice amongst live options is completely determined by their preference function over possible outcomes at the time in question, and cannot be influenced by the purely formal, past structure of the decision problem.

The best way to respond to Machina's prime case for the necessity of resolute choice, perhaps, is to highlight the fact that there is an alternative way to depict the mother's predicament; a way that accords with our intuitions that the mother really would prefer to toss a coin (once and once only) in order to award the treat, rather than give it to one of the children directly. Joyce (1999, p. 54) in fact uses this example as an illustration of why it is important to describe outcomes in all their relevant detail. The idea is that the outcomes in option *c* are not identical to the outcomes in *a* or *b*. Option *c* results in either "treat awarded to daughter via an obviously fair process" or "treat awarded to son via an obviously fair process". Options *a* and *b*, on the other hand, result in the treat being awarded to one or other of the children via an unfair process. In this case there is no failure of independence, and no need to appeal to resolute

choice to explain why the past matters, and consequently why the mother prefers to toss the coin just once.

A similar analysis can be given of Allais's problem, as depicted in Figure 1-2. I maintain that even an independence-violator will not choose options *a* and *c*, if the problem has the sequential form given.[32] Under the assumption that the agent's preference function remains stable (whether or not this preference function obeys independence), the agent should be expected to choose similarly at the choice node in problems *A* and *B* because the set of live options is identical in each case. That is, the agent will choose *a* and *d*, or *b* and *c*, or else they will be indifferent between both options in the two scenarios.

We need not accept, however, that Figure 1-2 accurately portrays the kind of choice situation that the typical agent faces. The risks already borne may have a real psychological effect on the typical agent, such that they predict that their preferences would not be identical in situations *A* and *B*, or equivalently, that the outcomes are not described in sufficient detail. It might be the case that the agent has already experienced anxiety about the unknown after the first chance node has been resolved in problem *B*, and so they might prefer the lottery to the sure $1 million. Given the lack of suspense in problem *A*, the agent in question might have the opposite preferences. Another way of putting the point is that the outcomes are not identical in the two decision scenarios. Problem *B*, for instance, involves "$1 million cash plus anxiously beating heart", while problem *A* involves "$1 million cash and no anxiety". If this is the case, we need not look further (i.e. to independence violation combined with resolute choice) in order to explain how a rational agent could choose the sure prize in problem *A* and the lottery in problem *B*.

---

[32] Note that there are other ways to construe Allais's problem in normal/sequential form that may well recommend options *a* and *c*, if the agent's choice function involves violations of independence. That is to say, I am not here trying to show that there is no way the Allais choices (*a* and *c*) can be explained by a theory that relaxes independence. (In fact, I consider this possibility in detail in Chapter 4.) It is just that this particular sequential representation does not allow an independence-violating theory to rationalise the Allais-choices.

The basic point being made here is that any aspect of the world that affects an agent's preferences should be included in the description of outcomes. (This is by no means a new claim. Joyce (1999, p. 52), for instance, expresses this principle in formal terms.[33]) In terms of my discussion about sequential choice, what is important is that past conditions can influence prospective outcomes in subtle ways, say by affecting an agent's psychology. The treat example and the sequential version of Allais's problem are cases in point. When outcomes are described in sufficient detail, so that the influence of past happenings on both the outside world and an agent's psychology is recognised, then sophisticated choice can accord with all our reasonable intuitions about choosing strategies. When it comes to the treat and Allais examples, it also so happens that choice behaviour that appears to violate independence can actually be reconciled with SEU theory when the outcomes are described differently. This on its own says nothing conclusive, however, about the plausibility of decision theories that relax independence.

I want to stress that it is important not to take the message about adequately describing outcomes too far. There should still be the possibility of an agent having irrational preferences. In other words, we shouldn't just re-describe act outcomes whenever preferences don't conform to the appropriate axioms, or whenever an agent seems to choose in a non-sophisticated sort of manner. I take up this issue of what is the appropriate way to describe outcomes in Chapter 4, when I examine the independence axiom of SEU theory in more detail. For now, let me just say that we don't want to trivialise the distinctions between SEU theory and its competitors, or between the sophisticated and resolute approaches to sequential choice, by allowing outcomes to reference purely formal aspects of the decision model.

---

[33] Joyce (1999, p. 52) states that "an outcome O is underspecified any time there is a possible circumstance C such that the agent would prefer having O in the presence of C to having O in C's absence. Whenever there is such a circumstance, O must be replaced by two more specific outcomes $O_1 = (O \,\&\, C)$ and $O_2 = (O \,\&\, \sim C)$".

1.6 Conclusions

In this chapter I have sought to address some key puzzles regarding sequential (or dynamic) choice. In particular, I have been concerned to reconcile sequential choice with the standard (static) Savage/Jeffrey-type decision problem. I have argued that the sequential and static models are best conceived as being alternative representations of the one decision problem, rather than two essentially different types of decision problem. The important thing to recognise is that the static and sequential models do not give the agent conflicting recommendations regarding what they should do *now*. Furthermore, I hold that an agent's choice of action at a given time should be sensitive to whatever the agent's best predictions may be with respect to their future behaviour. In other words, any satisfactory approach to sequential choice must respect the choice functions that an agent expects to hold at pertinent times in the future, whatever these choice functions turn out to be, and however disconnected they are from the agent's current beliefs and preferences. I have argued that only the sophisticated approach fits this bill. Both the naïve and resolute sequential-choice approaches involve unwarranted idealisations about an agent's future self. In particular, they restrict the kinds of belief and preference changes that the agent may expect to undergo. While the sophisticated approach is generally thought to lead to differences between static (normal form) and sequential (extensive form) decision solutions, I hold that this is simply a misleading way of looking at things. The sophisticated approach is the right way to unpack decision dynamics, and thus normal-form decision models should be brought into line with sophisticated findings.

While some find the resolute approach's attention to the past appealing, if outcomes are described in sufficient detail, the sophisticated approach is sensitive to how the past affects current choices. Indeed, only sophisticated choice gives a credible account of the influence of the past—past events inform an agent's current preference function and/or the character of final outcomes. The resolute approach, by contrast, involves an agent intentionally choosing at odds with their current all-things-considered preferences. This does not fit with any reasonable definition of "preference"; it is certainly at variance with the way preference is employed by Savage or Jeffrey. I

ended with a cautionary note about how far the concept of all-things-considered preferences can be pushed. As mentioned, this is an issue that will be treated in much more detail in Chapter 4. The problem is that while we want to be maximally permissive about what properties affect an agent's preferences, we do not want overly extensive outcome descriptions to artificially remove the distinctions between SEU theory and its rivals, or dilute the merits of sophisticated over resolute choice.

2 THE PRAGMATICS OF BELIEF

2.1 Introduction

In the previous chapter, I argued for a particular approach to sequential decision-making—the sophisticated approach. At least part of the reason for supporting sophisticated choice is that it is very permissive with respect to the beliefs and preferences that an agent might expect to have in the future. But one may well wonder whether this permissiveness with respect to how an agent may conceive of their future self is more of a weakness than a strength. For instance, some might hold that there are Bayesian norms governing how an agent should expect to change their beliefs with time. Contrary to my comments in Chapter 1, it might be contended that rationality is a temporally extended phenomenon, and as such, rational choice theory should be concerned with the ideal agent, both now and later. I do not think this is right, but to properly address this issue, it is necessary to consider just what the Bayesian epistemic norms stipulate, and how well any such claims are justified. To this end, I examine what many regard as the best defences of the key Bayesian norms—the Dutch book arguments (DBAs). I will initially examine the synchronic DBA for beliefs conforming to the probability calculus, as a way of setting the scene. Then I turn to the diachronic DBA for (strict) conditionalisation as the rule for updating beliefs,[34] which is more pertinent to my central question in Part I of this thesis: How should an agent account for their future self in decision-making?

The synchronic and diachronic DBAs are pragmatic arguments for the Bayesian

[34] I note in Section 2.7 that some have advanced a DBA for "Jeffrey-conditionalisation", but in this chapter I focus solely on strict conditionalisation.

epistemic norms just mentioned. In broad terms, they supposedly show that violation of these Bayesian norms leads to *unnecessarily* bad outcomes in particular circumstances. And bad outcomes that could have been avoided are argued to be indicative of an epistemic inconsistency.[35] Of course, there may remain other ways to fault an agent's belief set; it is important to be clear about the intended scope of the DBAs. Consider, for instance, the synchronic case. Let's say our agent's beliefs conform to the probability axioms (which I outline in Section 2.3). We call such an agent and belief set *coherent*. Our coherent agent may, however, have credence close to 1 in the flat-earth theory. It is not the role of the synchronic DBA to fault such an agent, however much their belief set is found wanting.[36] Likewise for the diachronic case—our agent may subscribe to the Bayesian belief-updating rule of conditionalisation (which I will outline in Section 2.7), but their beliefs might include some unusual inferences or conditional beliefs. For instance, the agent might hold that if it is true that the glass vase is dropped to the floor in 5 minutes time, it is less likely, rather than more likely, that the vase will soon be broken. Again, the diachronic DBA is not intended to find fault with this agent, however much we might think their conditional beliefs are misguided.

But even if we are careful not to expect more from the DBAs than what they were ever intended to show, there are reasons to be concerned about whether they provide convincing justification for the Bayesian norms of probabilism and belief-update via conditionalisation. In this chapter I will examine in detail the reasoning behind the two arguments to determine just what an agent who does not satisfy the Bayesian epistemic norms in question (a "non-Bayesian" agent we might say) is guilty of. For starters, both the synchronic and the diachronic DBAs rest on substantial assumptions about rational choice (in particular, rational betting) and its relationship with belief. These assumptions deserve investigation. For instance, when it comes to the

---

[35] Skyrms (1987) and Armendt (1993) present the DBAs in this way—they focus on why the bad outcomes that are highlighted by the DBAs are indicative of epistemic inconsistency. Indeed, I draw heavily on the work of both these authors in what follows. Both claim that Ramsey's (1926) intended interpretation of the DBA holds bad pragmatic outcomes to be mere indicators of epistemic inconsistency.

[36] Both Skyrms (1984, pp. 21–26) and Armendt (1993) are careful to point out that even if an agent satisfies the probability axioms, the agent is not beyond reproach.

synchronic DBA, I question whether the equating of degrees of belief with an agent's fair betting quotients requires explanation. The diachronic DBA brings even more questions to the table regarding how an agent should assess strategies. My arguments in Chapter 1 will be pertinent here. It is even less obvious in the dynamic context, as compared to the static, just what sort of sure losses are indicative of an epistemic inconsistency. Moreover, the rule of conditionalisation can be interpreted in two different ways, and we need to determine whether either or both of these interpretations can be supported by a diachronic DBA. But I will come back to the diachronic DBA in Sections 2.7 onwards. The first part of this chapter focuses entirely on the synchronic DBA for probabilism—the position that rational beliefs conform to the probability calculus.[37]

2.2 The synchronic Dutch book argument

At the base of the synchronic Dutch book (DB) story is the claim that an agent's degrees of belief can be directly ascertained from their fair betting odds.[38] The concept of a "fair bet" is very important here; it is a bet that does not confer any advantage to either the buyer or the seller. (Alternatively, it is the maximum price at which the agent would buy a bet, and also the minimum price at which they would sell the bet.[39]) It will be useful to consider the basic betting set-up in a little more detail than usual. Consider a bet on Q that pays \$1 if Q is true and \$0 if Q is false. Assume that the agent is willing to pay anything up to \$$p$ for this bet; for this maximal buying price the agent's net winnings will be \$$(1 - p)$ if Q is true and -\$$p$ if Q is false. Importantly, in the Dutch book story, the bet could be run with any stake S, and we

---

[37] Hájek and Ericksson (forthcoming) define "probabilism" in this way.

[38] Armendt (1993, p. 3) says, "the fair betting quotients are supposed to *be* the agent's degrees of belief".

[39] This is assuming that a rational agent's credences are "sharp/determinate" as opposed to "vague/indeterminate". Levi (e.g. 1986) relaxes this assumption in order to permit indeterminate belief, and in this case, the agent's maximum buying price may be less than their minimum selling price for a bet on some proposition. I discuss Levi's treatment of indeterminate belief in Chapter 5.

assume that, if the agent is willing to pay up to $\$p$ for a bet on Q that pays \$1 if Q is true and \$0 otherwise, then in general they will be willing to pay up to $(p \times S)$ for a bet on Q that pays any stake S if Q is true and nothing otherwise. At their maximal buying price, the agent's net winnings if Q turns out to be true amount to $(1 - p) \times S$, and if Q turns out to be false, the net winnings are $(-p \times S)$. This is essentially how the betting game featured in the synchronic DBA operates. For the case just described, we would say that the agent's fair betting odds for a bet on proposition Q are $p{:}1 - p$ (where p is the agent's "fair betting quotient"). It is important to the DB story that the agent will accept any bets that they deem fair or favourable.[40]

I have dwelt a little on the set-up of the standard betting game, because this is crucial to the Dutch book story. If an agent's degrees of belief are supposed to be equivalent to their fair betting odds, it is important to be clear about what fair betting odds are, and whether they have the right kind of relationship with credence. In fact, I don't think it is at all obvious how an agent's credences can be inferred from their judgments about fair bets, short of already assuming what the DBA is supposed to show—that credences obey the probability calculus. To illustrate, I will consider an example in which the stakes are fixed at \$1. In this particular case, let's say the agent is willing to pay up to 30c for a bet on Q. We know that this means the agent will receive a minimum of \$1 – 30c (= 70c) if Q is true, and they will lose the amount that they paid—a maximum of 30c—if Q is false. What does this say about the agent's degrees of belief in Q? To repeat, what does the agent's acceptance of a 70c win versus a 30c loss on Q tell us about the agent's credence in this proposition? The fact that the agent's betting quotient is supposedly invariant under different stakes turns out to be very important to the identification of belief with fair betting odds. But I will return to this puzzling aspect of the DB story later. Let me first recount the complete Dutch book argument, so we know where the appeal to fair betting quotients is heading.

---

[40] Some tell the synchronic Dutch book story in such a way that the agent is the bookie. That is, the agent sets the fair betting odds for all propositions, and the cunning bettor can make whatever bets they please at any stakes. Here I tend to depict the betting situation in the opposite way, but it makes no difference. In my case, the cunning bookie names the bets, and the agent must accept any that are fair or favourable at any stakes (in line with their fair betting quotients).

For now, I will just assume the connection between fair betting odds and degrees of belief. Let me continue with the Dutch book story. A Dutch book is a set of bets, each of which the agent regards as fair or favourable, which collectively guarantees that they suffer a loss. Hájek (2005) is careful to distinguish between the (synchronic) Dutch book theorem (together with converse theorem) and a further Dutch book argument. The theorem states that if an agent's fair betting quotients (credences) do not conform to the probability calculus, then there exists a Dutch book against them. That is, if the agent's fair betting quotients (credences) cannot be represented as a probability function (over a set of propositions assumed to be closed under negation and disjunction), then there is a series of bets, each of which the agent regards as fair or favourable, which taken together guarantees that they suffer a sure loss. A proof of this theorem is given in Appendix 2.[41] Without the converse theorem, this is not a very useful result. The converse theorem states that if an agent's fair betting quotients (credences) *can* be represented as a probability function, then there is *no* series of bets, each of which the agent regards as fair or favourable, that leaves them with a loss no matter how the world turns out. We are ready now for the synchronic Dutch book argument: given that only betting quotients (credences) obeying the probability calculus are shielded from Dutch book losses, on pain of irrationality, an agent's betting quotients (credences) should conform to the probability calculus.

Hájek (2005) draws attention to the fact that the DBA properly concerns any set of bets that the agent will accept, whether the individual bets are fair or better than fair. (Accordingly, in the above account of the DB theorem I was careful to refer to the agent accepting fair or favourable bets.) If we were only interested in bets that are exactly fair, then it is unclear whether having betting quotients that conform to the probability calculus would be a good thing—the incoherent agent would be open to sure loss, but, unlike the coherent agent, they would also be open to sure gain.[42] When

---

[41] It is shown that a Dutch book can be made against any belief set that fails to satisfy one of the three axioms of the probability calculus—non-negativity, normalisation and additivity. (In Section 2.3 of this chapter I state the probability axioms formally).

[42] Hájek (2005, p. 142) refers to a book of fair bets that guarantees the incoherent agent a sure gain a "Good book".

we focus on both fair and favourable bets, however, the incoherent agent does not have any special advantages. Sure gains can also be enjoyed by the coherent agent, given that they will accept any bets that are better than fair.

This brings me to an important general point regarding what the DBA does and does not show. We should not think that sure losses always point to an inconsistency. For instance, there could be a scenario in which an agent is forced to choose the better of two evils, and so takes an option that yields sure loss. Such a choice would not be irrational. The reason sure loss is a "pragmatically defective outcome"[43] in the DB story is that the sure loss is chosen over a *definite* zero or positive gain. The situation is not symmetrical when it comes to sure gains. Of course, if the choice is between a sure zero gain versus a sure positive gain, then a rational agent must take the sure positive gain. But a coherent agent can forgo a sure gain if the circumstances are such that there is an alternative available bet or set of bets that has higher expected value. For example, we don't think an agent irrational if they choose a bet that pays $10 if a fair coin lands heads and -$1 otherwise over a bet that pays a sure $1. In the same way, it is not necessarily irrational for an agent to accept only 4 bets out of a possible 5, even if the 5 bets together yield a sure gain. (The agent might deem the 4 bets to have higher expected value.) In a Dutch-book-type scenario, then, just saying that the agent rejects a sure gain is misleading; we must be clear about what the agent will take instead. It will not be the case that the coherent agent chooses an option that is *certainly* inferior to some other option. In other words, while the coherent agent may forgo sure gains, they will never accept a set of bets that is sure to make them worse off than what they might have been had they rejected some or all of these bets.

While its focus on sure losses is by no means arbitrary, the DBA can be criticised for failing to provide an agent with any constructive advice about how to align their beliefs. Granted the underlying assumptions of the story, (which I will go on to discuss shortly), all that the DBA establishes is that an incoherent belief set is vulnerable to some particular problems, and that coherent belief sets are not

[43] Armendt (1992 & 1993, p. 3) refers to the sure loss associated with a Dutch book in general terms as a "pragmatically defective outcome".

vulnerable to these same problems. Of course, as mentioned earlier, the best accounts of the DBA do not pretend that it is the last word on what constitutes a good belief set. The idea is that an agent should never set him/herself up for a sure loss that could have been avoided, and that such sure losses illustrate a fundamental epistemic inconsistency: when an agent evaluates the same propositional outcome differently, depending on how it is presented.[44] But the fact that the DBA is intended to show up only a certain kind of fault in an agent's belief set (admittedly, an *a priori* kind of fault) should not be taken lightly. It turns out that in some scenarios, an incoherent agent will do better than a coherent one, if their beliefs better match the objective probabilities at hand.[45]

Consider the following two agents, who are each offered a couple of bets—one on whether a red card will be drawn from a pack of well-shuffled cards, and the other on whether a black card will be drawn from the same pack. Agent 1 is incoherent, having Pr(red) = 0.6 and Pr(black) = 0.6. Agent 2, on the other hand, is perfectly coherent, with Pr(red) = 0.9 and Pr(black) = 0.1. The bets that are offered have terms shown in Figure 2-1.

Figure 2-1

	PRICE	STAKES
BET 1 ON RED	50c	$1
BET 2 ON BLACK	40c	$1

Clearly an agent who takes both bets will win a sure 10c. (The agent pays only 90c for the bets, and they are guaranteed $1 in return.) An agent who happens to have the objectively correct credences, i.e. Pr(red) = Pr(black) = 0.5, will regard the first bet as

[44] Skyrms (1987) and Armendt (1993) claim that Dutch book sure losses are indicative of a "divided mind"—such losses indicate that the agent is evaluating the very same proposition in different ways.

[45] I am simply assuming here that there is some suitable account of objective probabilities.

fair and the second favourable, which means they will indeed take both bets. Agent 1 will also take both bets, because they regard each of the buying prices as favourable. Agent 2, on the other hand, will only accept the first bet. They will think it wiser to accept just the one bet, because it has an expected value of $0.9 \times 50c - 0.1 \times 50c = 40c$, which is better than 10c. But if we look at the relevant objective probabilities, we see that Agent 2 is simply wrong in their assessment that taking Bet 1 alone has greater expected value than taking both bets. The true expected value of Bet 1 is zero, which is not as good as a sure 10c.

In the above betting situation, the incoherent agent is better placed than the coherent agent. Worse still, short of knowing the relevant objective probabilities, it is not clear how the incoherent agent can adjust their beliefs in line with the probability calculus, and still come out ahead with respect to this two-bet scenario. If the agent was to change only one of their credence values, then they had better be careful which one. Pr(red) = 0.6 and Pr(black) = 0.4 will give them the sure 10c, but Pr(red) = 0.4 and Pr(black) = 0.6 will not. It is not a case of the incoherent belief function being *clearly* dominated by any nearby coherent one. In the end, pragmatic considerations leave us in the lurch—an incoherent belief function has certain failings, but it will also have certain advantages over coherent belief functions that are not aligned with the "true" probabilities. And if faithfulness to the "true" probabilities is thought to be important,[46] it is not clear which way the incoherent agent should turn. Indeed only a coherent agent whose beliefs exactly match the objective probabilities will in all cases be better off than the incoherent agent. I regard this as a serious challenge to the usefulness of the line of thought underlying the DBA—that whatever we might think about the world, we should at least ensure that our beliefs are coherent. But perhaps the argument was never meant to provide direction to would-be rational agents. In any case, I will move on now to ways we might challenge even what the DBA *does* purport to show.

---

[46] Some Bayesians/Probabilists might hold that coherence is the only criterion for good partial belief. But I think the majority would agree that it is desirable for degrees of belief to match the way the world is, and specifically, it is desirable for degrees of belief to match objective chances, if there is any sense to be made of objective chance.

2.3 Does the synchronic DBA beg the question?

Some have argued that the synchronic DBA is not convincing because it is not at all realistic: it is simply false that there are always Dutch bookies waiting to exploit incoherent agents. But pointing this out is not a sufficient criticism of the DBA. Of course, conniving bookies are not ubiquitous enough to pose a threat to incoherent agents, and even if they were, they wouldn't be able to force an agent to place the required bets. Concerns of this sort are, I think, beside the point; fictional stories often help to illustrate norms. Moreover, when it comes to this particular story, if Dutch books really are so bad, then the mere possibility of making one against an agent is enough to cast doubt on the agent's belief set, no matter how idealised the betting circumstances are. In any case, I think the Dutch book narrative can be challenged in more substantial ways, and it is to these challenges that I now turn. The argument depends on some rather strong assumptions about an agent's betting behaviour, and how bets made in isolation should relate to one another. The synchronic DBA is not just a story about the threat of sure loss; it is rather a story about the potential for sure loss under very specific circumstances (that go beyond the mere existence of cunning bookies who are capable of forcing us to accept bets we see as fair or favourable). The question is how much weight we should give to any outcomes associated with betting under these special conditions.

I return to the issue of identifying degrees of belief with fair betting odds. My initial worry, hinted at earlier, is that this very relationship appeals to expected value calculations, and these calculations presuppose that credences obey the probability calculus. Consider the bet that our agent earlier was willing to place on Q—one that pays 70c if Q is true and -30c if Q is false. Assume that 30c is the agent's maximum buying price, or, equivalently, that the agent is willing to take either side of the bet (they regard it a fair bet). What does this tell us about the agent's degree of belief in Q? (Granted that belief is somehow mixed up with choice, if we were just looking at this isolated betting scenario, the agent's degree of belief in Q could well be 30/70 or 70/30 or some other manipulation of these numbers.) It appears that we can only work

out the agent's credence in Q if we know *why* they thought that paying 30c for a \$1 bet on Q was fair. If the agent must have a reason for thinking the bet fair, then surely it will be because the expected value of possible earnings is zero. So when the agent pronounces the bet fair, what they are really saying is

$$\Pr(Q) \times 70\text{c} - \Pr(\sim Q) \times 30\text{c} = 0, \text{ where } \Pr(Q) \text{ is the agent's credence in } Q.$$

We can solve for this credence if we assume that $\Pr(\sim Q) = 1 - \Pr(Q)$. Then

$$\Pr(Q) = 30\text{c}/(30\text{c} + 70\text{c}) = 0.3$$

Now let us return to the general scenario where the agent names their fair betting quotient and the stake can be any size. Recall that odds of p:1–p on proposition Q indicate that the bet pays $(1–p) \times S$ if Q turns out to be true and $–p \times S$ if Q is false. Again we might seek to explain the relationship between fair betting quotients and degrees of belief. When $\Pr(Q) = p$, a bet on Q at odds p:1–p is fair because we have:

$$\begin{aligned}\text{Expected value of bet} &= \Pr(Q) \times (1 - p) \times S - (1 - \Pr(Q)) \times p \times S \\ &= p \times (1 - p) \times S - (1 - p) \times p \times S \\ &= 0\end{aligned}$$

What I draw attention to is that, insofar as we need to know *why* an agent pronounces a particular bet fair, we must appeal to expected value calculations. But such calculations introduce some assumptions about the nature of credences.

The synchronic DBA will involve a significant amount of question-begging if an agent must explain why their fair betting quotients are one set of values rather than another. Clearly some probabilistic apparatus is assumed in order to do the necessary expected value calculations; this would mean that even before we start talking about Dutch books, a couple of assumptions are introduced regarding the probabilistic nature of partial belief. In particular, it must be presupposed that an agent's credence in the proposition Q—$\Pr(Q)$—obeys the following rule: $\Pr(Q) + \Pr(\sim Q) = 1$. Sobel (1987, p. 58) in fact refers to this rule as the "complement condition", and suggests that it is a fairly innocuous assumption about partial belief within the DB story. The "complement condition" doesn't seem too innocuous to me, however, given that the context is one in which we are trying to prove that degrees of belief should satisfy the

probability calculus.[47] Indeed, the "complement condition" is suggestive of credences satisfying all the key probability axioms[48]:

1. Non-negativity: $\Pr(X) \geq 0$ for all X in **S**.
2. Normalisation: $\Pr(T) = 1$ for any tautology T in **S**.
3. Additivity: $\Pr(X \vee Y) = \Pr(X) + \Pr(Y)$ for all X, Y in **S** such that X is incompatible with Y.

(Here Pr() is a probability function over a non-empty set of sentences **S** closed under negation and disjunction.)

It is clear that we must seek an alternative way to understand the relationship between betting quotients and degrees of belief. Proponents of the synchronic DBA cannot be appealing to expected value calculations, because this move is question-begging. As a way of moving forward, I note the important difference between examining a single bet in isolation, and looking at the agent's betting quotients for some proposition Q, where these betting quotients are independent of the stakes of the bet. It is hard to infer an agent's degrees of belief in a proposition when we are considering only a single bet. To use the example from before, what does an agent's belief that a bet paying 70c if Q is true and –30c otherwise is fair reveal about their degrees of belief? Not necessarily a great deal. But if we consider the more general framework, the relationship is perhaps more obvious. The Dutch book story stipulates that an agent specifies only their fair betting quotients, and the cunning bookie can name any stakes (perhaps within a reasonable range) for the bets of their choice. The agent is expected to accept any bets deemed fair or favourable in line with their nominated fair betting quotient. The condition that is being pushed here is that the agent's betting quotients are not dependent on the size of the stake. This means that there is some constant that characterises an agent's relationship with any proposition Q—the proportion of the stakes that the agent would be prepared to pay for the bet on Q. It is very plausible

[47] Colyvan (2004) raises similar concerns about Cox's representation theorem, which purports to show that any measure of belief is isomorphic to a probability measure. Colyvan does not just argue that reliance on something akin to the "complement condition" is question-begging, but that it is actually an inappropriate assumption in some domains of discourse where the logical principle of excluded middle fails.

[48] This presentation is directly from Hájek (2005, p. 140).

that this constant has something to do with the agent's credence in *Q*. Of course, there are still infinitely many ways that constant betting quotients and credence might be functionally related. But perhaps the simplest and thus the best theory is that an agent's constant fair betting quotient for *Q* just *is* their degree of belief in *Q*.

2.4 Making explicit the DBA assumptions

I grant then that we need not require an agent to explain *why* the betting quotient they nominate for a proposition *Q* is fair. It is sufficient that these fair betting quotients do not vary with the stakes of the bet. Then we can say that the betting quotient marks a constant relationship between the agent and the proposition *Q*, and plausibly, the best account of this relationship is that it is credence. This brings to the forefront a particularly weighty assumption underlying the synchronic DBA, however, and it is to this assumption that I now turn. It is the very condition that the maximum amount an agent will pay for a bet on some proposition *Q* should depend linearly on the value of the betting stakes. This is no uncontroversial point. In fact, for many people, money has diminishing marginal returns—it is worth less the more of it that one has. This means that while I may be willing to pay $5 for a bet on *Q* at stakes of $10, it is unlikely that I will be willing to pay as much as $5,000 for a bet on *Q* at stakes of $10,000.

There is another related assumption in the DB story that we can similarly challenge. Schick (1986) draws attention to the fact that there are other ways for an agent to avoid sure loss when making a combination of bets; the agent need not have beliefs conforming to the probability calculus (together with well-ordered preferences). The easy route to avoiding sure loss is to be cautious about making collective bets. It may not be the case that the sum of fair bets is collectively fair. Schick gives an example (that admittedly turns on the logic of preference rather than the logic of belief), of an agent who has the following intransitive preferences over prospects (where the prospects in question are *X*, *Y* and *Z*, and $X > Y$ indicates that the agent strictly prefers

X to *Y*):

$X > Y, Y > Z, Z > X$

The agent may well consider the trades below to be individually fair:

1. Pay *X*, receive *Z*
2. Pay *Z*, receive *Y*
3. Pay *Y*, receive $X - \partial$ (for some positive ∂)

The sum of trades 1–3, however, will result in the agent handing over *X* for $X - \partial$. The usual story holds that the cunning bookie can guarantee him/herself a sure ∂ because the agent goes into the deal assuming that the sum of fair bets will itself be fair. But our agent might not be so presumptuous. The agent might simply refuse to make the three trades in tandem. Although the agent may be confused about which of *X*, *Y* or *Z* they would rather have (due to their intransitive preferences), they will at least not be taken for a sure loss if they have any pragmatic sense at all.

Although he goes on to defend the synchronic DBA, Armendt (1993) presents very clearly just what is involved with the value additivity of bets assumption. It affirms the following series of equalities:

V(bet on *Q*) + V(bet on *R*) = V(bet on *Q* + bet on *R*) = V(bet on *Q* or *R*)

Where V here represents the agent's value function (in monetary units)

As Armendt himself notes, his "divided mind" interpretation of the DB story clearly supports the latter inequality: we should evaluate the same outcome in the same way, regardless of how it is presented. This means that the value of a book of bets consisting of a bet on proposition *Q* and a bet on proposition *R* should be equivalent to the value of a single bet on ($Q \vee R$). Failure of the first equality above does not, however, amount to "divided mind inconsistency". Here we are talking about adding the value of bets that are supposedly taken in complete isolation. Any such sum does not correspond to a real betting situation; whenever the agent places more than one bet, we should be talking about the value of a book of numerous bets, rather than the addition of bets that are supposedly made in isolation. Like constant betting quotients, value additivity of isolated bets is, in a sense, a superfluous assumption in the DB

story. Of course, both constant betting quotients and value additivity are necessary assumptions if we want the result that degrees of belief should be probabilities, but an agent who fails to respect them is not necessarily bound to lose money, and they need not be guilty of "divided mind" inconsistency.

2.5 Value additivity

We might say that constant betting quotients and value additivity of bets (I will refer to both as "value additivity") are aspects of the DB story, that, like the very existence of cunning bookies, we should just accept for illustrative purposes. But the two sorts of idealisations are really quite different. Cunning bookies are introduced just to set the scene, we might say. They provide for the *existence* of a Dutch book scenario. Value additivity, on the other hand, provides the very substance of the DBA, and so it is worth taking this sort of assumption seriously. The DB story is not just about avoiding sure loss (even sure loss in fictional betting scenarios). It is about guarding against any sure loss that could hypothetically arise if we assume that an agent's fair betting quotients are invariant under different stakes, and that the sum of individually fair bets is itself fair. We thus have a rather qualified test of whether an agent's belief set is coherent. In an effort to shore up these qualifications, I will proceed to offer some defence of the "value additivity" assumption. But I will also stress that any such defence touches on much broader territory than the synchronic DB story. If we are going to criticise this pragmatic argument for probabilism, then we are wise to keep things in perspective—it is important to consider where any such criticism leaves us in terms of alternative theoretical positions.

I will outline what I think is a promising motivation for value additivity. It requires appeal to a dynamic decision-making scenario in which the agent's beliefs and preferences are stable or unchanging over time. In these circumstances, value additivity allows the agent to make a series of bets, each assessed purely on its own individual merit, in the confidence that a sure loss will not be suffered after a finite

number of transactions. The idea is that an agent whose belief set supports value additivity will reap the rewards of this dynamic decision-making luxury. The agent need not engage in forwards or backwards planning; provided their beliefs/preferences remain stable, they can happily make individually fair or favourable bets in isolation and not worry about the collective result being a sure loss. This, however, is surely only a weak defence of value additivity. It does not seem sufficient basis for the strong interpretation of the DB result—that an incoherent agent exhibits a fundamental ("divided mind") inconsistency. We cannot say that an agent is *irrational* if their credence function does not facilitate easy dynamic decision-making (when beliefs/preferences remain stable). It is rather that the incoherent agent's credence function lacks this particular quality, if indeed it is a quality. All other things being equal, the agent will have some extra work to do (forwards/backwards planning) in the dynamic setting.

When it comes to seeking further defence of value additivity, I return to the point that this constraint on choice is not curious to the DBA justification of probabilism. Recall from the introduction to this thesis that the SEU representation theorems (e.g. those of Savage (1954) and Jeffrey (1983)) also purport to provide a pragmatic defence of probabilism, amongst other things. The theorems show that an agent who satisfies the SEU axioms of rational preference can be represented as an expected utility maximiser with respect to a unique probabilistic belief function and a utility function that is unique up to positive linear transformation. The claim here is that a non-probabilistic belief function will not yield rational preferences. But SEU theory has its own, rather similar version of the "value additivity" assumption—the independence axiom.[49] This axiom appears in various forms in all of the expected utility representation theorems. As stated in the previous chapter, it essentially holds that "a rational agent's preference between (any two acts) A and A^* should not depend on what happens in circumstances where the two yield identical outcomes" (Joyce, 1999, p. 86). It is independence that requires utility to be additive, or in other words requires that an agent's utility function be "linear in the probabilities". This is to say that the

[49] Armendt (1993) notes the close relationship between the DB value-additivity assumption and the independence axiom of SEU theory. My conclusions in this section are similar to those of Armendt.

amount of utility any individual outcome contributes to the total utility of an act should be proportional to the agent's subjective probability for that outcome.[50]

We could say then that any defence of the Dutch book value-additivity constraint will amount to a limited defence of the independence axiom. (It is a defence of independence in domains in which utility is proportional to monetary value.) Conversely, if independence is considered a plausible axiom of rational choice, then value additivity stands on firmer ground. Note that independence does not actually imply value additivity, because the latter concerns monetary units, while the former deals with the more general notion of "utility". But if there is no good case for independence being a constraint on choice, it would be difficult to argue that the value of monetary bets should be additive. In any case, we can say that independence and value additivity reinforce each other to at least *some* extent.

There is clearly much intuitive appeal to the independence axiom. (Indeed, many decision theorists, including Savage (1954), Raiffa (1961) and Broome (1991) suggest that independence is more or less a self-evident constraint on rational preference.) The axiom has come under challenge, however, by those who claim that a rational agent need not have preferences that are "linear in the probabilities"; agents might, for instance, be generally risk-conservative, and so underweight probabilities in their decision-making.[51] It is my opinion that the dynamic choice framework is critical for determining the merits and/or demerits of relaxing the independence constraint on choice. The details of this debate need not concern us here, however; I will take up these issues in Chapter 6. For now, I think it is sufficient to point out that the DBA should not be singled out for dealing in value additivity. The representation-theorem or "decision-theoretic" defences of probabilism rest on a somewhat similar constraint that is also subject to challenge.

---

[50] I will examine the independence axiom of SEU theory in detail in Chapter 4.

[51] Machina (1989, p. 1631) lists a number of such candidate "non-expected" utility theories that have arisen in the decision theory literature. (Recall that I will discuss the role of the independence axiom in Chapter 4.)

2.6 A powerfully simple defence of probabilism?

If we pitch the DBA against the SEU representation theorems, then there is a case to be made for the former being the simpler and thus the better story. The two arguments for beliefs conforming to the probability calculus involve similar kinds of preference constraints, but for the DBA, we could say that the constraints are more limited in scope. Some perceive this in a negative light and view the DB story as overly simplistic. After all, the representation theorems set out their assumptions clearly, and aim to justify a general (additive) measure of value that is not tied to any particular commodity. The DBA, on the other hand, depends uncomfortably on monetary bets, when, as mentioned, it is commonly accepted that money has diminishing marginal returns. This latter fact certainly casts doubt on both the relationship between credence and fair betting quotients, and the value additivity of bets. On the other hand, we could say that it is to the DBA's advantage that it deals with only a modest domain of choice—monetary bets. It is arguably better to introduce fewer or narrower assumptions about rational choice, if we just want to prove a result about the nature of rational credence. The representation theorems are very ambitious—they require an agent to have an infinitely rich preference ordering that everywhere satisfies the SEU axioms, including independence. Of course, it is no good trying to get by with less grandiose assumptions if these assumptions have no plausibility. For the Dutch book story to get off the ground, however, arguably all we require is that there is *some* betting domain in which it is plausible that betting quotients are constant and the value of bets is additive. Perhaps this domain is very narrow—value additivity might only hold for stakes between, say, $0 and $100. I don't think this means that the DBA only says something about rational belief within this limited domain. After all, we will continue to assume that an agent has some particular epistemic relationship with a given proposition. It is just that the 0–$100 betting domain, for instance, gives us a sufficiently controlled choice experiment whereby we can isolate degrees of belief.

I do not then regard the DBA as an overly simplistic argument for probabilism that is

merely useful for pedagogic purposes; in fact, I think its relative simplicity makes the synchronic DBA usefully different from the SEU representation theorems. I still have some general concerns, however, about these pragmatic defences of probabilism.[52] Recall my claim in Section 2.2 that the DBA is not a constructive argument—it does not direct the incoherent agent to an all-round pragmatically better credence function. (A similar point can be made about the representation theorems, but I will not explore the issue here.) Under conditions in which the objective probabilities are not known, neither coherent nor incoherent agents are clearly better placed when it comes to hypothetical betting outcomes. It is just that the incoherent agent can suffer *sure* loss in some circumstances, while the coherent agent whose beliefs do not match the objective probabilities is vulnerable only to *expected* losses. Importantly, there will be cases in which the incoherent agent will suffer less *expected* loss than the coherent agent. In short, only the objectively correct credence function[53] is superior to an incoherent belief set in every imagined betting scenario.

The other general worry about both the synchronic DBA and the SEU representation theorems is their dependence, as discussed in the previous section, on some kind of value additivity constraint. When it comes to the representation theorems, while there are some strong arguments for independence (particularly when we look to the sequential-choice setting), there will always be room to dispute the status of this axiom.[54] Likewise, the more specific money-based value-additivity assumption can never be defended beyond all doubt. Of course, as Armendt (1993) points out, we are not going to derive any substantial results about partial belief from a pragmatic story unless we start off with some substantial assumptions about how an agent values options. But in appealing to pragmatic arguments, it looks like we are trying to make a claim about the logic of partial belief by appeal to an axiom of preference, or a standard for betting behaviour, whose status is no more secure. In other words,

---

[52] The following comments gesture towards a non-pragmatic justification of probabilism (in the vein of Joyce 1998), or otherwise, acceptance of the probabilistic nature of rational belief as a brute fact.

[53] As noted earlier, I am assuming here that in some cases there will be objective probabilities relevant to a decision situation.

[54] Recall that I will engage with this debate in Chapter 6.

pragmatic justifications of probabilism appeal to the fact that preferences must satisfy independence, or that bets (in at least some domain) should satisfy value additivity, in order to show that belief must conform to the probability calculus. I suggest, however, that the latter may well be the more self-evident norm of rationality.

2.7 Introducing the diachronic Dutch book argument

I will now move on to the diachronic Dutch book argument for conditionalisation as the rule for updating beliefs. This argument is more pertinent to my chief interest here as to how an agent should account for their future attitudes in decision-making. The synchronic DBA for probabilism is somewhat of a side issue, but it does prepare the ground for investigating the diachronic argument, and in any case, the significance of pragmatic defences of probabilism is an important background question that comes up again in Part II of this thesis. My parting comment about the synchronic DBA is the suggestion that a non-pragmatic justification of probabilism might be more compelling. But as discussed, this is not to say that the synchronic DBA does not have merit. To the extent that its assumptions about betting behaviour seem reasonable in at least some betting domains, the synchronic DBA provides some defence for probabilism. While the two sorts of Dutch book argument are really very different, I end up drawing similar kinds of conclusions in the diachronic case. I should qualify here—there is not really one well-accepted diachronic DBA. There are, rather, different formulations of the diachronic DBA, and I will argue that the standard formulations are invalid, or else support epistemic rules that do not seem right. Here I will defend a particular formulation of the argument that I refer to as the sophisticated version. But while I think the sophisticated diachronic DBA provides some defence for conditionalisation (suitably interpreted), at the end of the chapter I question whether this updating rule might be better justified in a more straightforward, non-pragmatic way.

Let us begin by considering the rule that the diachronic DBA is supposed to defend—

(strict) conditionalisation. The controversies surrounding the diachronic DBA could be said to start here (or end here, depending on which way you want to look at it.) The problem is that there are different ways to interpret the demands of conditionalisation. (Compare with the synchronic DBA: it is fairly clear what it is supposed to show—that beliefs at a time should conform to the probability axioms). I will start with a basic outline of conditionalisation before drawing attention to its nuances. The rule specifies how some evidential input E should affect an agent's credence function. That is, conditionalisation specifies the relationship between an agent's initial credence function $\Pr_{initial}$, and their posterior credence function $\Pr_{final}$, if they were to become certain of the proposition E. This relationship accords with Bayes' rule, and can be stated as follows:

$$\Pr_{final}(Q) = \Pr_{initial}(Q|E) = [\Pr_{initial}(E|Q) \times \Pr_{initial}(Q)] / \Pr_{initial}(E)$$

We assume that the initial credence function obeys the probability calculus. (In such case the final credence function will also obey the probability calculus because it is equivalent to the initial credence function conditional on E.) Note that there is a modified belief-update rule—"Jeffrey-conditionalisation"—that handles cases in which the evidence is probabilistic over some partitioning of the possibility space, rather than known for certain. Some have even advanced a diachronic DBA for "Jeffrey-conditionalisation".[55] In this chapter, however, I focus solely on strict conditionalisation.

The presentation of conditionalisation just given suggests that it dictates how an agent's beliefs should actually change in real time. Given a particular credence function at time t_1, it seems that the rule specifies what the agent's credence function should be at the later time t_2, if everything the agent has come to know in the interim time period can be summarised by the evidence proposition E. But we need not interpret the demands of conditionalisation in this strong sense. Rather than dictating how an agent's beliefs should actually change in real time, the rule can be understood to merely concern an agent's premeditated *plans* for updating their beliefs. An agent can be said to satisfy conditionalisation if they plan to update in line with their

[55] See Jeffrey (1983) for this modified rule of conditionalisation. Armendt (1980) and Skyrms (1987) give diachronic Dutch book arguments for "Jeffrey-conditionalisation".

conditional probabilities.[56] Whether or not they actually do update their beliefs in this way might be regarded as another matter altogether. Clearly this will be an important thing to question about any proposed version of the diachronic DBA: does it justify the strong or the weak interpretation of conditionalisation? In fact, in the process of examining what is a plausible version of the diachronic DBA, we will become better acquainted with conditionalisation; in particular, we will come to see which interpretation(s) of the rule can be defended.

There is another epistemic rule that bears some relationship to conditionalisation—the rule of "reflection". There is even more controversy surrounding the latter rule. In brief, reflection holds that current beliefs should match the beliefs one expects to hold in the future. Consider an agent who currently has a credence function represented by $\Pr_1$. The function $\Pr_2$ represents what their credence function could be at some later time. An agent obeys "reflection" if for any proposition Q in the set of sentences **S**,

$$\Pr_1(Q \mid \Pr_2(Q) = x) = x$$

The above is a constraint on the agent's conditional probabilities.[57] If the agent assigns varying probabilities to what their credence function will actually be at time t_2, then "reflection" requires that their current degree of belief in Q equal the *expectation* of their possible future degrees of belief in Q. It can be shown that in most cases, if an agent expects that they will satisfy conditionalisation in the strong sense, i.e. if the agent thinks they will actually update their beliefs in accordance with their conditional probabilities, then they will satisfy reflection.[58] We might think then that conditionalisation and reflection are interlocked, in the sense that a defence of one will serve as a defence of the other. But the two principles clearly focus on different aspects of belief, and again, investigating the diachronic DBA will help us to understand the distinctions between them.

---

[56] This is the interpretation of conditionalisation that Skyrms (1987, 1993) and Armendt (1992) focus on.

[57] The "reflection principle" is so-named and defended by van Fraassen (1984). Goldstein (1983) presents a similar theorem stipulating the relationship between present and future beliefs.

[58] Maher (1992, pp. 133–4) makes this point. (Exceptions are when the agent's credence function is non-conglomerable. See Arntzenius, Elga and Hawthorne (2004, p. 274–80).)

From what I have said thus far about the epistemic rules that the diachronic DBA has been thought to defend, one might expect that the argument is not easy to formulate. Indeed, there are a number of subtly different versions of the argument, and each tells a slightly different story. In the next few sections, I will work my way towards what I argue to be the only defensible version of the diachronic DBA. Firstly, (in Section 2.8), we must consider what sort of hypothetical sure losses are worth worrying about in the diachronic setting. I argue that the diachronic DBA is most promisingly framed as a story about successful sequential decision-making. Even then, it is not obvious how to formulate the argument in such a way that the sure losses involved are plausible indicators of epistemic failure. I criticise the standard sequential version of the argument—the "naïve" version—in Section 2.9. This can be contrasted with the "sophisticated" version that I come to defend in Section 2.10. The sophisticated diachronic DBA justifies only the weaker interpretation of conditionalisation outlined above, and it provides no defence for reflection. I discuss why this is a good result. A further test of the sophisticated diachronic DBA is whether it has suitable scope. If it is worth its salt, the argument should only justify conditionalisation in those cases where we think the rule does in fact properly apply.[59] I claim that the argument does indeed pass this test. However, there remain ways to challenge the sophisticated diachronic DBA. In Section 2.11, I suggest how we might provide an alternative defence of conditionalisation on purely epistemic grounds. With a view as to what is to come, then, let me begin my analysis by considering what sort of "pragmatically defective outcome" is potentially indicative of a faulty belief-update plan.

2.8 What sort of sure loss is important?

If there is going to be a valid diachronic DBA, then we know this much: given that it is a pragmatic argument, it will use sure loss in the dynamic setting as an indicator of epistemic inconsistency. But of course, not just any sure loss should be considered a

[59] I thank Alan Hájek for suggesting this line of questioning.

sign of failure. The conditions under which we should regard sure loss as a telltale bad outcome need to be carefully thought through. We might look to the synchronic DBA as a point of comparison. In this case, the sure loss to be avoided is that which might arise from grouping together bets that an agent regards as individually fair or favourable. Note that this is a somewhat artificial condition, because in any real-life scenario, the agent would have the opportunity to assess packages of bets in their own right, and would not be bound to judge the sum of fair bets as itself fair. There is a legitimate question (as discussed in the earlier part of this chapter) as to whether sure loss under these conditions is really indicative of epistemic inconsistency. When it comes to the diachronic DBA, I think it is even harder to work out just what sort of sure loss amounts to a pragmatically defective outcome. The rational agent is allowed to be vulnerable to some kinds of hypothetical betting losses, but not others. With this in mind, I will approach the diachronic DBA in the following way: I will start out with a betting scenario that leads to sure loss, but which clearly does not point to any epistemic fault on the part of the agent. This will give some perspective to my subsequent criticisms of the standard formulation of the diachronic DBA.

We can easily generate a sure loss scenario by requiring an agent to accept all bets that they deem fair or favourable at different times, where the agent may receive new evidence about the world between bets. For instance, let's say at time t_1 my credence in rain is 0.25. According to these rules I will accept a bet from a bookie of \$1 to my \$3 that it will rain (he wins \$3 if it rains; I win \$1 if it does not rain). Let's say at time t_2 my credence in rain has changed (because I have updated my beliefs on new evidence) to 0.50. Now I am supposed to accept a bet that pays me \$2 if it rains and costs me \$2 if it doesn't rain. Overall, if it rains, I make -\$3 + \$2 = –\$1. If it doesn't rain, I make \$1 – \$2 = –\$1.[60] I will be guaranteed a loss by accepting fair bets at different times, just because I have updated my beliefs on new evidence. (We can assume that I updated in accordance with conditionalisation.) Christensen (1991) toys with a diachronic DBA of this kind and indeed finds that it supports "calcification" (no change at all) in the belief set, which of course precludes any kind of belief-

[60] This example is from Christensen (1991, pp. 239–40), although he uses it to illustrate the fact that we don't expect the beliefs of two separate agents to be collectively coherent and thus immune from Dutch booking.

update rule. Clearly this cannot be right. The example effectively shows that if there is any story to be told about belief change, then it cannot be the case that an agent is able to accept any sequence of fair or favourable bets judged in isolation and be confident not to be bilked of cash.

The original version of the diachronic DBA retains the idea of an agent accepting any bets deemed fair or favourable, but emphasises that the cunning bookie has to name the bets from the outset that will guarantee sure loss for an agent with a faulty updating plan, regardless of what piece of evidence from some partitioning of the evidence is received. It is important that the bookie knows, at the outset, only what the agent knows—the agent's current belief set and their rule for updating on new evidence. It turns out that if the agent's updating rule is not conditionalisation (in the strong sense), then they can be taken for a sure loss—the cunning bookie can set a trap of bets, and then simply watch and wait while the agent accepts bets deemed fair or favourable at the initial and later time. This version of the diachronic DBA, including proof that a non-conditionalising agent can be caught for a sure loss, are given by Teller (1973, 1976), who attributes it to Lewis. (I detail the sort of bets that are involved later in Figure 2-2). Skyrms (1987) importantly proves the converse theorem: An agent who updates their beliefs in line with conditionalisation cannot be taken for a sure loss, in any sort of betting set-up where the cunning bookie knows only as much as the agent. These are important proofs, and will be relevant to any favoured version of the diachronic DBA. With the requirement that the cunning bookie must detail the sure-loss bets ahead of time, we have here an argument that actually supports (the strong interpretation of) conditionalisation, rather than "calcification". This is a significant result, but for reasons that I will outline shortly, I find aspects of this diachronic DB story rather unsatisfactory.

My main concern is that it is not clear why an agent need worry about the kind of sure loss that is described above. In effect, I do not think reference to a cunning bookie who knows only as much as the agent is very well motivated in the diachronic DB story. Why should a rational agent's plan for updating their beliefs be such that no-one could lay a trap guaranteeing their loss? The fact remains that the agent is

supposed to judge bets at different times on their individual merits.[61] But, as was made clear by the diachronic betting scenario that supports "calcification", an agent should never judge bets at different times on their individual merits, if there is any possibility that their belief set has changed in any way whatsoever. So I think it is unreasonable for this sort of constraint on an agent's assessment of bets to feature in the diachronic DB story. Putting constraints on what the cunning bookie is allowed to know does not change the fact that the agent should not be accepting individually fair or favourable bets in the first place.

Just what is at stake, I think, is made more obvious if the case is made in terms of an agent confronting a sequential decision problem. Recall from the previous chapter that this kind of decision problem is one where the agent expects to make a series of choices before receiving a final outcome. If the diachronic DBA can show that a non-conditionalising agent may suffer unnecessary losses due to the way that they assess strategies, then this would make for a more credible story. There would be no need to refer to cunning bookies that know only so much. It would rather be a matter of whether the agent's updating rule makes for problems in their assessment of strategies. In other words, what is important in the sequential-choice setting is the agent's current relationship to their future betting behaviour—how the agent now thinks they will respond to future betting offers, and what this means for their current choice of strategy. It does not matter whether the agent actually follows through with their chosen strategy at the future times in question. The important thing is whether the agent's initial plans, at least, are rational, or whether they involve "pragmatically defective outcomes". (Note that this means a sequential-choice version of the diachronic DBA can only ever provide justification for the weak interpretation of conditionalisation.)

---

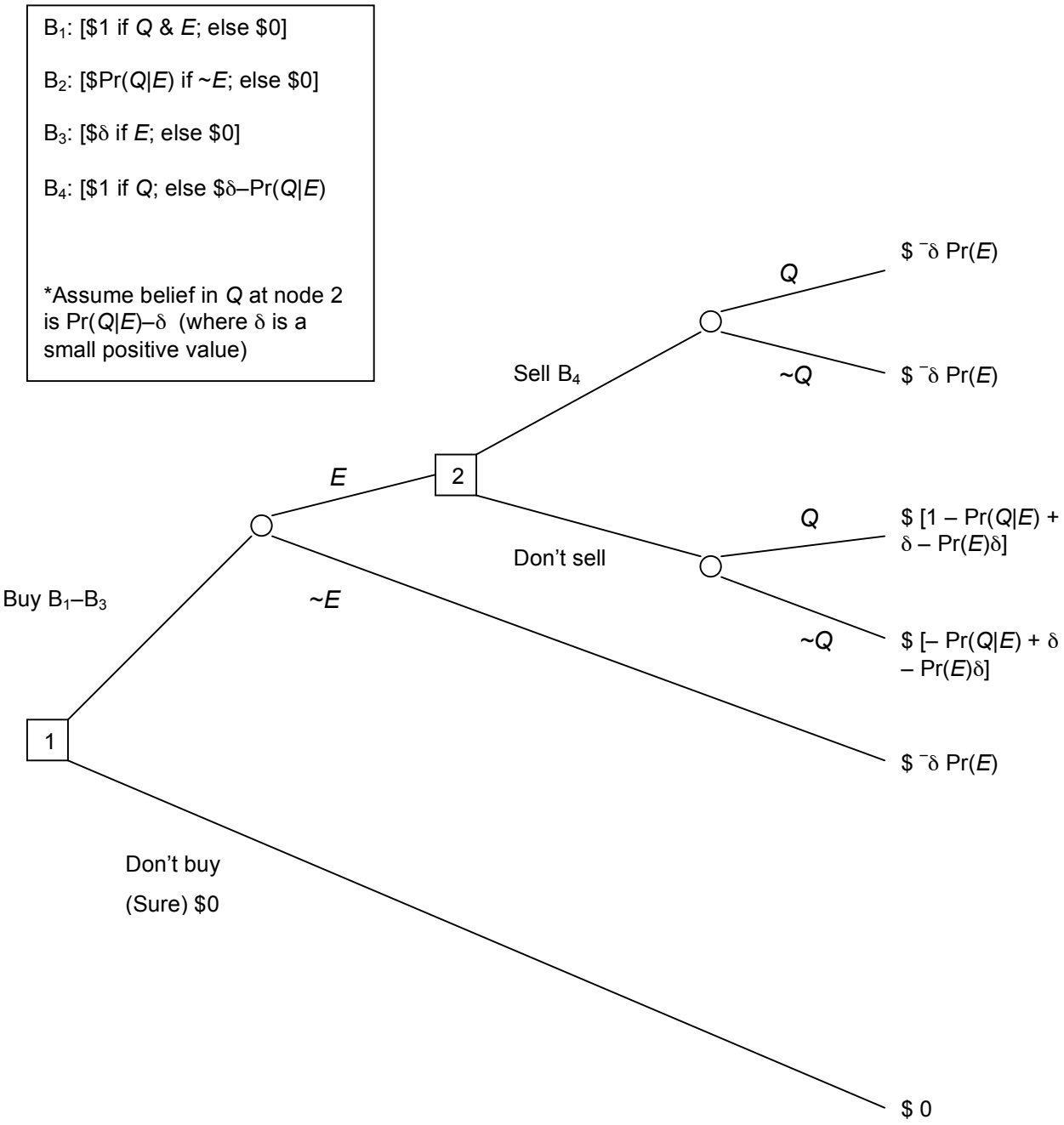
[61] Skyrms (1987 & 1993) tells the story somewhat differently. It is the agent who sets the odds ahead of time (acting as the bookie), and the cunning bettor (as opposed to bookie) then has the opportunity to place any bets of their choosing. If the agent's odds do not reflect a plan to update via conditionalisation, then the bettor can take them for a sure loss, however the evidence turns out. This reading is admittedly more convincing than the alternative—where the agent does the betting—but I still think the idea of an agent having to set odds ahead of time is somewhat unmotivated, or at least less motivated than a sequential-choice version of the argument, which I will discuss shortly.

In the next couple of sections, I will contrast two possible sequential-choice accounts of the diachronic DBA. Section 2.9 argues that the more standard account, which depends on naïve choice, has some serious shortcomings. In the previous chapter, I argued that in ordinary decision-making contexts, an agent is unwise to pursue naïve choice. While in the context of an *argument* for conditionalisation, the naïve approach to strategy assessment may nonetheless play a role, I show that there is not sufficient motivation for championing this approach to strategy assessment, even in idealised circumstances. Once again, the problem is that the sure losses featured in this version of the DB story are not convincing. This does not, however, put the diachronic Dutch book argument to rest. In Section 2.10 I use a result from Skyrms (1993) to argue that the credibility of the diachronic DBA surprisingly rests on the agent being a sophisticated chooser.

2.9 First pass: Does the diachronic DBA depend on naïve choice?

The standard sequential account of the diachronic DBA assumes a particular method for assessing strategies that I refer to as naïve choice, and which I outlined in the previous chapter. A naïve sequential decision maker assesses strategies entirely from the point of view of their current or initial beliefs and preferences. Importantly, such a decision maker assumes that they will make choices at future times that accord with their current choice dispositions. In order to assess the "naïve" version of the diachronic DBA, we need to consider the sort of betting scenario that the agent is thought to face. Figure 2-2 depicts the sequential decision problem that is typically thought to confront the agent in the DB story. The diagram closely resembles those found in Earman (1992, p. 48) and Skyrms (1993, p. 323). (The bets are identical to those featured in the original Teller-Lewis story.)

Figure 2-2



Note that for a naïve chooser at node 1:

Expected value of buying = $1–δ Pr(*E*)

Actual value of buying = (Sure) $¯δ Pr(*E*)

At node 1 the agent decides whether or not to accept a set of three bets, where these bets effectively amount to a bet on evidence proposition *E*, and a bet on proposition *Q* conditional on *E*. (The book of bets in fact includes a bet on *Q & E*, plus a bet on *E*, plus a bet on ~*E*.) Clearly the agent's conditional probability for *Q* given *E*—Pr*(Q|E)*— plays a role in the assessment of these initial bets. There is a fourth bet that is available at node 2—it is the option to sell a bet on *Q*. A naïve chooser assesses the choice at the second node from the perspective of their current beliefs and preferences. So the plan about whether to accept Bet 4 is also based on the agent's current conditional probability Pr(*Q|E*). Not surprisingly, problems will arise if the agent's belief updating plans are such that they do not expect their belief in *Q* to equal Pr(*Q|E*) when they get to node 2. The agent assesses the strategy from the point of view of their current beliefs and preferences, despite conditionally planning to hold an alternative belief function in the future. This is a recipe for sure loss. Referring to Figure 2-2, the agent assesses their chosen option as having positive expected utility (because they falsely think that they will reject Bet 4), when in fact, if everything goes according to plan, they are bound to accept Bet 4 and suffer sure loss, however the evidence turns out. (Note that Figure 2-2 really only tells half the story—the problem is designed to catch out the agent who plans to update their belief in *Q* to a value that is *less than* the relevant conditional probability. A slightly different problem will bring sure loss to an agent who plans to update to a value greater than the relevant conditional probability.)

A well-known and important challenge to this account is that restricting the agent to naïve choice does not seem well motivated. If what we are interested in is whether the plans that an agent commits to are consistent, then surely we should allow the agent all the planning tools that are on offer. As noted in Chapter 1, two main alternatives to naïve choice have been suggested in the literature, one "resolute" and the other "sophisticated". I argued in Chapter 1 that there are some serious problems with resolute planning; hence, I will stick to the "sophisticated" alternative to naïve choice. Recall that this is essentially a process of backwards planning, whereby the agent considers what they will choose at future nodes and works backwards to determine which strategies are actually possible and which are rather pipe dreams that should be recognised as non-options. Maher (1992, p. 125) and Levi (1987) show that an agent

who does not plan to update via the rule of conditionalisation can still make sound assessments of strategies. When it comes to assessing the decision problem in Figure 2-2, for instance, the agent realises that they will accept the bet offered at time 2, and that when combined with the other bets, this will result in sure loss overall, so they had better not even venture down this path. The agent chooses not to bet at all.

Skyrms (1993, p. 323) suggests that the appeal to sophisticated choice "unfairly prejudices the case against dynamic coherence arguments". Skyrms also answers to the sophisticated challenge in a more substantial way, and I will go on to discuss this move in Section 2.10. For the time being, I want to explore whether the diachronic DBA can plausibly restrict an agent's decision-making to naïve choice. While I do not think it can be argued that the naïve approach to sequential decision-making is the only rational approach (as per my claims in Chapter 1), the DB defender might focus on the advantages to be had if one's decision-making apparatus allows naïvety. In the appropriate situations naïve choice affords the luxury of being able to evaluate strategies from the perspective of present beliefs and preferences, in the confidence that this will match what one actually plans to do at future choice nodes. There is perhaps less work to be done in the naïve sequential decision-making process—we do not need to engage in backwards planning (as per sophisticated choice). The agent need only think about their current attitudes. The diachronic DBA might thus be interpreted as saying that only one rule for updating on new evidence supports the luxury of naïve strategy assessment, and that rule is conditionalisation.

On this reading, the DBA provides defence for conditionalisation, then, only to the extent that naïve choice really is a luxury; to the extent that naïve choice has benefits that other dynamic decision-making methods (i.e. sophisticated choice) cannot boast. As mentioned, there are arguably some benefits of this kind. But it is highly questionable whether these benefits are sufficient to make conditionalisation the preferred updating rule, all things considered. For instance, the fact that conditionalisation supports naïve choice might go in its favour if this approach requires less computational power than sophisticated choice, but we would not want to put too much weight on such a consideration. After all, there may be alternative

rules for updating beliefs on new evidence that require less computational power than conditionalisation. A further complication to the naïve choice story is that it is unclear how often an agent will be able to reap the benefits of this dynamic decision-making luxury. In fact, I will go on to argue that this complication is in fact devastating to the naïve diachronic DBA. There are plenty of ordinary circumstances in which a well-intentioned conditionalising agent is unwise to engage in naïve choice. The luxury of naïve choice afforded by conditionalisation is surely dampened if the agent has to at least sometimes avert to sophisticated planning.

To begin with, in the majority of cases, an agent will only avoid sure loss in the diachronic naïve-choice scenario if they plan on updating via conditionalisation, and they also predict that they will later stick to these updating plans. This means that the naïve version of the diachronic DBA supports a very particular account of conditionalisation. It does not go so far as to support the "strong" interpretation of the rule, i.e. it does not require that an agent's beliefs actually change (in real time) in accordance with conditionalisation. It is only the agent's current projections about the future that matter—the agent must merely *plan* to update via conditionalisation. Importantly, however, the naïve diachronic DBA appears to require that the agent also predict (from their current position) that their plans will be successful. In such case, the agent will in ordinary cases satisfy the reflection principle, which I outlined earlier. But there is something disingenuous about requiring an agent to predict the success of their plans—surely we can only expect the agent to have rational belief-updating intentions, and we should not take them to task for predicting that in some circumstances, these intentions may not be realised. Indeed, there are some obvious counterexamples to reflection that make the principle look confused, and this in turn should make us suspicious of the naïve diachronic DB defence of conditionalisation.

Let us consider an example. Mary plans to drink generous amounts of champagne tonight. She predicts that at some point in the night, on account of the champagne, she will believe that she has extraordinary talent as an opera singer. Mary does not now believe that she has such talent, and for obvious reasons, she is not inclined to shift her current degrees of belief to match her predictions regarding her future degrees of

belief on the matter.[62] While Mary is definitely violating reflection here, it is not so clear that she is violating conditionalisation. We need not call her champagne plans a violation of conditionalisation because it could well be the case that Mary does not *plan* to update her beliefs via some rogue rule. It is just that despite her best intentions, she predicts that she will simply be incapable of updating via conditionalisation. In other words, Mary's plans do not fly in the face of conditionalisation; it is just that, according to her best predictions, any plans that she cares to make with respect to updating her beliefs will go astray.

The reflection principle is not so well positioned as the rule of conditionalisation to deal with a distinction between an agent's *plans* versus their *predictions* with regard to belief change. Reflection says something stronger than conditionalisation—it is not just about plans for belief change, but rather about how present beliefs should mesh with future beliefs, regardless of how the different possible future belief sets might come to pass, and with what probability. This is problematic because sometimes there is reason to predict changes of belief that one does not now endorse from an epistemic (if not a pragmatic) point of view. These considerations suggest that conditionalisation is an appropriate epistemic norm whereas reflection is not. Of course, defenders of reflection can argue that the rule does not apply in these sorts of cases that involve a departure from ideal rationality. But Mary is rational at the time that she is assessing the champagne plan. She can make well-informed predictions about what her opera-singing beliefs will be at the end of the night, and she will pursue the champagne strategy if the payoffs are good enough. So if reflection rules out such cases, it must be because Mary's future self, as opposed to her present self, is not ideally rational. We have then a puzzling situation: reflection is supposedly a synchronic norm, one that constrains an agent's current beliefs, conditional on having particular belief functions at some time in the future. But the norm is only valid if the agent predicts that they will always be ideally rational from the point of view of their current belief set. And this assumption calls into question much more than just the agent's current rationality.

---

[62] Maher (1992) gives a more morbid example that concerns an agent getting drunk and thinking they can drive safely!

Just as reflection does not handle the distinction between plans and predictions with respect to beliefs, I do not think a justificatory story that champions naïve choice handles this distinction well either. Pragmatics must come first; it is legitimate to pursue a strategy in which one's predicted beliefs do not match one's planned beliefs, if there are suitable payoffs involved. Alternatively, an agent might predict that their tastes or desires for outcomes will change in the course of pursuing a particular strategy, and moreover, such a change does not seem irrational. (I will discuss this possibility in the next chapter.) In such cases, sophisticated choice is the only robust method for assessing strategies. Where an agent predicts a belief change that does not accord with conditionalisation, they must abandon the naïve approach if they want to be sure to make sound assessments. Again, we can accommodate the exceptions by claiming that naïve choice is a luxury to be enjoyed by conditionalisers just in those cases where predicted belief change matches planned belief change (and where utilities over outcomes are stable). But perhaps the "exceptions" to these circumstances are not so rare, in which case sophisticated choice should be pursued the majority of the time. Indeed, it might be argued that we should always make allowances for changes in taste, or for belief change not going to plan, in which case naïve choice is never a viable approach, even for the well-intentioned conditionalising agent. And if we agree to this much, we should be sceptical about whether there is any reason to consider naïve choice the superior approach to sequential decision-making.

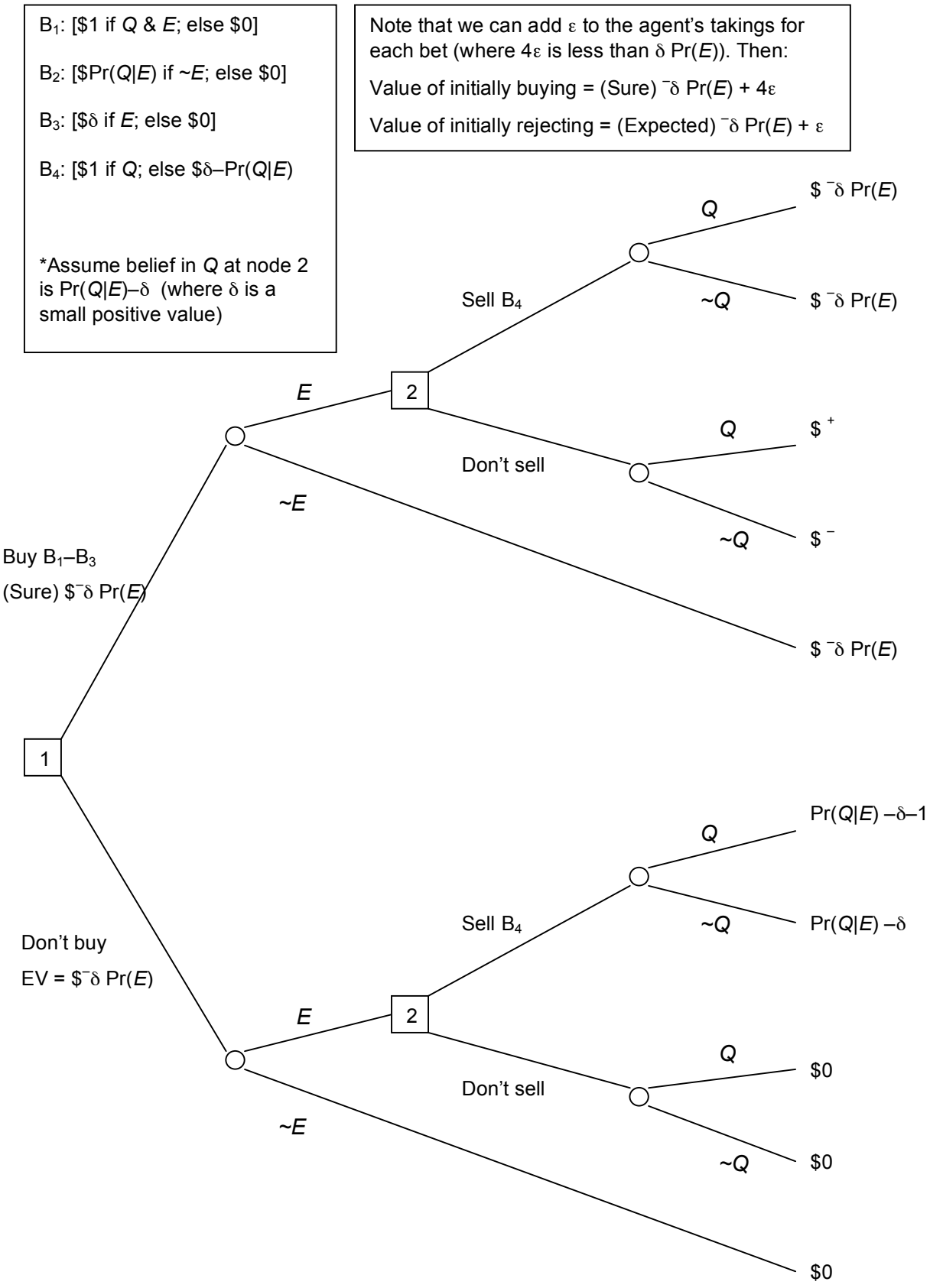
2.10 Second pass: The "sophisticated" agent

Skyrms (1993) makes an important advance on the diachronic DB story to counter the claim that a sophisticated chooser can avoid any pragmatically defective outcome, whether or not they subscribe to conditionalisation. Skyrms points out that regardless of whether the agent decides to bet in the first instance, the cunning bookie can in any case offer the agent the second round of bets (i.e. the bookie can in any case offer Bet 4 in Figure 2-2). And this changes everything for the non-conditionalising agent. The revised sequential decision problem is depicted in Figure 2-3 below. (It mirrors the

model outlined in Skyrms (1993, p. 325), and is again designed to catch out an agent who plans to update their belief in Q upon receipt of new evidence E to a value that is less than $\Pr(Q|E)$.)

According to this version of the story, our agent is sophisticated and they can thus see the sure loss coming if they opt to buy the three bets. The trouble with this case is that initially rejecting the bets is calculated to have the same negative expected value. To repeat, rejecting the bets does not lead to *certain* loss, but it leads to an equivalent expected loss all the same. So the sophisticated agent may well end up taking the option that leads to sure loss. Skyrms suggests a slight modification of the story that puts the agent well and truly in line for a sure loss. We can simply specify that the agent will receive an extra small positive amount ε per transaction (refer again to Figure 2-3). So if the transactions were already thought fair, then they will now be each slightly better than fair. It turns out that this modification makes the sure loss a slightly better option for the sophisticated agent, and so they will definitely pursue this path.

Figure 2-3


B_1: [$1 if Q & E; else $0]
B_2: [$Pr($Q$|$E$) if ~$E$; else $0]
B_3: [$δ if E; else $0]
B_4: [$1 if Q; else $δ–Pr($Q$|$E$)
*Assume belief in Q at node 2 is Pr(Q|E)–δ (where δ is a small positive value)
Note that we can add ε to the agent's takings for each bet (where 4ε is less than δ Pr(E)). Then:
Value of initially buying = (Sure) ⁻δ Pr(E) + 4ε
Value of initially rejecting = (Expected) ⁻δ Pr(E) + ε
$ ⁻δ Pr(E)
Q
~Q
$ ⁻δ Pr(E)
Sell B_4
E
2
Q
$ ⁺
Don't sell
~E
~Q
$ ⁻
Buy B_1–B_3
(Sure) $⁻δ Pr($E$)
$ ⁻δ Pr(E)
1
Pr(Q|E) –δ–1
Q
~Q
Pr(Q|E) –δ
Sell B_4
Don't buy
EV = $⁻δ Pr($E$)
E
2
Q
$0
Don't sell
~E
~Q
$0
$0

The question is whether the sure loss that the sophisticated agent walks into is a significant one. Although this kind of loss is an oddity, it is not clear that it is indicative of epistemic irrationality. Refer again to Figure 2-3. Here we have an agent choosing what is predictably a sure loss, over another option that has lower expected value for the agent. We might say that there is nothing outright inconsistent about this choice. The agent is not choosing a sure loss over another *available* option that has sure zero or positive gain. And they are not making any faulty assessments of plans. The agent is simply taking what looks to them to be the better of two evils. At least, that is one way of explaining their behaviour. We might look more closely at the concept of an *available* option. This is the key to determining whether or not the non-conditionalising agent takes an option that is effectively dominated by another. Given the belief-updating plans of such an agent, a sure zero gain is not an option *for them*. But the agent's belief-update plan is the very thing that we are trying to assess. So a more objective account of what are the available options seems appropriate in the context of the argument. It is surely important that *if* the agent had planned to update via conditionalisation, then a sure zero gain would have been available to them. The non-conditionalising agent can only blame him/herself for the sure loss. The agent is not forced into sure loss by any external circumstances (the agent has only to accept bets that they deem fair or favourable); they rather bring this outcome upon them self, due to their own belief-updating plans. This is a good reason to take the plans to be faulty.

Given that the sophisticated version of the diachronic DBA shows promise, let's see how it fares with the exceptions. To begin with, there are the cases where an agent predicts that their belief-updating plans will go astray. I have argued that these cases are more or less fatal for the naïve diachronic DBA, because this version relies on an agent aspiring to naïve choice, but we can see that future psychological limitations sometimes make naïve choice defective. More importantly, we might question why naïve choice should be thought superior to its sophisticated alternative in the first place. When the diachronic DBA is framed in terms of sophisticated choice, there is not, of course, the problem that an agent may assess plans naïvely (and mistakenly) in those cases when they predict that their best intentions with respect to belief-updating will not come to pass. A future psychological limitation is treated just like any other

external constraint—it can sometimes lead to outcomes that could have been bettered were the constraint removed, but at least the agent assesses the strategies before them accurately. The sophisticated diachronic DBA cuts to the real problem. What is of concern is any sure loss that the agent brings upon him/herself by subscribing to one belief-update plan rather than another.

Some other examples might seem to present greater challenge to the diachronic DBA justification of conditionalisation. These are counter-examples to the updating rule itself, even when it is interpreted in the weak sense as a constraint on an agent's belief-updating *plans*. I do not think the exceptions put the status of conditionalisation in jeopardy. It is perfectly reasonable for a rule to hold only in some specified domains, so long as we can give an account of why the rule should hold in these domains and not in others. In fact, I claim that the sophisticated version of the diachronic DBA, unlike the naïve version, is able to distinguish these different domains.

The following is an example where the agent does not just predict a rogue change of belief, but in fact plans to update their beliefs other than via conditionalisation: Let's say there is some kind of social institution that rewards agents who have a particular set of beliefs at a given time.[63] (For example, extremely nationalistic people might have a better chance of procuring desirable jobs in the public service.) We will assume that our agent attaches greater value to whatever rewards are associated with the socially encouraged set of beliefs than to being a free-thinking individual. According to my story, to receive what they think are the greater rewards, the agent must actually plan to update contrary to their conditional probabilities. To be frank, I am not convinced that such a plan can count as rational, or even possible. But that is because I have other non-pragmatic reasons for endorsing conditionalisation, which I

[63] Maher (1992, p. 133) discusses a fantastical example in which some superior being rewards an agent for having particular beliefs at some given time. Presumably, Maher appeals to a superior being because we might think that an agent can otherwise simply *pretend* to have the rewarded set of beliefs. I will use a more ordinary example, however, because I think it is often not possible to convincingly pretend to believe something that one doesn't in fact believe.

will discuss in Section 2.11. There is certainly scope to argue that the agent can plan to update their beliefs irrespective of the evidence to a function that will be more advantageous to them. Naïve choice, of course, is not conducive to any belief-update plan that does not accord with the agent's current conditional probabilities. But the sophisticated version of the diachronic DBA can distinguish these cases in which an agent is rewarded for having particular beliefs, from other sorts of cases. In fact, the sophisticated version actively recommends updating in such a way as to reap the best rewards. To the extent that the agent has a say in the matter, their decision-making plans should not lead them to an inferior outcome. So in this particular case, if the agent is indeed capable of planning to update to the socially rewarded set of beliefs, then the sophisticated diachronic DBA indicates that this is the rational plan for the agent to adopt.

To close this section, I want to briefly consider a more general issue concerning the relationship between the form of a dynamic decision model and the sort of updating plan that is warranted. As stated at the outset of this chapter, I have been concerned only with strict conditionalisation, and the diachronic DBA for this belief-update rule. Recall, however, that at least one modification of strict conditionalisation has been proposed, and a modified diachronic DBA has been developed to support it. I am referring to "Jeffrey-conditionalisation", and the associated diachronic DBA articulated by Armendt (1980) and Skyrms (1987). One might wonder how the possibility of an alternative updating rule can be reconciled with the diachronic DBA that I have outlined here. Well, it comes down to the way the agent depicts the dynamic choice problem that confronts them. Here I have been assuming that the agent expects to become certain of one evidence proposition amongst a particular partitioning of the possibility space. But this is not to say that this is the only kind of model for learning new evidence. Jeffrey-conditionalisation, for instance, is based on the idea that an agent does not expect to become certain of anything; the agent expects, rather, to merely change their probabilities for the evidence propositions over some partitioning of the possibility space. The diachronic DBA that I have outlined is just not designed to address belief-updating rules in these kinds of cases. Importantly, the sophisticated diachronic DBA for strict conditionalisation that I have outlined does not give the wrong results in these alternative evidential circumstances; it simply

does not address them.

2.11 Is there a more straightforward justification of conditionalisation?

While I have argued that the sophisticated diachronic DBA has much merit, there are reasons to be sceptical of even this version of the pragmatic argument for conditionalisation. As mentioned, it could be denied that the sophisticated DBA reveals any outright inconsistency in a non-conditionalising agent's sequential decision-making. There is certainly something odd about a sure loss associated with a strategy that involves an agent freely assessing bets at different times, especially when the agent predicts that their beliefs will change in accordance with their best plans. But one might insist that that is all the situation amounts to—an oddity. We might explain to the agent that if only they were prepared to change their updating plan then they could receive a sure zero gain, rather than suffer a sure loss. But our rogue-updating-agent might simply shrug their shoulders at all this, and maintain that their updating plan is correct. Given their expected future beliefs, the zero gain is *simply not a live option*. (I pursue this debate about the significance of the losses in the sophisticated DBA in Chapter 6, because the issue is also pertinent to sequential-choice analyses of the independence axiom of SEU theory.)

In any case, it seems strange to select a rule for updating beliefs on the basis of pragmatic outcomes. The story goes like this: given that conditionalisation makes more favourable betting options available to us, it must be the right way to update our beliefs. It is not obvious why this link between pragmatic considerations and a rule for updating belief should hold. I do not think it is, for instance, as compelling as the relationship between fair betting odds and degrees of belief that underpins the synchronic DBA. Moreover, the diachronic DBA really says very little in the end. As stressed, the argument is purely about an agent's *plans* for updating their beliefs, and not about whether they predict that these plans will come to pass, or about how their beliefs do in fact end up transforming in real time. But surely there are more direct

ways to justify a mere plan for updating beliefs on new evidence?

Indeed, I think the best argument for conditionalisation (interpreted in the weak sense) is that it *just seems unreasonable* (in normal circumstances) to plan to update one's beliefs/preferences in a manner that does not match one's current conditional beliefs/preferences. My conditional probability Pr($Q|E$) expresses my belief in Q, were I to know E. In standard cases, then, it would seem very strange for me to entertain a plan whereby upon finding out E for sure, I will change my belief in Q to something other than Pr($Q|E$). Such a plan simply would not make sense! It is close to contradictory for an agent to assert that they have Pr($Q|E$) = x, and yet they intend to update their belief in Q upon finding out that E is true to something other than x. Surely the latter is just what a conditional probability is supposed to express. Either the agent is not reporting their true belief in Q given E, or else they don't understand what this conditional credence is supposed to represent.[64]

I note that Skyrms (1987, p. 3) argues that a mere understanding of conditional probability is not sufficient reason for planning to change one's beliefs in accordance with such conditional probabilities. (Hence the need for the diachronic Dutch book argument.) Skyrms later cites Hacking's (1967, p. 315) related point that "…*Prob(h|e)* stands merely for the quotient of two probabilities. It in no way represents what I have learned after I have taken *e* as a new datum point." I do not think this statement of Hacking's is quite right, however, and I think Skyrms is too quick in his claim that we cannot infer an agent's updating plans from their interpretation of conditional probability. In Skyrms' Dutch book story, *Prob(h|e)*, or Pr($Q|E$), to use my original notation, is not just the quotient of two probabilities; these quotients or conditional probabilities are already interpreted in terms of conditional bets. If there is a jump in the argument, then surely it is at this initial stage, when we claim that Pr($Q|E$) should be an agent's fair betting quotient for a conditional bet that is won if Q and E are both true, lost if Q is false and E is true, and called off if E is false. (Skyrms notes that de

[64] According to this way of understanding belief-update plans, an agent cannot simply plan to update contrary to their conditional probabilities, even if there are pragmatic rewards involved (recall the fairy Godmother case from Section 2.9). To do so would amount to having two conflicting interpretations of conditional credence.

Finetti gives a coherence argument for the ratio definition of conditional probability that hinges on this basic claim that a conditional probability stands for an agent's fair betting quotient for the associated conditional bet.) Once we have accepted this pragmatic interpretation of a conditional probability, then I think we are bound to a particular plan for updating belief. If an agent is prepared to pay an amount x for a conditional bet on Q given E, then they should be prepared to pay the same amount x for a bet on Q, having just found out that E is in fact true.

In other words, I think it is rather a matter of correctly applying the betting interpretation of conditional probability that we should plan on updating our beliefs in accordance with these probabilities. The diachronic Dutch book argument, with its threat of pragmatically defective outcomes, seems rather extraneous. It is simply indicative of a *divided mind* or a lack of integrity to plan on updating (in normal circumstances) in any non-conditionalising way. I borrow the term "divided mind" from Skyrms (1987) and Armendt (1993), but I think this is a much stronger or more obvious sense of *divided mind* than what Skyrms and Armendt had in mind in relation to the synchronic DBA. In the synchronic setting, an incoherent agent can be said to have a divided mind only if a substantial extra assumption is granted—that the sum of fair bets should itself be fair. When it comes to planning belief change, no such extra assumptions are necessary. If the agent plans to update Q in the case that E is learnt other than via conditionalisation, then they have a genuinely divided mind in the sense that they essentially misapply the standard interpretation of the conditional probability $\Pr(Q|E)$.

2.12 Diachronic DB conclusions

I have here focused on two different ways to understand the diachronic Dutch book argument in terms of sequential decision-making. The first is the standard naïve version; the second is the sophisticated version, which is inspired by Skyrms' (1993) defence of the diachronic DBA. While Skyrms might have intended just to shore up

the argument, I claim that he significantly transformed it. The naïve version of the diachronic DBA suffers from some serious flaws. The main problem is that it is rather tenuous to champion naïve choice, given that there is at least one other perfectly reasonable dynamic decision-making strategy on offer. Where an agent predicts that their updating plans will go astray, they simply must choose according to the sophisticated approach. And it is unclear why an agent should aspire to naïve planning just in those cases where they predict that their belief-updating plans will be realised (and their preferences will remain stable). Furthermore, the naïve approach recommends updating via conditionalisation in those cases (involving rewards for particular belief sets) where this is arguably not the appropriate rule. In such cases, not only do some argue that a rational agent should not conditionalise, they would additionally be foolish to assess strategies naïvely.

One may well ask why it is not easier for the agent to always act as a sophisticated chooser, given that this sequential-choice approach always leads to accurate assessments of plans. If the rational agent must at least sometimes appeal to sophisticated choice to be immune to sure loss, then simplicity would suggest that this is the superior dynamic decision-making approach, even in ideal circumstances. Importantly, and perhaps surprisingly, Skyrms (1993) shows that it is possible to frame the diachronic DBA in terms of sophisticated choice. What is at issue is whether the agent's epistemic plans prevent them from sure gains that they might otherwise have enjoyed. Of course, external constraints (including future psychological limitations) might also prevent the agent from obtaining pragmatic benefits, but such losses are not relevant because they are unavoidable. The sophisticated version of the diachronic DBA focuses attention on sure losses that could have been avoided were the agent to select the correct belief-updating plan. Moreover, the sophisticated diachronic DBA is sensitive to those cases in which it is advisable to follow an alternative belief-update plan (when there are special rewards for updating to specific belief sets).

As per the synchronic DBA, one might maintain that we are pursuing the wrong path in trying to justify an epistemic norm on pragmatic grounds. Indeed, I personally

think that (the weak interpretation of) conditionalisation can be justified at the level of interpreting conditional probabilities, as discussed in Section 2.11. Put simply, an agent who understands the standard interpretation of conditional credence should in normal cases plan to update via conditionalisation. Even if this more straightforward justification of conditionalisation is accepted, we can nonetheless regard the diachronic DBA as lending support to the belief-update rule. (I drew a similar conclusion about the synchronic DBA in Section 2.6.) Of course, it all depends on whether we finally want to claim that the sure loss featured in the sophisticated diachronic DBA is genuine, or whether if involves an unfair comparison with something that is not really a *live option*. I have not yet said anything decisive on this issue, but it will become very important in Chapter 6, when I use the sequential-choice framework to assess core axioms of SEU theory, notably independence. I show that an agent who plans to violate independence can be caught by the same kind of sure loss as an agent who plans to update their beliefs contrary to conditionalisation. So if we buy the diachronic DBA, then we must also buy a similar sort of argument for upholding the independence axiom. This only strengthens the case that the pragmatic defences of Bayesian epistemic norms are intimately tied to SEU theory. But let us save the final analysis of diachronic Dutch-book-style arguments for the end of Part II. Now I want to consider an issue I touched on a couple of times in the discussion so far: the issue of updating desire or preference.

3 PLANNED PREFERENCE CHANGE

3.1 Introduction

The story about how an agent should account for their future attitudes in decision-making would not be complete without an analysis of preference change. In the previous chapter I concentrated exclusively on the belief side of the story. My aim was to determine just what sort of Bayesian epistemic norms can be justified, whether on pragmatic or other grounds. And, of course, the further consideration is how these norms constrain the modelling of belief in decision models. The basic probabilist norm—that degrees of belief should conform to the probability calculus—is pretty straightforward to interpret. It also sits very well with SEU theory, whether or not we think the norm is ultimately justified on pragmatic grounds. But it is not obvious just what the rule of conditionalisation requires. The version of the rule that I have been supporting is rather nuanced; it holds that an agent should *plan* on updating their beliefs in accordance with conditionalisation, but when it comes down to it, there is still the possibility that the agent's best-laid plans will go astray. So conditionalisation does constrain how an agent should think about their future beliefs, but only in a limited sense. My goal in this chapter is to consider how these lessons about belief translate to the modelling of desire. While I think there is an important sense in which the decision model should be able to cope with any predictions that an agent might have with respect to their future beliefs or preferences, I do not want to play down the importance of having the right sort of updating plans or intentions. I will argue that the desire and belief stories are not completely analogous in this respect; a rational agent might intend changes of desire that are not in line with their current conditional desires/preferences.

By way of putting this discussion in context, I want to point out that while there has been some substantial debate about the kinematics of belief in the decision theory and related literature, much less attention has been paid to preference/desire change. Perhaps this is no surprise, since, in general, there has been much greater concentration of effort on the belief side of the decision-modelling story, as compared to the desire side. And there are some good reasons for thinking that this imbalance is quite reasonable. After all, desire seems to lack the anchor that belief has in facts about the world. Or else, even if there is some objective standard for desires (think moral facts)[65], this is a matter of some controversy, and there will not be the neat bridge between moral facts and an agent's desires that frequency data offers between objective chance and the agent's beliefs.[66] Notwithstanding the greater controversy surrounding desire, however, it is unclear why a formal representation of desire should be less rich or useful than a formal representation of belief, when it comes to the relationship between desires, and the impact of new evidence. Indeed, the similarities between the two representations run deep. If we take seriously the SEU representation theorems of Savage (1954) and Jeffrey (1965 & 1983), for instance, we find that the numerical representations of belief and desire have the same humble beginnings in ordinal preference.

Here I will attend to the desire side of the decision-modelling story, and, in particular, to the question of how a rational agent might plan to update their desires. Some important inroads have already been made on this issue. A good place to start is Jeffrey's (1965) account of the desirability of prospects or propositions. Indeed, Jeffrey's account of desire grounds my discussion in this chapter, so it is worth pausing a moment to outline his "axiom of desirability". We assume that an agent has a probabilistic belief function Pr and a utility function U over possible worlds (w_i).[67] (According to Jeffrey's evidential decision model, the utility function over possible

---

[65] Colyvan et al. (to appear) explore what constraints an agent's utility function must satisfy to be considered morally correct, according to each of three key ethical theories.

[66] I am assuming here that there is some sense to be made of objective chance.

[67] I might otherwise refer to the possible worlds as possible outcomes. Importantly, the set of possible worlds/outcomes contains all the possible outcomes for all of the actions available to the agent at the time.

worlds measures how much the agent desires to learn (or would be happy to learn) that the world in question is the actual world.[68]) According to the "axiom of desirability", the desirability of any proposition Q amounts to "a weighted average of the possible ways Q can be true, where the weighting on each possible way is its conditional probability of truth, given that Q" (Bradley 2006). This can be stated formally as follows:

$$\mathrm{U}(Q) = \sum_i \mathrm{U}(w_i) \times \Pr(w_i \mid Q) \quad \text{(where the } w_i \text{ are all worlds in which } Q \text{ is true)}$$

According to Jeffrey, the desirability of Q, conditional on some proposition E being found to be true, just amounts to the desirability of Q & E. Stated formally:

$$\mathrm{U}(Q|E) = \mathrm{U}(Q \,\&\, E)$$

$$= \sum_j \mathrm{U}(w_j) \times \Pr(w_j \mid Q \,\&\, E)$$

(Here the w_j are all worlds in which Q & E is true.)

$$= \sum_j \mathrm{U}(w_j) \times \Pr{}_2(w_j \mid Q) \qquad \text{where } \Pr{}_2 = \Pr(\cdot|E)$$

(Here the w_j are all worlds in which Q is true.)

On this account of conditional desire, the desirability of learning that one will go to the beach tomorrow *given* that it will be raining is just the desirability of learning that one will go to the beach *and* it will be raining tomorrow.

I note that Bradley (2005 & 2006) has expanded Jeffrey's original work on

---

[68] I acknowledge the argument that Jeffrey's evidential decision theory sometimes gives the wrong results. It advises the agent to choose the act that will bring them the most desirable news. But such an act may not be optimal when "there is a *statistical* or *evidential* correlation between the agent's choices and the occurrence of certain desirable outcomes, but no *causal* connection between the two" (Joyce 1999, p. 146). (The assumption here is that the most optimal act is in fact the one that can bring about the best news, rather than merely betoken the best news.) Decision theories that that capture this requirement that acts be ordered in terms of their causal efficacy in bringing about good outcomes are referred to as causal decision theories (see, for instance, Joyce 1999). While the distinction is very important, I do not think it is a major concern for my treatment of desire in this chapter (nor for my discussions in other chapters, for that matter). I refer to Jeffrey's "axiom of desirability", but I am assuming that the axiom can be modified to reflect causal considerations via an adjustment to the probability term in the expression, and that such a modification would not affect my findings.

conditional desire—he discusses updating the desirability of propositions in response to changes in credence, where the belief changes originate either in the agent learning some certain evidence proposition (as per classical/strict conditionalisation), or else in the agent revising their subjective probabilities over some partitioning of the possibility space (as per Jeffrey-conditionalisation). Bradley also defines conditional desirability slightly differently to Jeffrey—he prefers to think of conditional desire in terms of the *extra* satisfaction that an agent would have if they were to learn, say, Q, given that they already know E, rather than the desirability of learning that $Q \,\&\, E$. In the context of this discussion, however, such details will not be important; it is sufficient to give one plausible definition of conditional desire. To keep things simple, as with my discussion of belief in the previous chapter, I here assume contexts in which the agent's evidential experience lends itself to classical Bayesian belief-updating—when what the agent comes to know through experience can be summarised by a single evidence proposition.

My initial question (in Section 3.2) is whether there is some plausible modification of the diachronic DBA that does for desire what the original DB story does (or at least attempts) for belief—that an agent should only ever plan to update their desires (in the appropriate evidential circumstances) in accordance with their current conditional desires. The question about the diachronic DBA is a loaded one, because it seems reasonable for an agent to plan to update their desire function in a manner that is independent of any change in credence. Surely a genuine change in taste is not irrational, even if, as Bradley (2005) suggests, such a change is nonrational, or outside the scope of rational principles. It seems overly procrustean to think that planned preference changes can only be in response to new relevant information, as, for example, when someone plans to revise their preference for going to the beach over going to the movies if they were to come across a reliable forecast of rain. People apparently do experience actual changes in taste, i.e. changes in desire that are not triggered by changes in belief. Moreover, not all such changes appear random; in some cases, at least, the desire revision appears to be the result of conscious deliberation. In this case, we would hope that our favoured version of the diachronic DBA does not in fact do for desire what it does for belief. Indeed, it seems most plausible that a rational agent should at least have the option of planning genuine

changes in taste.

In the remaining Sections of the chapter (3.3 & 3.4), I consider how we might make sense of planned changes in taste.[69] Our agent might, say, currently prefer driving to work than riding their bike, but they aspire to be the sort of person who cares more about reducing air pollution. In Section 3.3, I examine in some detail the concept of "higher-order" preference (or "higher-order" desire), because I think this is central to a number of puzzles in the modelling of desire, including the one I'm focusing on in the chapter. I take the position that "higher-order" preferences are really just ordinary preferences about one's desires/utility function at future times. It seems very plausible that such preferences about the future play a role in any conceivable planned change in taste. While we may not be able to capture the mechanics of any such change, I think there is a lot to be said about the phenomenon in the context of dynamic/sequential choice. For instance, if I choose one strategy over another because I believe that the former makes it more likely that I will develop a taste for jazz music (and with all other things being equal), then it can reasonably be said that I plan a change in desire in favour of this style of music. I pursue this line of thinking in Section 3.4, giving particular attention to the role that "higher-order" preference can play in sequential decision-making. So having given an overview of how I will pursue the issue of planned changes in taste, let me first examine what the diachronic DBA has to say about the rationality of this kind of desire change.

3.2 The diachronic DBA: an asymmetry between belief and desire

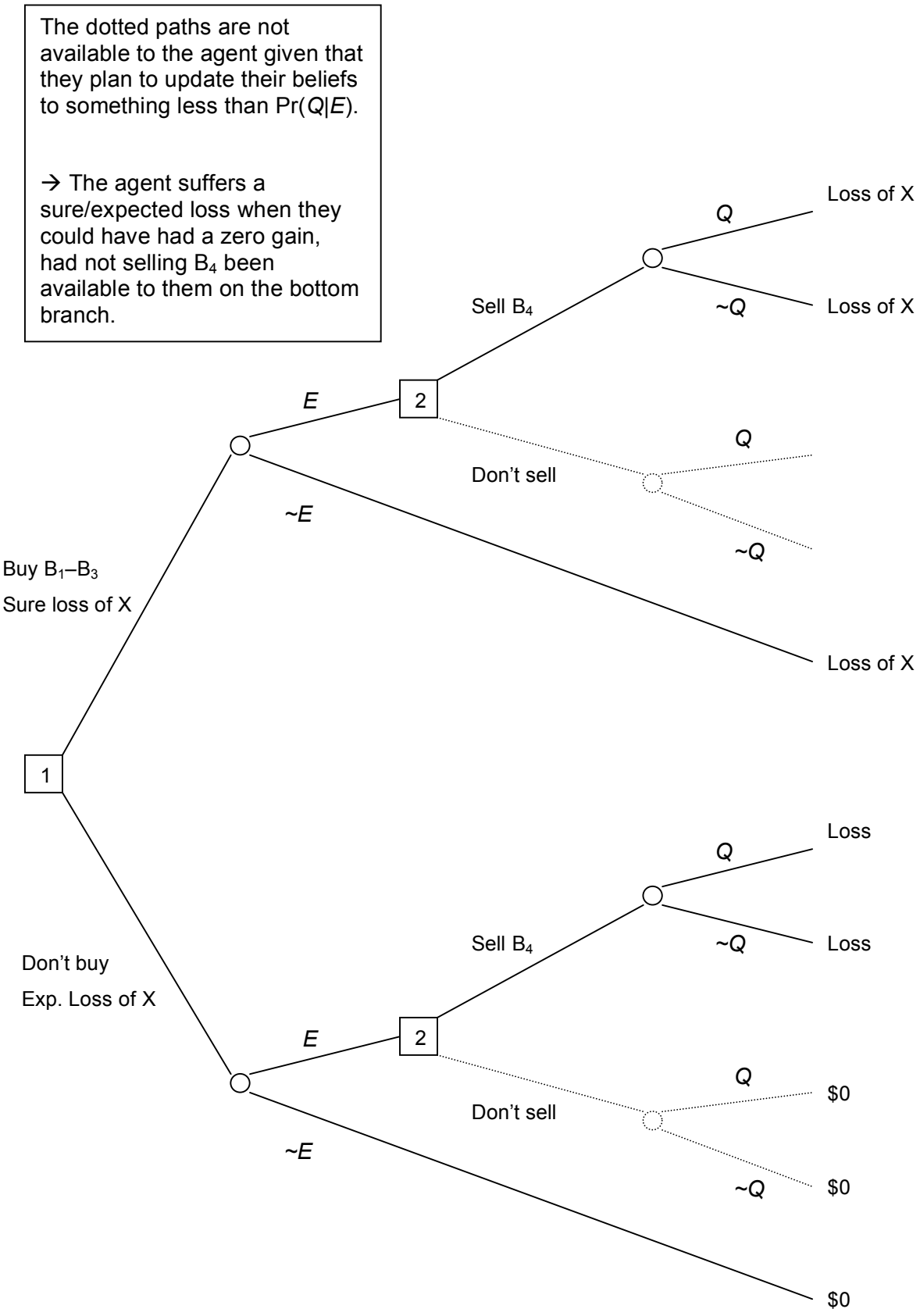
I have argued in Chapter 2 for a particular reading of the diachronic Dutch book argument (diachronic DBA) for strict conditionalisation. This reading of the DBA appeals to the sequential-choice framework, and specifically to a context in which what the agent learns about the world can be summarised in a proposition that they

[69] At present, there does not seem to be any well-accepted model in the economic or philosophical literature for handling this sort of desire change.

come to regard as certain. The (sophisticated) diachronic DBA shows that the non-conditionalising agent sometimes chooses a strategy that is dominated by another strategy. Admittedly, the dominating strategy is, in these instances, not a live option for the agent, but its inaccessibility is entirely due to *the agent's own updating plans.* In other words, a sure gain may be out of reach for the non-conditionalising agent because they plan to update their beliefs in such a way that they do not expect to make the requisite choices at various points in the sequential decision. The path to the better outcome is simply not a possibility, given the agent's planned credence functions at future times. Figure 3-1 depicts this diachronic Dutch book lesson.

I have argued that this is the only defensible reading of the diachronic DBA for strict conditionalisation. Unlike other versions of the story, it allows the agent all the sequential planning tools that are on offer. We assume that the agent assesses strategies in a "sophisticated" fashion, that is, that the agent is realistic about what path they will pursue at later choice nodes, and so doesn't make the mistake of setting off on a journey that is expected to fail. When the diachronic DBA is told in this way, it is not undermined by cases where the agent predicts that their beliefs will change (despite their best intentions) in a manner contrary to conditionalisation, or by cases where the agent has pragmatic reasons for consciously planning to update in some alternative way. The sophisticated diachronic DBA can handle these complications because it makes clear that only sure losses resulting directly from an agent's updating plans matter. Any gains that the agent effectively forfeits because she predicts that her conditionalising plans will go astray are regrettable, but not relevant. What matters is that the agent assesses strategies accurately (which is assured if the agent is a sophisticated chooser), and that they don't forfeit gains on account of their premeditated belief-updating plans.

Figure 3-1

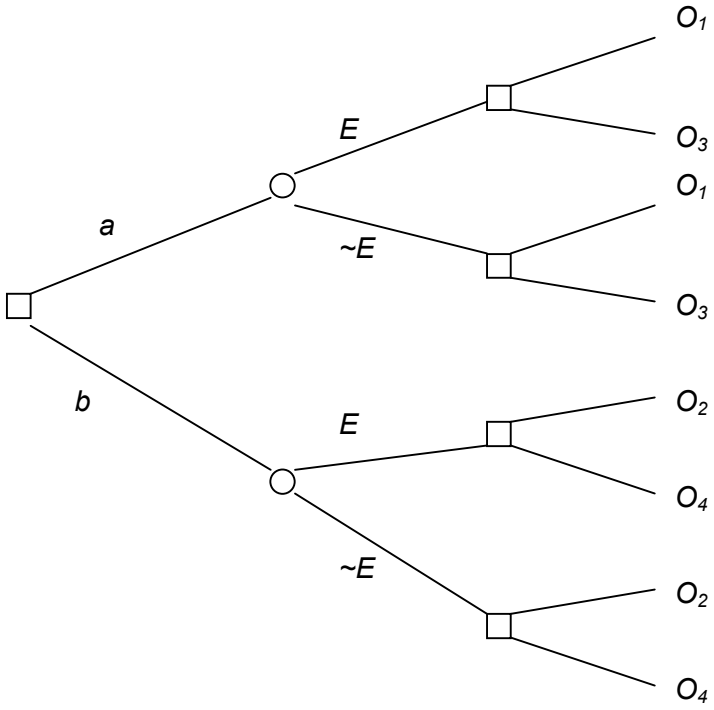

The dotted paths are not available to the agent given that they plan to update their beliefs to something less than Pr(*Q*|*E*).
→ The agent suffers a sure/expected loss when they could have had a zero gain, had not selling B_4 been available to them on the bottom branch.
Buy B_1–B_3
Sure loss of X
Don't buy
Exp. Loss of X
1
E
~*E*
2
Sell B_4
Don't sell
Q
~*Q*
Loss of X
Loss of X
Loss of X
Loss
Loss
$0
$0
$0

There remain ways to dispute the diachronic DB justification of conditionalisation as the rule for updating beliefs. But the sophisticated version of the argument is at least a very credible story. Let us now turn to the desire side of the decision-modelling story. I have introduced the diachronic DBA because it presumably says something about updating preference or desire, if its conclusions about what are appropriate plans for updating beliefs are valid. If we want to address the question of how a rational agent should plan to update their preferences or desires, then the diachronic DBA is surely a good place to start. Presumably, some version of the argument recommends that an agent should update their preferences/desires in line with their current conditional desires. For ease of reference, let us call the plan to update one's current desire/utility function in response to new evidence to a function that matches one's current conditional desires, "conditionalisation$_D$". That is, if an agent subscribes to "conditionalisation$_D$", then they plan to update their current desire/utility function U upon learning the truth of some proposition E to a new desire/utility function given by $U_{new} = U(\cdot|E)$. Now if there is going to be a desire-change version of the diachronic DBA, it must recommend that an agent subscribe to conditionalisation$_D$, on pain of sure loss. The question is what sort of sure losses these will amount to. If the threatened sure losses do not seem compelling, then this might be reason to be suspicious of the whole diachronic DB story, as it applies to belief as well as desire. On the other hand, we might conclude that the diachronic DBA is more pertinent to belief change than desire change, and that this amounts to, or otherwise is indicative of, a significant asymmetry between belief and desire.

The diachronic DBA for updating beliefs via conditionalisation assumes that the value of monetary amounts in the story is constant, or in other words, independent of any evidence received. (It is actually only necessary to assume that the conditional desirability of the monetary amounts is constant, but it is reasonable to think that even the unconditional value of money remains constant in the confined betting circumstances that feature in the DB story.) The uncomplicated nature of desire in the DBA allows conclusions to be drawn about planned belief change. I want to consider whether we might be able to reverse this scenario, and tell a diachronic DB story about planned desire change. The idea is to fix the belief side of the story, and see what conclusions can be drawn about updating desire. (We can just assume that the

agent updates their beliefs according to strict conditionalisation.) The problem depicted in Figure 3-2 below helps illustrate the desire version of the DBA that I have in mind.

Figure 3-2



Let us assume that according to the agent's current desire function U, the final outcomes (O_1, O_2, O_3 and O_4), conditional on E and conditional on $\sim E$, have comparative desirabilities as follows: $[U(O_1|E) = U(O_1|\sim E)] > [U(O_2|E) = U(O_2|\sim E)] > [U(O_3|E) = U(O_3|\sim E)] > [U(O_4|E) = U(O_4|\sim E)]$. We will further assume that, whether the agent learns the truth of E or whether they learn $\sim E$, then they will update to a new desirability function U_2 for which the following inequalities hold: $U_2(O_2) > U_2(O_3) > U_2(O_4) > U_2(O_1)$. In this case the sophisticated agent will choose strategy b, because this way they expect to end up with O_2, whether E or $\sim E$, whereas if they had chosen strategy a they would expect to end up with O_3, whether E or $\sim E$. According to their current desirability function U, however, the agent misses out on a sure $U(O_1|E) - U(O_2|E)$, due to the fact that their desire-updating plans do not permit the choice of O_1 at the second choice node .

The above is far from being a general desire-change version of the diachronic DBA—to begin with, I am assuming here that the agent's current desires for the outcomes, conditional on E, are identical to their current desires for the outcomes conditional on $\sim E$. (I make this assumption in order to generate a *sure* loss.) I am also assuming that there are three outcomes (O_2, O_3 and O_4) with unchanging conditional value that can be used to test the agent's plans for updating their desire for a particular outcome (O_1) in response to some evidence proposition (whether it be E or $\sim E$). The three extra propositions have desirability values that lie between the current conditional desirability for outcome O_1 given E (or $\sim E$), and how the agent plans to value outcome O_1 once they have learned the truth of E (or $\sim E$). (These intermediate outcomes might be monetary amounts.) But this is a good enough sketch, I think, for the purpose of examining the kind of conclusion that a diachronic DBA would yield with respect to desire: An agent who plans to update their desire function contrary to their current conditional desires (contrary to conditionalisation$_D$) can forfeit sure gains by the lights of their own *current* desire function. While I have not attempted to achieve any generality in the sequential-choice problem depicted above, it is reasonable to think that an agent who *does* subscribe to conditionalisation$_D$ cannot make an effective sure loss by the lights of their current desire function. In any case, the question remains as to whether this is the sort of sure loss that we should be alarmed about.

Assuming the above story really is a suitable kind of analogue of the diachronic DBA for conditionalisation as the rule for updating beliefs, it does not seem to have the punch of its belief counterpart. The agent loses a certain amount by the lights of their current conditional desires, but why should this matter if the agent's desires at the time of actually receiving the outcome are satisfied? As mentioned, the diachronic DBA for belief change assumes that the value of the monetary outcomes is constant at different times in the sequential decision problem. So any sure loss that figures in the story is at least a straightforward loss. The ordering of outcomes is not something that is at issue because it is invariant with time. (What *is* controversial in the belief DB story is whether the better option that is inaccessible to the non-conditionalising agent should be counted as a *live option.*) I think the above sketch shows that the desire analogue of the diachronic DBA does not really succeed as a defence of updating

desire in accord with one's current conditional desires. It is just not clear how we should recognise a sure loss when an agent's very opinions about the desirability of prospects are expected to change.

In fact, I think it is to the credit of the sophisticated version of the diachronic DBA (the version that I endorse) that this style of argument doesn't look so good as proof that conditionalisation$_D$ is the only legitimate way to plan on updating desire. Intuitively, it seems far from irrational for an agent to plan a change in their basic tastes or conditional desires. How we should actually model this kind of planned desire change is, I think, the more difficult question. In the next couple of sections, I will explore this issue. It seems most plausible that "higher-order" preferences motivate such desire changes. I therefore begin my investigation with a detailed examination of what "higher-order" preferences can plausibly amount to. I do not hope to uncover any specific mechanism for planned transformations in basic tastes or desire. The aim is, rather, to provide a more indirect account of the phenomenon. To this end, I later consider the role of "higher-order" preference in sequential decision-making, and how we might come to a broader understanding of what constitutes an "updating plan" in this context.

3.3 Making sense of higher-order preference

The concept of "higher-order preference" is interesting in its own right, but, as just mentioned, I focus on it here because I think an account of higher-order preference is key to the puzzle of planned changes in basic tastes or desire. Both Jeffrey (1974) and Sen (1977) have written on the topic and it will be useful to start with their accounts. Sen (in particular) examines higher-order preference with the greater aim of understanding the kinds of personal struggles that an agent experiences when their selfish, short-sighted wants conflict with their moral or other long-term commitments. Jeffrey focuses rather on the pure logic of higher-order preference. Indeed, we require an account of how higher-order preference claims can consistently mesh with regular

first-order preference claims, and what impact this mixture of preference types has on decision-making. I argue that neither Jeffrey's nor Sen's account of higher-order preference can provide satisfactory answers to these questions. I thus develop an alternative account of higher-order preference, borrowing from the work of Bolle (1983). While somewhat deflationary, my account makes "higher-order" preferences comprehensible in terms of standard decision terminology. Furthermore, my account allows "higher-order" preferences to play an explanatory role when it comes to understanding an agent's inner battles.

I first consider Sen's (1977) remarks on higher-order preferences. It is easy to misread Sen because his use of preference terminology is, in my opinion, not standard. To begin with, Sen does not identify an agent's preferences with their choice function. It seems that he stresses the distinction between preference and choice so that he can depict an agent as having multiple preference orderings. Sen claims that the agent might have one preference ordering over options based entirely on personal welfare considerations, another that is based on both personal welfare and the welfare of people that matter to the agent, and yet another preference ordering that reflects the agent's purely moral judgments. Indeed, the agent may have any number of preference orderings that each reflects their evaluation of options with respect to specific aspects of the world. I do not think there is any great problem with entertaining different preference orderings of these kinds,[70] but at the end of the day we need an account of how the agent goes about determining their final choice function or all-things-considered preferences (which may well be a partial ordering). The different preference orderings that Sen is concerned with do not seem to be like the separate criteria that are appealed to in the context of multi-criteria decision analysis. (Ideally these criteria are weighted to give an overall preference ranking.[71]) For starters, the first two preference rankings referred to above overlap in the sense that they both involve considerations of personal welfare, so assigning each a

[70] I note that the issue depends on how we want to understand ordinal preferences, and how we think they can be elicited. If an agent's preferences can only be identified via their choice behaviour, then arguably we can only ever determine an agent's all-things-considered preference ranking.

[71] I note that Levi (1986) argues that the different criteria may sometimes be incomparable, in which case we cannot assign them relative weights, even in the ideal scenario.

weighting would not be a suitable way to combine them (because the personal welfare considerations would be "double-counted"). Perhaps moral and self-centred considerations could be weighted against each other (if these were the only things that mattered to the agent), but this does not seem to be what Sen has in mind with respect to negotiating the two types of values. Sen is in fact rather vague on this issue; he simply says that commitment (or moral judgment) "drives a wedge between personal choice and personal welfare" (1977, p. 329).

Sen goes on to introduce higher-order preferences to this structure. He replaces the idea of singular criteria-based preference rankings with higher-order rankings of possible utility functions that essentially serve the same purpose. So, for instance, instead of having just one maximally moral preference ordering, the agent expresses their moral judgments through a ranking of possible preference orderings that each vary in their moral merit. To give another example, instead of having one preference ranking representing personal welfare considerations, the agent ranks all possible preference orderings according to how well they reflect differences in personal welfare alone. While this is an interesting way of depicting an agent's various judgments, I do not think that it sheds any light on the choice process. In fact, given that we can already represent an agent's personal welfare judgments by a single preference ordering over possible worlds or outcomes, higher-order personal welfare preferences seem redundant. Utility functions that are more faithful to the true personal welfare-oriented utility function are preferred. Nor has any headway been made here with respect to determining the agent's overall choice function. The problem remains that the agent must negotiate various conflicting higher-order preference orderings, one based on personal welfare, another based on moral judgment, and so on. What we are really interested in is the agent's actual all-things-considered preference ordering, and what sort of higher-order preferences are consistent with particular first-order preferences. Sen's account, as it stands, does not deliver this.

Jeffrey's account of higher-order preferences, in one sense at least, answers better to these demands. Jeffrey shows how first-order and higher-order preferences can mesh

together, essentially because he thinks they do not occupy completely distinct planes of desire. According to Jeffrey's model, an agent has preferences over propositions, and although these propositions are generally thought to describe possible external events, they might just as well concern the state of mind of the agent herself. So, for instance, the agent has some preference for the proposition "I prefer smoking to not smoking" relative to all other propositions in the set (assuming a complete ordering), including such propositions as "it rains tomorrow". The puzzle for Jeffrey is how propositions describing events and propositions describing preferences can fit together consistently in an agent's preference ordering (and associated utility function). For instance, Jeffrey does not think that the following ordering is legitimate (where desirability of propositions decreases as we go down the page):

~*S* pref *S*

S

~*S*

S pref ~*S*

(*S* might be interpreted as the proposition "I smoke" and "*S* pref ~*S*" stands for the proposition that the agent prefers *S* to ~*S*.) If all of the above propositions are genuine options, the ordering cannot be consistent because here the agent prefers smoking to not smoking, but they would prefer to have the opposite preference. Importantly, the agent prefers preferring not smoking to smoking to the event of smoking itself (apologies for the tongue-twister!) Interestingly, Jeffrey thinks a slight modification of the above preference ordering would be consistent:

S

~*S* pref *S*

~*S*

S pref ~*S*

The important difference for Jeffrey between the two orderings is that in the first, the higher-order preference for not smoking is more desirable than smoking itself, and in the second, the opposite is the case.

While Jeffrey brings higher-order preferences down to earth, so to speak, his model does not really illuminate the interplay between first-order preferences and these loftier ambitions. What we see are the results of the agent's struggle between "appetite and will" (Sen 1977). Just how the agent arrives at this final preference ranking remains mysterious. More importantly, I am not convinced by Jeffrey's analysis of the above preference orderings. It is hard to comprehend the significance of the relative ranking of the act of smoking itself and having the preference of not smoking over smoking. Indeed, the interplay of these two kinds of preferences is puzzling. Perhaps Sen was right that first-order and higher-order preferences indeed occupy different planes (as their names suggest). In fact, I think the general problem here—the reason why it is difficult to compare propositions about events and propositions about one's own preferences in the way Jeffrey suggests—is that this handling of higher-order preferences seems to be at odds with Jeffrey's own evidential decision theory. According to the evidential decision model, propositions have value according to their news value; it is a question of how much satisfaction the agent would derive from finding out that a proposition is in fact true.[72] But in the above preference orderings, it seems that the agent already knows whether the propositions describing their current preferences are true or not. In both cases, "*S* pref ~*S*" is true and "~*S* pref *S*" is false, by virtue of the placement of *S* above ~*S* in the preference ranking. So contrary to Jeffrey's analysis, this means that "*S* pref ~*S*" and "~*S* pref *S*" cannot be located just anywhere in this agent's preference ranking. The former should have the same desirability as any tautology, and the latter should have the same desirability as any contradiction.

Jeffrey himself notes that given his commitment to evidential decision theory, his

[72] In what follows, I assume that we want our account of higher-order preference to be accommodated by the evidential decision model. Alternatively, we might criticise the evidential decision model for holding that all propositions that are known to be true should have the same news value (i.e. the value of no news).

account of higher-order preferences relies on the agent being somewhat unsure about their first-order preferences. (The agent is indifferent between all propositions that they know for sure are true, because in these cases, there is no sense in which the agent derives pleasure from finding out something new about the world.) While Jeffrey is not unaware of this issue, I do not think his response is adequate. It is certainly reasonable to think that an agent may be unsure about aspects of their own mental state; we all appreciate the difficulties of introspection. Indeed, it is commonly argued that an agent cannot intuit the relative strengths of their own beliefs and desires, hence the decision theory representation theorems that require only ordinal preferences as input. And there is certainly a place for arguing, as Jeffrey suggests, that an agent can even be unclear about their ordinal preferences—that preferences do not simply equate to, or else cannot simply be ascertained from, observed choice, as the behaviourist would have us accept. In the context of Jeffrey's discussion of higher-order preferences, however, scepticism about the accessibility of an agent's preference ordering seems unwarranted. The examples given above are candidate preference orderings, and *in this context*, any propositions that describe the preference ordering itself must have no news value. In other words, here we are simply assuming that the agent has a specifiable preference ordering. Perhaps this is indeed a point that Jeffrey would dispute. He might be of the opinion that agents do not generally have a specifiable preference ordering, but rather have indeterminate preferences, or are at least uncertain about what their preference ordering is. While others have developed thorough accounts of indeterminate preference, however, this is not a position that Jeffrey consistently maintains.[73] In any case, it seems an overly restrictive account of higher-order preferences to have them depend on the agent having indeterminate first-order preferences.

While Jeffrey might be right that an agent can never narrow their preferences down to a single ordering, I think there is a better way to understand the idea of higher-order preferences. I take up a suggestion of Bolle's (1983, p. 197) that "second order preferences are derived from first order preferences".[74] In fact, I will argue that a

[73] Notably, Levi's (1986) decision theory can accommodate indeterminacy.

[74] Bolle is led to an account of higher-order preference in his investigation of how ethical

"derivative" account of higher-order preferences (which I will refer to as "quasi-higher-order preferences") can account for our intuitions, while avoiding the problems associated with Jeffrey's and Sen's respective models. Like Sen, the derivative account of higher- (second-) order preference understands these preferences as occupying a different plane to first-order preferences. What we are talking about is a preference ordering over primary preference orderings/utility functions that does not lend itself to being integrated within these very orderings, contrary to Jeffrey's proposal. But here is the catch: such second-order preferences are just fictions. We *think* that we have these higher-order preferences because what we *do* have are first-order preferences over our possible future preference functions. (In other words, the idea of higher-order preference is derived from first-order preferences about the future.) In Jeffrey's language, we may prefer that the proposition "tomorrow I will prefer working efficiently to surfing the internet" come true, rather than its negation. But I claim that there is no sense in having propositions of the kind "I *now* prefer working efficiently to surfing the internet" in my current utility function. There is still the possibility that such preferences occupy a separate, higher plane, but this would not be very useful because it is unclear how a genuinely higher plane of preference could have any impact on an agent's choices.[75] At the very least, the relationship between the two is unclear.

If quasi-higher-order preferences are indexed to future times, then they can be accommodated by the evidential decision theorist without them having to make sweeping claims about whether an agent is able to identify, or be identified with, a single preference ordering.[76] It is much more likely that the regular agent has some

---

judgments influence an agent's preference ranking and associated choices. He argues for a very specific model of the uptake of social norms, but I agree with his general stance on second order preferences, as well as his comments on the explanatory power of dynamic decision models.

[75] I admit the possibility that genuine higher-order preferences might affect the way an agent perceives the possible outcomes before them. For instance, "I smoke at time *t*" might be more properly described as "I smoke at time *t* and contradict my higher-order preference for not smoking".

[76] As noted, we might think that there are some serious problems with the way that prospects are evaluated in the evidential decision theory framework. It could be argued that an agent *should* be able to differentiate between propositions that are known to be true. In this case, maybe it is perfectly reasonable for an agent to have opinions about their own current

uncertainty (however slight) about their future preferences, as compared to their current preferences. Indeed, it would be almost foolish for an agent to be certain about their future preferences, because any number of events could conceivably cause an unpremeditated change in outlook. Moreover, there is another sense in which an agent's preferences with respect to her *future* utility function, as opposed to her *present* utility function, fit naturally within her present preference ordering. There will not be the same problems of consistency that Jeffrey was concerned with, as per my earlier discussion. For instance, it does not really matter how the following propositions are ordered in a preference ranking, provided the quasi-higher-order preferences have a different (future) time index. Indeed, the following ordering is perfectly legitimate:

$(\sim S \text{ pref } S)_{\text{tomorrow}}$

S

$\sim S$

$(S \text{ pref } \sim S)_{\text{tomorrow}}$

It is not even clear that the preference orderings an agent favours at some future time must satisfy the usual axioms. Must we consider the agent irrational, for instance, if they desire a non-transitive future utility function over a transitive one? Surely some non-transitive utility functions can be considered better than some transitive ones. In any case, there will be only limited restrictions on the relationship between an agent's current preferences and their hopes about their future preferences.

Quasi-higher-order preference can also make sense of the struggle between appetite

---

preference ordering/utility function (as per Jeffrey's account of higher-order preference). Moreover, it seems plausible that I should have opinions about the past—I may regret something that happened yesterday, or I may wish that I had held different preferences a week ago. I do not dismiss these sorts of arguments, but I think it is worth considering an account of higher-order preference that can be accommodated by evidential decision theory. Furthermore, the "derivative" account that I favour ("quasi-higher-order preference") seems sufficiently powerful for analyzing sequential decision problems and the concept of planned changes in taste.

and will that Sen, in particular, alludes to. Indeed, this is intuitively why we are led to talk about higher-order preferences—so that we can better understand akrasia or weakness of will on the one hand, and on the other an agent's concern for self-improvement. Importantly, I don't think a genuine higher-order preference model succeeds in this respect. It is unclear how genuine higher-order preferences could possibly interact with first-order preferences. For starters, the two kinds of preference are supposed to exist on separate planes. Secondly, an agent's first-order preferences *by definition* ground their choices at a particular time; so there is no sense in which the process of choosing how to act involves some further struggle with higher-order commitments. Recall that Sen does not use preference terminology in this way. His talk of preferences is reminiscent of, but not to be identified with, multi-criteria decision analysis—the agent may have various rankings over propositions depending on what criterion or aspects of the world they choose to focus on. As discussed, in the end Sen owes us an account of how an agent should determine their all-things-considered preferences. As I see it, this can only amount to some kind of combination of first-order preference rankings (each based on a particular "criterion"). Mixing and matching different orders of preference just seems unnecessarily complicated.

When it comes to the quasi-higher-order preference model, the struggle between appetite and will can be thought to happen at the point where an agent shifts from one utility function to another with the passage of time. What is in tension is the agent's former idea about the kind of person they want to be in the future, and the inertia of their actual desires. It is plausible that the agent's new desires are some function of their old desires and their old hopes about how their desires might change. I do not hold too much hope, however, that such a function could actually be formalised, or the transformation mechanism made more specific than what I have just outlined. But we need not try to understand changes in desire in such direct terms. Quasi-higher-order preferences can in any case shed light on the agent's decision-making process when the strategies that are on offer are predicted to lead to distinct changes in the agent's utility function. Clearly, the agent's preferences regarding their utility function at future times will have an impact on their current choice of strategy. In this way, we can model a more indirect kind of planned desire change. It is not that the agent plans the actual mechanism by which they will undergo a change in taste;

rather, they give themself the best possible chance of undergoing a preferred desire change by pursuing a path that makes this change most likely.

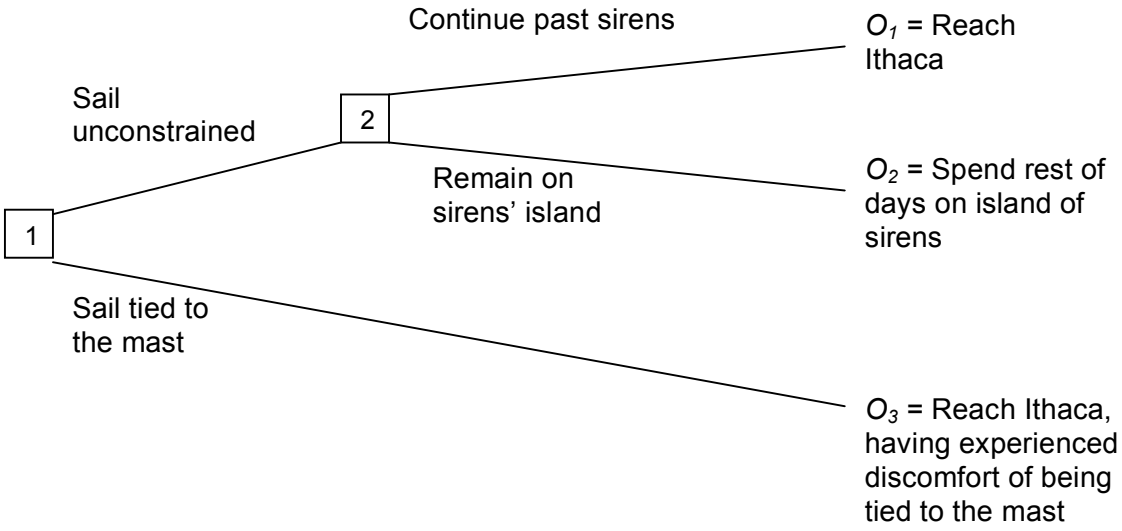
Just how we should employ quasi-higher-order preferences to assess decision strategies that involve changes of taste is no straightforward matter. The evaluation of strategies in these kinds of circumstances is not something that has been particularly well addressed in the sequential-choice literature. Both Hammond (1976, 1988b, 1988c) and McClennen (1990), for instance, suggest that the ideal agent should not expect to undergo changes in their basic tastes. Levi (1991) acknowledges predicted desire changes of this sort, but does not address how such changes might complicate the assessment of strategies. This is the question that I will pursue in the next section. If we want to understand an agent's choice to pursue a predicted change in desire, we need to be clear about how "higher-order" preferences (of the derivative or quasi kind) and regular first-order preferences might play out in the sequential-choice context.

3.4 "Higher-order" preferences and the evaluation of strategies

In order to examine the role of "higher-order" preference in evaluating sequential-choice strategies,[77] it will be useful to revisit Ulysses' well-known predicament, as illustrated in Figure 3-3 below. As the narrative goes, Ulysses is on route to his home kingdom of Ithaca. The decision he faces is whether to instruct his sailors to tie him to the ship's mast just before they sail within earshot of the island of the sirens. Either way, Ulysses predicts that he will undergo an uncontrollable change of taste at this point of the journey. While he now prefers going home to Ithaca more than spending the rest of his days on the island of the sirens, Ulysses has reason to believe that in the presence of the sirens' sweet song, he will in fact prefer spending the rest of his days on the island to continuing his journey homeward.

[77] From now on, assume that "higher-order" preferences are quasi-higher-order preferences.

Figure 3-3



Now it is generally agreed that Ulysses should indeed instruct his sailors to tie him to the mast, despite the discomfort associated with this. The assumption is that Ulysses currently prefers sailing home to Ithaca and perhaps suffering a bit of humiliation and rope-burn to staying on the island of the sirens. More precisely, we assume that Ulysses' preference ordering is $O_1 > O_2 > O_3$ at node 1 in Figure 3-3 above, and $O_2 > O_1 > O_3$ at node 2. That is, Ulysses expects his attitude towards staying on the island of the Sirens to be different at the initial and later time (even though there will be no new evidential input). In assessing the strategies before him at the initial node, it is thus important that Ulysses acts as a sophisticated chooser who is realistic about likely consequences. Recall that the sophisticated chooser essentially works backwards through a dynamic decision problem in order to determine the real consequences of a current choice of strategy. (I argued in Chapter 1 that only the sophisticated approach to dynamic decision-making is defensible.) Ulysses cannot simply hope that he will be able to overcome the song of the sirens and sail on past their island without the constraint of ropes. He must take seriously his own predictions about how his preferences are likely to change (even if the change is somewhat mysterious to him), and what the upshot of such changes will be in terms of his choices at future times.

Sophisticated reasoning reveals then that Ulysses has a choice between not being tied to the mast, which will inevitably lead to his spending the rest of his days on the island of the sirens, and suffering the humiliation of being constrained to the mast with the outcome that he is likely to eventually reach Ithaca. (In other words, the choice is effectively between outcomes O_2 and O_3.) Given that Ulysses currently prefers the latter consequence, most think he should go with the mast-tying strategy. While I agree with this popular diagnosis of Ulysses' predicament, I think it involves various assumptions about how an agent should handle conflicts between their present preferences and their predicted future preferences, conflicts that do not seem to have been acknowledged in the dynamic choice literature. My aim here is to make these assumptions more transparent. In the process I will consider slight variations of Ulysses' story that might lead us to draw different conclusions about how he should evaluate the strategies before him.

The main issue I am interested in is why it is generally supposed that Ulysses should not take seriously his potential future preferences upon hearing the song of the sirens—why shouldn't Ulysses respect a future desire to remain on the sirens' island? A general attitude among commentators seems to be that this preference change is only temporary and so not truly indicative of Ulysses' character. Indeed, McClennen (1990, pp. 202, 232) points out that Ulysses' change can be considered not a genuine preference change at all, but rather a sudden succumbing to weakness of will. (According to McClennen, this position is suggested by a number of dynamic decision theorists, including Strotz (1956), Hammond (1976) and Elster (1979).) The idea is that Ulysses still prefers to go home to Ithaca when he hears the sirens, but this option is psychologically unavailable to him. If this is right, then my general concerns about the evaluation of strategies given potential conflicts between current and future preferences will not be warranted in Ulysses' case. According to this reading, Ulysses' preferences are in fact stable, in which case his evaluation of final outcomes at the initial node should be straightforward. While I do not want to dismiss the possibility of there being psychological barriers to acting on one's true preferences, I think this reading of Ulysses only serves to push aside important questions about strategy assessment that only arise when an agent predicts a change in basic

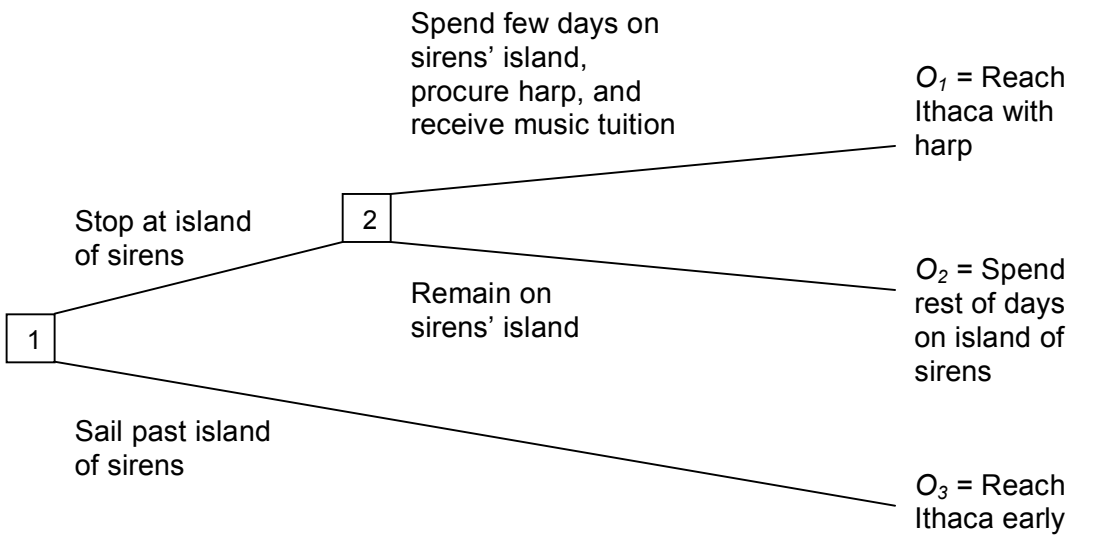
preferences.

While genuine changes of taste may be somewhat mysterious and non-rational, we regularly witness them, and I think it is at least plausible that this is indeed what Ulysses predicts to happen, as opposed to an episode of akrasia. Furthermore, I don't think it is entirely fair to describe the potential preference change as merely "temporary", implying that the resulting subjective state need not be taken seriously. This is not to say that there is no good reason for regarding Ulysses' desire to stay with the sirens as rather feeble. Indeed, so far as the story goes, Ulysses has this desire only when he is within earshot of the sirens' sweet song. Many would regard this as too fragile a preference state to take seriously in decision-making. Having said this, however, I draw attention to the fact that if Ulysses did decide to sail unconstrained—the option that would likely result in his staying on the island of the sirens—the desire to stay there might yet be classed as flimsy, but it seems confused to call it temporary. Once there, Ulysses would presumably hold the desire to remain on the sirens' island for the rest of his days. Surely after 50 or so long years, this preference would no longer be considered transient. My point here is that, while we may share the overwhelming intuition that Ulysses should pay little heed to what his preferences dictate while under the "spell" of the sirens, this change in outlook need not be regarded as temporary, and it is not necessarily undesirable either. What seems important is whether, in this situation, Ulysses himself prefers his tastes as they stand to the change he would undergo if exposed to the sirens. The story indicates that Ulysses does favour his tastes as they stand, in which case, he is surely right to opt for being tied to the mast.

It just so happens that in Ulysses' case, Ulysses' own preferences about his future tastes favour him making the commitment to sail home to Ithaca. But "higher-order" preferences need not always adjudicate in favour of the agent's utility function remaining stable. Let me now consider a scenario in which an agent expects his or her core tastes to change for the better. I will argue that in such cases, it is not so obvious how the agent should assess competing strategies. Should final outcomes be evaluated on the basis of the present utility function, or rather on the basis of some more

preferred potential future utility function? Consider the decision problem in Figure 3-4 below. It is a variation on Ulysses' story. Ulysses again predicts that his tastes will be transformed by the sirens' sweet song, but this time he is of the opinion that it will be for the better. Let's say that the choice is between sailing straight home to Ithaca, or first spending a couple of days on the island of the sirens. Ulysses in this case thinks the outcome of visiting the sirens' island is that he will be able to acquire a unique harp, as well as some music tuition that will transform his appreciation for harp music. The negative aspect is that he will lose time in getting home to Ithaca, and the harp will be bothersome to transport.

Figure 3-4



The important thing to note in this example is that, given his current disinterest in harp music, Ulysses currently prefers going straight home to Ithaca without the instrument. (At node 1 in Figure 3-4 his preferences are as follows: $O_3 > O_1 > O_2$.) At the same time, our hero predicts, however, that if he were to stop at the island of the sirens and receive some music tuition, he would regard procuring the harp as a worthwhile deviation from his path homewards. (In other words, at node 2 his preference ordering will have changed to $O_1 > O_3 > O_2$.) The predicted change in preference is not the kind that would result from learning some new information—Ulysses' attitude towards the harp would not be different were he to know for sure,

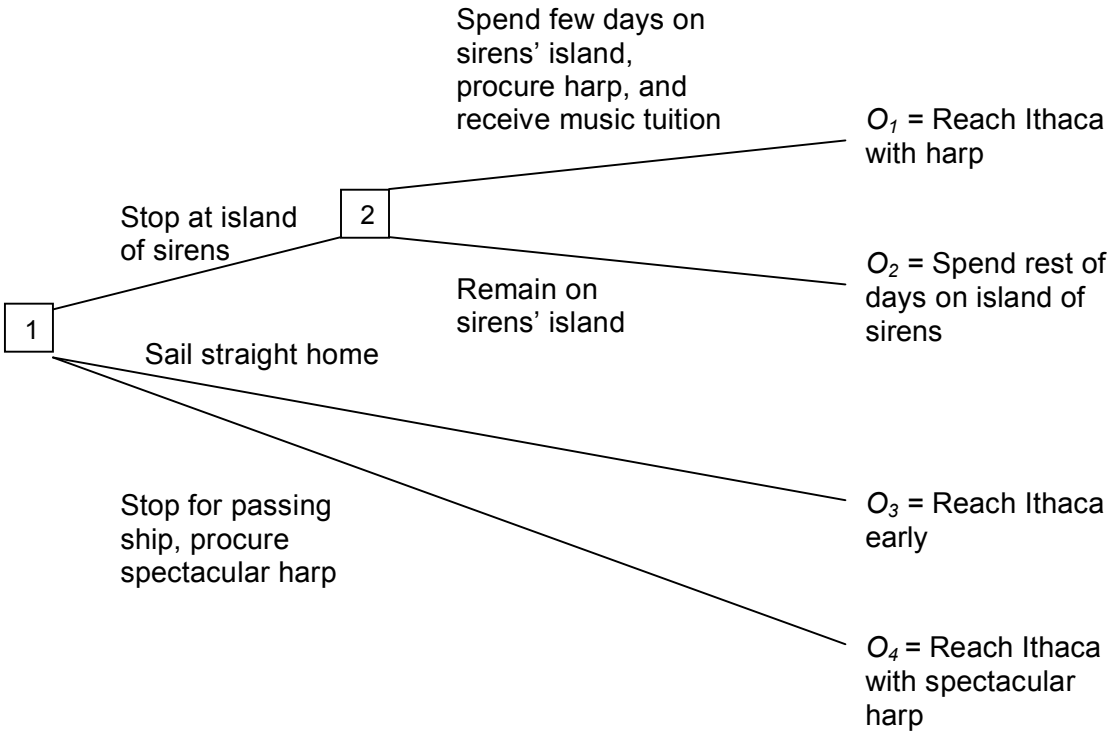
say, that it was the result of 1000 hours of careful craftsmanship, as opposed to being a 2-day knock-up job. Rather, Ulysses predicts a genuine and somewhat inexplicable change in taste due to a new experience. The intuition here, or mine at least, is that Ulysses should make the effort to get the harp. But interestingly, according to this way of modelling the scenario (I will go on to discuss variations in the model later), this means that Ulysses should choose according to his future preferences, rather than his present preferences. The question I pose is whether this is plausible, and how it might complicate the study of dynamic choice.

Comparing the two Ulysses stories above, we might come to the conclusion that an agent should assess decision strategies according to the utility function that they would most prefer to end up with. In the example depicted in Figure 3-3, Ulysses prefers his current tastes to those he would endorse were he to remain on the island of the sirens, so he should choose his strategy according to what his current preferences dictate—he should opt for being tied to the mast. In the second problem depicted in Figure 3-4, on the other hand, Ulysses prefers the more cultured preferences he would acquire were he to temporarily visit the island of the sirens, so it seems that he should evaluate the strategies from this perspective, and stop a few days with the sirens on his way home to Ithaca. I do not think this amounts to asking Ulysses to choose against his own current wishes. After all, Ulysses' current wish is to better appreciate harp music in the future. If he predicts that taking a particular strategy will result in him acquiring such preferences, then it seems plausible to assess final outcomes in accordance with these future improved preferences. As mentioned, in the problem in Figure 3-4, it is a difference between Ulysses pursuing the strategy that leads to outcome O_3, because that is what he now prefers, and pursuing the strategy that leads to outcome O_1, because that is what he *would* prefer were he to go down this path.

While the above might seem a reasonable recipe for assessing strategies when it comes to these fairly straightforward examples, however, we are going to run into problems when the decision scenarios become more complicated. I am thinking of possible cases in which the agent prefers the utility function that they would acquire if they pursued a particular strategy, but the utility function in question recommends

choosing an alternative strategy, so the agent would never in fact acquire these preferences. More generally, if a number of possible strategies all lead to distinct changes in the agent's tastes, and if the ranking of final outcomes varies with these differing tastes, then what perspective should the agent assume when deciding how to act? The problem depicted in Figure 3-5 illustrates the kind of complexities that might arise.

Figure 3-5



Ulysses' predicament is essentially the same as that depicted in Figure 3-4, but this time there is an extra option—Ulysses can stall his journey homeward in order to wait for a passing ship that carries on it the most beautiful harp of all. He will be able to procure this harp, but the loss of time will mean that he won't be able to also visit the island of the sirens, and so he will miss out on the music tuition. In this case, Ulysses' most preferred future utility function is the one that he would acquire were he to visit the island of the sirens. The complication, however, is that this utility function

recommends that Ulysses wait for the passing ship; his preference ordering post-music tuition is expected to be $O_4 > O_1 > O_3 > O_2$. But if he were to wait for the passing ship, he would never acquire the coveted preference for arriving home in Ithaca with the most beautiful harp of all; he would continue to hold his current preference ordering of $O_3 > O_4 > O_1 > O_2$. In this case it seems plainly misguided for Ulysses to pursue the strategy leading to O_4.

While it might be objected that the example described above is rather crude, the problem is only intended to illustrate a logical possibility. There will be a problem with my suggested approach to evaluating strategies if it is ever the case that the agent prefers the utility function they will acquire if they pursue one strategy, but this utility function in fact recommends an alternative choice of strategy. We could try to come up with some refinements to the assessment process I have outlined thus far, but I think the potential for the sort of complexity illustrated in Figure 3-5 casts doubt on this whole general approach to evaluating strategies. It simply does not look plausible for an agent to evaluate strategies according to some utility function that they do not currently hold, even if they later hope to hold such a utility function. At the same time, surely an agent's evaluations of final outcomes should be sensitive to what their desires might be at the time in question. So it seems that we have a genuine puzzle here.

In what follows, I suggest how we should approach this puzzle. I have shown that it does not look promising for an agent to assess strategies according to some preference ordering or utility function that they do not currently hold, even if this utility function is the most preferred amongst those the agent could potentially hold in the future. Whatever the details turn out to be with respect to how an agent should evaluate strategies, where changes of taste are involved, it appears fundamental that such evaluations rest on the agent's *current* preference ordering or utility function. It is not just the difficulties highlighted in the decision problem in Figure 3-5 that caution against tampering with this principle. The idea that an agent's actions reflect only their current preferences is intuitively very compelling and basic to standard decision theory. Of course, in the discussion above I was trying to maintain a strong link

between an agent's current preferences and their evaluation of strategies. The idea was that an agent should defer to a future utility function just in case they have a current "higher-order" preference for these future tastes. This might be likened to a kind of reflection principle for preferences—the idea being that an agent should choose according to the preferences they would most like to have in the future, which means that the agent effectively acts as if they hold those preferences at the current time. But the short of it is that the agent does not actually have the desired future preferences at the current time. And my analysis of "higher-order" preferences shows that this does not make the agent irrational. Moreover, we are wise to be sceptical about it being a rational requirement that an agent act purely out of deference for a potential future self. A sensible agent will of course take future attitudes into account, but when it comes down to it, it is surely the contemporary agent, with all their existing quirks and foibles, who is the driver of the decision.

The way to take altered future preferences into account, I think, is just to include these kinds of psychological facts in the description of final outcomes. This may seem like an uninteresting solution to the problem of changes in taste, but of course that should not count against it. We are familiar with the fact that act outcomes must be described in all their relevant detail.[78] For instance, assume that some act leads to the outcome that I am at the beach with some drinking water and a large hat, while some other act also leads to me being at the beach, this time without the water and hat. If I feel differently about these two circumstances, it is no good representing them as the same outcome, i.e. "I spend the day at the beach". The presence of the water and hat makes a difference to me, and so must be acknowledged in the description of the outcomes. When an agent anticipates possible changes in taste or desire, they must attend even more carefully to describing outcomes in all their relevant detail. For instance, being at the beach with water and a large hat will mean different things depending on whether I expect to hold my current attitude of caution about sun cancer, or a possible future desire to have a tan. In general, the particular combination of psychological and physical properties of an outcome will affect the agent's final evaluation of it.

[78] In Chapter 1 (Section 1.5) I gave Joyce's formal statement of this requirement. I also pursue the issue further in the next chapter.

I note that even though an agent's utility or desire for particular outcomes will depend on what combination of properties is present, we can still speak of the utility or desirability of just one kind of property. As discussed in Section 3.2, Jeffrey's "axiom of desirability" holds that the overall value of a state of affairs or property, like going to the beach, amounts to a weighted average of the values of all the possible ways that this state of affairs could turn out to be true. Some of these "worlds" will include hat and water, and others will not. The same applies to psychological properties of outcomes. We can say that the strength of my "higher-order" desire for preferring tanning to sun-caution is just a weighted average of the value of the possible ways that this preference might be realised.

The fact that we can talk about the desirability of any general property does not remove the difficult problem of evaluating the component individual worlds or outcomes. This is a substantial issue in real-life decision-making, and yet it is not addressed by formal decision theory. As mentioned, the SEU representation theorems assume that the agent has existing raw preferences over possible worlds, and if these preferences obey the SEU axioms then we can represent the agent with a probabilistic belief function and an affine desire function. When we introduce the possibility that an agent may expect a change in their basic desires, and that this should be included in the description of final outcomes, we give the agent even more to think about with regard to their relative preferences for these individual outcomes or worlds. And yet the formal theory does not offer any assistance when it comes to these basic evaluations of outcomes. For instance, the problem rests with Ulysses to determine how much he currently values having a beautiful harp but a lack of musical sensibility at time *t*, as compared to having an ordinary harp but much greater appreciation for music. One might say that in making these kinds of issues a matter of an agent's subjective opinion, I have not really responded to the problem of how an expected change in basic desires should be dealt with in sequential decision-making. But I think it is important to at least get the basic framework right. I have argued that an agent can only ever assess strategies from the point of view of their current preference ordering, and that expected future preferences will be taken into account to the extent that they affect the agent's current evaluation of outcomes.

3.5 Planned changes in desire

We might try to say more about how an agent should assess outcomes—for instance, surely there are some commonsense principles governing the weight that should be given to one's future attitudes towards potential states of affairs, in the light of "higher-order" preferences. In the least, we might introduce some systematicity to the evaluation of outcomes by way of some ceteris-paribus-type rules. For instance, it seems reasonable to say that if an agent is confronted with two outcomes that are identical in every way except that one involves a more preferred future utility function, then the agent should prefer this outcome. This rule is fairly straightforward, but the conditions under which it applies are more limited than what might first appear to be the case. To give a specific example, it is not clear that the prospect of living in a small country town and having an appreciation of orchestral music should be thought better than living in a small country town without this desire, even if the agent in question generally aspires to develop this fine taste. The problem is that live orchestral music may not be available in the small country town, and so the associated desire may not be well satiated. It is only when "all other things are equal" that we have a clear-cut case in which an agent should prefer the outcome that involves a more coveted future utility function; and "all other things being equal" refers not only to the physical characteristics of outcomes, but also to how well future desires are satiated.

Though it may be hard to distinguish what an agent's motives are for valuing outcomes in the way that they do, and thus why they choose to pursue particular strategies, it is safe to say that expected changes in taste may contribute to these evaluations. This is to say that an agent may pursue a particular strategy at least in part because they predict that it will involve a desired change in their preference function. It seems reasonable to call this a "planned" change in preference. (I note that not all expected changes in preference are planned in this sense, because sometimes an agent will pursue a strategy that results in a less desired preference function, if the

other aspects of the final outcomes make up for the bad change in taste.) Of course, here we are not talking about the same kind of planned change as what the rule of conditionalisation is supposed to amount to when it comes to updating beliefs/preferences upon receipt of new evidence.[79] Conditionalisation is more like a conscious mechanism for revising one's belief/preference function at some point in time, when one expects to learn something about the world that can be summarised by a single proposition. In this discussion of changes in taste, there is no presumption that the agent will have conscious control over how their preferences will change. It is simply predicted that certain decision strategies will lead to a change in basic desires, due to some kind of experience. The agent can be said to plan such a change in desire to the extent that it is this feature of the final outcomes that leads them to pursue the strategy in question. In other words, the agent plans a change in desire to the extent that their "higher-order" preferences direct their choice of strategy.

3.6 Conclusions

I want to conclude this chapter, and indeed, the first part of this thesis, by comparing the dynamics of desire with that of belief. In many ways, the two can be treated similarly. When it comes to assessing decision strategies, for instance, an agent may have reason to predict any kind of perverse change in either their beliefs or their preferences. This is largely why I recommend the sophisticated approach to sequential choice, i.e. because it does not assume that an agent will be ideally rational at any time beyond the current moment of choice. (I also argued in Chapter 1 that the sophisticated approach makes sense of how the past affects choice at any given time.) While I think the decision model should permit any kind of predictions about future choice behaviour, this does not mean that an agent is entitled to be indifferent about how their beliefs/preferences will change. The upshot of my investigations in these

[79] Skyrms (1987) and Armendt (1993) refer to conditionalisation as a *plan* for updating beliefs and preferences. They argue, rightly I think, that a rational agent's beliefs/preferences need not actually change in accord with conditionalisation; it is sufficient that the agent plan such a change (in the right evidential circumstances).

last couple of chapters is that there are compelling constraints on how a rational agent may *plan* to update their beliefs or preferences, even if they have good reason to *predict* that their plans will go astray. For the most part, an agent should plan to update in accord with their conditional beliefs/preferences. This rule for updating is arguably proved by the diachronic DBA, at least as far as the belief side of the story is concerned. We might otherwise argue that updating via the rule of conditionalisation is just a matter of common sense, or in other words a matter of correctly interpreting conditional beliefs and preferences.

In this final chapter of Part I, I have concentrated on the desire side of the story. In particular, I claimed that there is an important asymmetry between belief and desire, when it comes to planning a change in attitude. We might look to the diachronic DBA: If an agent was able to consciously update their beliefs contrary to conditionalisation, then there is a sense in which any such updating plan will result in inferior outcomes in carefully contrived decision situations. But the same kind of story does not look so compelling when the agent's desire function over outcomes is what is at issue. In this case we must ask what makes particular outcomes inferior to others, i.e. by whose standards are we judging the results of the decision process? As a matter of fact, I do not think an agent can plan a conscious change in their basic tastes—the kind of change that is contrary to conditionalisation. But, as discussed, the agent might be pleased about an accidental change in taste if their "higher-order" preferences are so aligned. The epistemic story is quite different in this respect. It is permissible for a rational agent to pursue a strategy that is expected to involve some unorthodox belief change, but any such changes are regrettable (from the current point of view) and detract from the worth of the strategy. So we can conclude that, for the case of preference or desire, there are at least two different ways in which a rational agent can "plan" a change in outlook, but where beliefs are concerned, only update-via-conditionalisation is worth aspiring to.

II RISK-SENSITIVITY

4 ALLAIS'S PROBLEM AND THE INDEPENDENCE AXIOM

4.1 Introducing the concept of "risk"

This chapter serves as an opening to the second part of this thesis. My general preoccupation in these chapters (4–6) is whether the subjective expected utility (SEU) theory treatment of "risk" is the only rational approach. In other words, is there scope for an agent to evaluate acts according to a different calculus, or does rationality require that one should always seek to maximise expected utility? This basic question has lead to the development of a number of alternative decision models. (There have been even more doubts raised about the descriptive adequacy of SEU theory, but recall that I am concerned only with the normative case here.) Importantly, the various alternative decision models involve different sorts of adjustments to SEU theory, depending on how the concept of "risk" is interpreted, and how it is thought to affect choice. I cannot here give a comprehensive survey of the ways that "risk" has been defined, and all the associated modifications of SEU theory. I focus, rather, on a particular class of risk measures—those that pertain to an individual act and its distribution of outcomes (or, more precisely, its distribution of outcome utilities). This class of risk measures might include the spread of outcome utilities for an act, or the utility of the worst-case outcome. My interest is whether it is rational to take account of this kind of risk in decision-making.

A simple example might help to make the above comments about "risk" clearer. Consider the decision problem in Figure 4-1 below.

Figure 4-1

	Coin lands heads	Coin lands tails
Bet 1	Gain $50	Gain $50
Bet 2	Gain nothing	Gain $100

If we assume that the utility-money relationship is linear (even though this assumption generally doesn't hold), the two bets in Figure 4-1 have the same expected utility. In such cases, SEU theory demands that a rational agent be indifferent between them. But the two bets involve varying degrees of risk, that is, if we define "risk" in terms of the distribution of outcomes for an act. In the first case, the agent is assured of $50 no matter how the world turns out. In the second case the agent will win $100 if the coin lands tails, but the coin may land heads, in which case the agent will come away with nothing. We might say that the second bet is more "risky", either because its worst-case outcome has less utility than that of the first bet's worst-case outcome, or because its outcomes have a greater spread in utility (or greater variance). In any case, the general class of risk measures I am interested in are those that concern a single act and its distribution of outcome utilities. By way of contrast, it is worth noting that the "Regret Theories" of Bell (1982) and Loomes and Sugden (1982) define "risk" in terms of the relationship between outcomes of different available acts that fall under the same state. According to this alternative class of risk measures, the two bets in Figure 4-1 are similarly risky, because when measured against each other, they each involve the same gains and losses, given the two ways that the coin might land. (If the coin lands heads, then Bet 1 is up $50 compared to Bet 2, but if the coin lands tails, Bet 2 is up $50 compared to Bet 1.)

We can associate the idea of "regret" with the concept of "risk". Indeed, the two compliment each other because both have to do with uncertain outcomes, and hinge on the relationship between a realised outcome, and the alternative outcomes one might otherwise have feared or hoped for. Above, I contrasted the class of risk measures that will be my interest in these next few chapters with an alternative

concept of risk that involves comparisons between acts. We might make the same distinction using the language of regret. In the contrast case, the focus is on how the agent rates the act that they actually chose against the other acts that they might have chosen. For example, if the coin in the Figure 4-1 decision problem ends up landing heads, there is regret associated with the choice of Bet 2, because if the agent had chosen Bet 1, they would have been $50 better off. As stated, the sort of regret attitude that I am concerned with depends rather on the relationship between the outcomes of an individual act. The agent may regret receiving a particular outcome, not because they might have performed a different act, but because they are considering how the world might otherwise have turned out (in which case the same act would have led to a different outcome). There may be yet other ways to conceive of "risk/regret", but I have surely described two of the major classes of risk/regret measures. From now on, however, whenever I refer to "risk" or "regret", I am talking about a measure of the distribution of outcomes for a single act.

4.2 Allais's problem and the independence axiom

There are two well-known problems in the decision theory literature that press the issue as to whether it is rational for risk/regret considerations to affect an agent's preferences/choices. I am referring to the Allais and Ellsberg problems. In this chapter I will focus on Allais's (1953) problem, and in particular, the relationship between "risk-sensitivity" and the "independence" axiom of SEU theory. The Ellsberg problem raises some further questions about the relative precision of our beliefs; I will take this up in Chapter 5.

I turn then to the details of Allais's problem, and, in particular, to Savage's presentation of the decision scenario (Resnik, 1987, p. 105): we have two choice situations, *A* and *B*, and in each choice scenario, there is a choice between two lotteries (*a* or *b* in situation *A*, and *c* or *d* in situation *B*). For each lottery there are 100 tickets; one ticket is chosen at random and different ticket numbers yield different

monetary amounts, as per Figure 4-2 below. The punter is asked to nominate their preferred lottery in each of the two choice situations *A* and *B*.

Figure 4-2

	Ticket Number		
	1	2-11	12-100
A:			
a	1 million	1 million	1 million
b	0	5 million	1 million
B:			
c	0	5 million	0
d	1 million	1 million	0

A number of empirical studies have shown that many people make inconsistent choices, by the lights of SEU theory, when faced with Allais's problem—they opt for the sure 1 million in problem *A*, but then choose the 5 million gamble in problem *B*. (I will call this combination of choices the "Allais-choices", and the people who select these options the "Allais-choosers".)[80] The Allais-choices are inconsistent according to SEU theory because the choice of the sure 1 million in problem *A* indicates that the agent holds the following to be true:

[80] Allais's initial empirical finding was that the majority make the "Allais-choices" (MacCrimmon and Larsson 1979).

$$U(1\text{ million}) > 0.10 \times U(5\text{ million}) + 0.89 \times U(1\text{ million})$$

$$0.11 \times U(1\text{ million}) > 0.10 \times U(5\text{ million})$$

The choice of option *c* over option *d* in problem *B*, on the other hand, indicates that the agent holds this expression to be true:

$$0.10 \times U(5\text{ million}) > 0.11 \times U(1\text{ million})$$

Clearly these statements are contradictory.[81] But we might wonder why the Allais-choices seem so reasonable to many. Some argue that the Allais-choosers are sensitive to the varying riskiness of the options in problems *A* and *B*, and that SEU theory does not adequately accommodate this kind of risk-sensitivity.

Before continuing, I want to point out that the above analysis highlights a particularly valuable aspect of Allais's decision problem—the problem isolates, as far as possible, the influence that risk/regret considerations might have on choice. In fact, both Allais's and Ellsberg's respective problems are formulated in such a way as to rule out the relevance of the diminishing marginal utility of goods/money phenomenon. We can see from the expressions above that the Allais-choices are inconsistent, whatever utility the agent attributes to the stated outcomes. The same cannot be said of the simple example that I gave in Figure 4-1 above. In that case, we could try to argue that SEU theory is inadequate, and make some appeal to the varying riskiness of the bets in order to defend a preference for Bet 1 over Bet 2. But a SEU theory-defender could simply reply that there is a much more obvious explanation for the preference for Bet 1—the agent's utility for $100 might be less than double their utility for $50, in which case the two bets do not have equivalent expected utility after all. While a concave money-utility curve like this is sometimes referred to as a "risk-averse" function, I think it captures an attitude towards the good in question (in this case money) rather than an attitude towards risk. In any case, this is not the type of risk-sensitivity that I am interested in here; my question is whether it is legitimate for an agent's choices to be sensitive to the spread of outcome utilities associated with an

[81] I am assuming that the agent's subjective beliefs correspond to the (arguably) objective probabilities for each of the three outcomes. But this assumption is not necessary. The agent's choices can be shown to be inconsistent regardless of what probabilities they assign to the three states.

act. Allais's problem, in particular, is designed to press this very question.

Savage presents Allais's problem in the manner depicted in Figure 4-2 so as to highlight the relevance of the independence axiom. This axiom is the backbone of SEU theory's approach towards risk. But I will say a bit more about the connection between the handling of risk and the independence axiom shortly. Let me first give a formal expression of the "independence" axiom. Joyce (1999, p. 85) presents Savage's version of it as follows:

> Suppose that (acts) A and A^* produce the same outcomes in the event that E is false, so that $A_{\sim E} = A^*_{\sim E}$. Then, for any act $B \in \mathbf{A}$ (where $\mathbf{A}$ is the set of all acts, including constant acts), one must have
>
> $A > A^*$ if and only if $A_E \,\&\, B_{\sim E} > A^*_E \,\&\, B_{\sim E}$
>
> $A \geq A^*$ if and only if $A_E \,\&\, B_{\sim E} \geq A^*_E \,\&\, B_{\sim E}$

In other words, independence holds that "a rational agent's preference between A and A^* should not depend on what happens in circumstances where the two yield identical outcomes." (Joyce, 1999, p. 86)[82]

In exposing agents' risk-sensitivity, Allais's decision problem effectively challenges the independence axiom of SEU theory. If we model the problem as per Figure 4-2 above, the lottery pairs in situations A and B are identical in the last column (a and b both have winnings of 1 million for tickets 12–100, and c and d both have winnings of 0 for tickets 12–100). So the choice between a and b in situation A should depend solely on whether the agent prefers 1 million for all 11 tickets, or whether they want to take a gamble on one ticket yielding nothing and the other 10 yielding 5 million. But this is exactly what the choice in situation B depends on. So if you choose a in situation A then for consistency you should choose d in situation B. And if you choose b in A then you should choose c in B. (Or else you can be indifferent between both

[82] I gave this same presentation of the independence axiom in Chapter 1.

lotteries in each choice scenario.)

The reason Allais's problem is considered paradoxical is that while many people think it is reasonable for choice to be affected by the spread of outcome utilities, as well as the expected utility, of an act, many also regard independence as a compelling constraint on rational choice.[83] Now there is clearly much intuitive appeal to the independence axiom when it is stated in general terms as the rule that "choice between acts should not depend on circumstances in which the acts yield identical outcomes". But it must be noted that it is precisely this axiom that precludes any kind of risk measure from playing a formal role in the decision calculus. The risk measures that I have been referring to rest on "global" properties of acts. If such a measure was incorporated in the decision calculus then acts could not be evaluated on a state-by-state, or an outcome-by-outcome, basis. Individual outcome utilities would be important, but so too would be the relationship between outcomes. This entails a violation of independence because it allows preferences between acts to shift, depending on how the outcomes that are common to the acts in question affect their respective spreads of outcome utilities.

4.3 The significance of empirical results about choice behaviour

It is important to be clear about just what we can learn from experiments like those involving Allais's decision problem. Since Allais first proposed his famous problem, there have been various empirical tests to ascertain what factors affect the choices people make in Allais-type decision situations. MacCrimmon and Larsson (1979, pp. 350–59) reference the historical results of Allais and Hagen, in addition to outlining some experiments of their own. As might be expected, the results are sensitive to the particular group of people who are sampled, how the problem is presented to them, and the parameter values that are used. Both Allais and Hagen found that the majority

[83] MacCrimmon and Larsson (1979) cite empirical findings that support this combination of sentiments.

of their subjects made choices that seem to violate the independence axiom (and in particular, the majority of choices showed an "aversion towards risk"). But MacCrimmon and Larsson later found that this result is highly sensitive to the probabilities and utilities involved. When monetary amounts are extreme (in the order of millions of dollars) and the probabilities are very disparate (such that the agent must compare a 100% chance with a 5% chance), a number of tests confirm that there is a tendency to make choices that appear to violate independence. But when the parameters have values that the average person is better able to comprehend, such that they can make more meaningful comparisons between acts, we do not see the same trend.

These empirical results are of course very valuable, and are particularly pertinent to the descriptive study of choice. But it is important to question what we should ultimately take away from findings about how people actually choose, when considering the normative question of how people should choose. Kahneman and Tversky (e.g. 1982), for instance, have drawn attention to a number of mistakes that the majority of people tend to make when reasoning with probabilities. The lesson here is that we should be very cautious about basing the properties of a normative decision model on the choice behaviour of the majority. But this is not to say that trends in actual choice behaviour cannot offer any insights into the normative study of choice. For one thing, convincing empirical results serve to focus attention on aspects of the normative decision model that have implications we might want to challenge. This is precisely what has happened in the case of Allais's problem and the independence axiom of SEU theory. Secondly, while broad statistical results about choice behaviour in Allais's betting scenario are ultimately neither here nor there when it comes to the normative decision model, principles of rational choice must at least appeal to reasonable-seeming people, upon reflection. And a number of reasonable people (in particular, some decision theorists) have declared that they would choose the combination of bets in Allais's decision scenario that appears to violate independence, even after reflecting on this fact. In my opinion this means that we cannot take the logical necessity of the independence axiom for granted.

4.4 Two readings of Allais's "paradox"

Even if we restrict our attention to the reasonable and reflective Allais-choosers, it is hard to know what conclusions to draw from the apparent disparity between their choice behaviour and what the independence axiom prescribes. Indeed, this is a highly contested issue, and it is the focus of this chapter from here on in. We could categorise the different positions in a number of ways, but for present purposes I am interested in whether those who opt for the sure 1 million in problem *A*, but who would take the 5 million gamble in problem *B*, do indeed violate the independence axiom. Accordingly, the first response is simply to acknowledge that the Allais-choosers violate independence. (Then there are divisions in this camp as to what such a violation means.) The second response is that the decision problem must be incorrectly specified, because any well-specified problem would not make a violation of independence seem attractive to reasonable agents. I will discuss these two main responses to Allais's problem in turn.

As briefly stated above, our first reading is to affirm that there really is a "paradox" here. This is to accept that a significant number of reasonable people (after reflection) indeed violate the independence axiom when responding to Allais's problem. Of course, even if we accept the violation of independence, there are different things that can be said about this. Savage (1954, p. 101 ff.) for instance maintains that Allais's problem simply exposes a common flaw in people's reasoning. The fact that many people stick to their faulty choices after reflection only goes to show how seductive the inconsistent choices are in this particular kind of scenario. Unerringly rational decision makers, Savage would maintain, do not violate independence in this way; they would choose *a* and *d* or *b* and *c* (as per Figure 4-2), or else they would be indifferent between the two lotteries in each situation. The paradox is dissolved in this way: the seemingly reasonable Allais-choosers are not so reasonable after all. But as stated above, I think this response is a bit quick. In my opinion, Allais's problem demands a more substantial defence of the independence axiom.

There are others who agree with Savage that many people violate independence in the Allais scenario, but who take the moral of the story to be quite the opposite from what Savage takes it to be. They claim that Allais shows a fault in the independence axiom rather than a fault in people's reasoning. In other words, Allais's problem is a genuine challenge to SEU theory. According to this line of thought it is then a question of finding a suitable replacement theory that involves some relaxation of the independence axiom. Machina (1989, p. 1631) lists a number of potential alternative decision models, each with its own risk/regret rationale (that will allow for an explanation of the Allais-choices). While I will not pursue the details of these models, in the next Section I do discuss how there are more and less defensible ways to violate independence. For now, all we need to know is that there are a number of alternatives to SEU theory that involve some relaxation of independence.

Then there is the second reading of Allais's problem. A number of decision theorists (e.g. Jeffrey 1982, Weirich 1986 and Broome 1991) seek to reconcile the paradox with SEU theory.[84] Broome (1991, p. 107) claims, not unlike Savage, that violation of the independence axiom is just outright irrational. But unlike Savage, and in common with those just referred to, Broome seeks ways of redescribing Allais's problem (specifically the act outcomes) so that the common response to the problem turns out to be consistent with SEU theory after all. Broome's conclusion is much the same as others in this camp—as far as the typical agent is concerned, the proper way to

[84] I note that there is a third response to Allais's paradox that I am not emphasizing here because I am interested in the tension between the other two responses. The third path is to explain the common responses to Allais's problem by relaxing some other axiom of SEU theory (arguably an axiom more dispensable than independence). For instance, Levi (1997) argues that his decision theory, which relaxes ordering (and thus allows for incommensurability of acts), can explain both the Allais and Ellsberg paradoxes. I will give the details of Levi's theory in Chapter 5, and will apply it Ellsberg's problem. When it comes to Allais's problem, in brief, Levi thinks it is plausible that the agent's utility function is indeterminate in such a way that in both choice situations the two acts are incommensurable with respect to expected utility comparisons. Where there is no preference between options, Levi claims that we can appeal to secondary security considerations—a person who shares his intuition to maximise the worst-case scenario will choose options *a* and *c*. (I do not agree with Levi's treatment of secondary security considerations, as will become clear in Chapter 5.) Others also claim that Allais's paradox can be explained by relaxing ordering. Schervish *et al.* (1990) make the claim in passing. Bell (1982) and Loomes and Sugden (1982) claim that their respective versions of "regret theory" (which also relax ordering) can account for the "Allais-choices". (I don't see how relaxing ordering alone will give the risk-sensitive choices, but again, this general issue will be taken up in Chapter 5.)

describe the stakes involved in Allais's problem is to include in the description/evaluation of act outcomes the agent's attitudes towards risk/regret. For instance, we might factor into the relevant outcomes the typical agent's comparative happiness when it comes to outcomes that are more secure or closer to a sure deal. Accordingly, the outcomes for act *a* (refer to Figure 4-2) might be specified as 1 million + ∂, rather than 1 million apiece, where the term ∂ represents the extra satisfaction that the agent experiences given that they know the 1 million monetary gain is a certainty. Once this move is made, the symmetry of the choice situations *A* and *B* is broken, and we no longer have a case where the independence axiom applies. The rest of this chapter is mainly concerned with whether this move—including risk/regret sentiments in outcomes—is legitimate, or whether it is at odds with SEU theory. In the latter case, if we persist in claiming that risk/regret considerations legitimately affect choice, then we are effectively demanding a relaxation of the independence axiom of SEU theory.

4.5 Why not relax independence?

Before I discuss the latter "redescribing outcomes" response to Allais's paradox, let me elaborate on the motivations for this line of response. Broome and others in this camp think that the "Allais-choices" are reasonable, and yet want to explain them without sacrificing independence. So why is it so important to retain the independence axiom? Many claim that the reasonableness of independence is self-evident. Indeed the axiom is intended to be an intuitive requirement of rationality. As mentioned, I do not think this is a good enough defence of the axiom. Given that Allais's problem and the broader phenomenon of risk-sensitivity presents a challenge to independence, we might say that it is begging the question to respond with a restatement of the intuitive reasonableness of the axiom. In a moment, I will consider possible avenues for a more substantial defence of independence. But it is worth bearing in mind throughout this discussion that even if there is no decisive argument one way or the other, independence remains a compelling constraint on rational choice. So if Broome and co. can accommodate risk/regret attitudes within the SEU model, then surely this is

the path of least resistance for anyone who wants to take seriously the Allais-choices, and risk/regret sensitivity in general.

A more substantial defence of independence will require some further appeal to the consequences of violating the axiom. It is generally agreed that not all violations of independence are on the same footing, and some violations lead to obviously bad outcomes. Consider the original "Prospect Theory" proposed by Kahneman and Tversky (1979).[85] This decision calculus has the form:

$$\mathrm{U}(X) = \sum_{i=1}^{m} \mathrm{U}(O_i) \times \pi(\Pr(S_i))$$

where U is the agent's utility function, and option X has possible outcomes $O_1 \ldots O_m$ corresponding to states $S_1 \ldots S_m$, and

$\pi(.)$ is the agent's subjective risk function, subject to $\pi(0) = 0$ and $\pi(1) = 1$.

Machina (1989, p. 1634) notes that, if the function $\pi(.)$ is not linear (because in that case the agent's preferences would just conform to the SEU axioms), then there will exist probabilities over outcomes $\Pr(S_1) \ldots \Pr(S_m)$ summing to unity, such that

$$\sum_{i=1}^{m} \pi(\Pr(S_i)) \neq \pi(1)$$

Take the case where

$$\sum_{i=1}^{m} \pi(\Pr(S_i)) > \pi(1).$$

(The reverse case yields a similar result.) Then there will exist outcomes

$O_1 < O_2 < \ldots < O_m < O_\#$ such that

$$\sum_{i=1}^{m} \mathrm{U}(O_i) \times \pi(\Pr(S_i)) > \mathrm{U}(O_\#) \times \pi(1)$$

The agent is thus directed to choose the left-hand option, which means that they will miss out on a sure amount d, where $d > \mathrm{U}(O_\#) - \mathrm{U}(O_m)$! Another way of putting the

[85] The comments in this paragraph closely follow Machina's (1989, p. 1634) discussion of the original Prospect Theory.

above point is that Prospect Theory, as outlined above, violates independence in such a way that it also violates "first-order stochastic dominance", or what is simply referred to as "dominance"—there will be cases where, no matter how the world turns out, one act yields a preferred outcome, and yet the agent opts for the alternative act. We could say that the agent's decision calculus robs them of a sure amount, and this can only be considered a bad thing.[86]

But independence is a stronger constraint on choice than simple dominance, and so it is possible to relax independence while still respecting dominance. The major difference between the two is that dominance deals in simple or "basic" outcomes, while independence deals in "compound" outcomes. (The difference between the two is that compound outcomes are described in terms of a probability distribution over basic outcomes. Basic outcomes cannot be described in terms of a probability distribution over more basic outcomes.) Dominance stipulates that if you prefer the basic outcome O_x to another basic outcome O_y, then when you are faced with two acts that are exactly the same, except that the first yields outcome O_x in one particular state (or with some probability p), while the other yields outcome O_y for the same state, then you should prefer the first act. The independence axiom looks very similar to this, except that it applies to compound outcomes as well as basic outcomes. As per Joyce's formulation of the axiom given above, independence states that, for any two acts A and A^*, if you prefer A to A^*, (note that these are acts rather than basic outcomes), then you should prefer the compound act that for some partition of the state space yields A, to an otherwise identical compound act that yields A^* for the given set of states.

Whatever we might think about the rationality of risk/regret considerations affecting choice, it is surely unwise to accept a decision calculus that violates simple dominance, as the above analysis of a simplistic version of Prospect Theory shows. But more sophisticated decision theories have been developed that explicitly

[86] Kahneman and Tversky (1979) were not unaware of this problem with Prospect Theory. They thus recommended a two-stage decision process, where the first stage would involve removing first-order stochastically dominated options from the option set.

incorporate risk/regret considerations, including "Cumulative Prospect Theory".[87] These theories relax independence while retaining dominance, and thus are not so easy to dismiss. It might be thought that we could mount a Dutch Book argument against these more sophisticated independence-violating theories. But this will not be a fruitful avenue for criticism, because the Dutch Book argument assumes something akin to independence, and so it is likely that anyone who challenges independence is going to challenge the Dutch Book argument as well, and for similar reasons. In Chapter 2, I discussed this relationship between independence and the "value additivity" assumption of the Dutch Book argument, i.e. the assumption that the sum of fair bets is itself a fair bet. (My discussion draws on the work of Armendt (1993) and Schick (1986).) In short, there is no obvious problem with relaxing the independence axiom while retaining dominance.

It could be said, then, that the case for independence being an inviolable constraint on choice needs some further support. There are indeed some arguments to this effect that make reference to the sequential-choice framework. An analysis of these arguments will require some considerable space, and indeed, this will be my concern in Chapter 6—I consider what, if anything, the sequential-choice framework can reveal about the qualities of the SEU axioms of preference, in particular the independence and ordering axioms. But even if the arguments in support of independence are not entirely conclusive, I want to emphasise that if all other things are equal, it is surely better to retain the axiom. After all, there is a great deal of intuitive appeal to the independence axiom. And the position of Broome and Weirich is precisely that all other things are equal—given that SEU theory can accommodate risk-sensitivity, why move to a theory that relaxes the independence axiom? Recall that this is precisely the issue I want to pursue here: can SEU theory be shown to accommodate the kind of risk/regret attitudes that seem to motivate the search for a

[87] Machina (1989, p. 1631) notes that most of the independence-violating decision theories he lists (the latter five, anyhow) respect simple dominance, provided their component functions respect some reasonable monotonicity conditions. Cumulative Prospect Theory is just one example (and in fact Machina refers to this model as "Anticipated Utility"). I will not detail the model here, but note that it involves the transformation of cumulative rather than individual probabilities. A version of the theory is outlined by Tversky and Kahneman (1992), who credit the earlier work of Quiggin (1982), Schmeidler (1989), Yaari (1987) and Weymark (1981).

less stringent normative theory of choice? If so, there is no need to pursue a relaxation of the independence axiom. If, on the other hand, it is shown that SEU theory cannot accommodate the full spectrum of risk-sensitivity, then we are faced with a choice—we can rule that some kinds of risk-sensitivity are simply irrational, or we can pursue a relaxation of independence, despite whatever arguments there may be against this move. So having described what is at stake, let me return to the question of whether SEU theory can indeed account for risk-sensitivity.

4.6 Savage's theory and the content of outcomes

We can continue to use Allais's problem as the focus for investigations of risk-sensitivity. As mentioned, both Broome (1991) and Weirich (1986) seek to explain Allais's problem while not giving up the independence axiom. Their common strategy is to redescribe the relevant outcomes in Allais's problem. Interestingly, Broome and Weirich both argue that Savage's theory, to its detriment, doesn't allow such a move. I want to consider the general issue here: Do specific SEU theories (taken in all their detail) constrain the content of outcomes, and in particular, do they prohibit risk/regret properties in outcomes? In such case, we would have a very clear-cut answer to the question of whether SEU theory (or at least the versions of SEU theory in question) can accommodate the Allais-responses. Savage's (1954) theory is a good place to start on this issue, seeing as Broome and Weirich have already argued that it does not permit act outcomes to involve sentiments towards risk.

The argument is not that the independence axiom itself rules out risk sentiments being incorporated in outcomes or prizes. That would be begging the question. It is another assumption in Savage's theory that has been brought under scrutiny; both Broome (1991, pp. 115–117) and Weirich (1986, p. 424) point the finger at what Broome calls the "rectangular field assumption".[88] The assumption is not an intuitive requirement of

[88] Broome uses the term "rectangular field assumption" because the assumption in question concerns a product set, and "a product set occupies a rectangle or a series of rectangles in

rationality. It is what Joyce refers to as a "structure" axiom in Savage's theory, or what Hájek (2006 manuscript) calls "an idealisation of our theory of rationality" as opposed to "an ideal of rationality itself." It is described as follows:

> …there is a set of possible states of the world and a set of possible consequences, and any function from possible states to possible consequences is an option. It follows that a possible consequence can be produced by a variety of options—by options that yield the consequence in every state and by options that yield it only in a single state.[89]

Broome (1991, p. 115) states what he thinks are the consequences of this assumption. He considers a lottery, and how we might describe the outcome of losing. If the agent is sensitive to risk or regret, it might seem best to describe the outcome as "receive no money and feel disappointment at not winning". Broome (1991, p. 115) goes on to say why such an outcome is at odds with Savage's SEU theory:

> The rectangular field assumption says your preference ordering includes all arbitrary prospects. Amongst them is the prospect that leads to this particular outcome for sure. This prospect determines, whatever lottery ticket you draw, that you get no money and also feel disappointment. But this feeling of disappointment is supposed to be one you get as a result of bad luck in the draw. It is hard to see how you could feel it if every ticket in the lottery would lead to the same boring result. So this prospect seems causally impossible, and that may make if doubtful that it will have a place in your preferences.

To summarise, the claim that Broome makes here (and that Weirich also effectively makes in his further discussion of the issue) is that we cannot just randomly attribute outcomes involving risk sentiments to states (as per the "rectangular field

---

vector space" (1991, p. 80).

[89] This is a direct quote from Weirich (1986, p. 424). Weirich attributes the first sentence to Savage (1954, end papers and p. 14 f).

assumption") because such risk properties intimately depend on the precise combination of states/outcomes involved in an option.

But I think Weirich and Broome over-interpret the "rectangular field assumption", and single out Savage's theory unnecessarily, at least with respect to risk/regret. I agree that the assumption strongly suggests that the description of outcomes should preclude risk sentiments, but I do not think that we should interpret it so literally. As discussed, the "rectangular field assumption" is an idealisation; it describes the preference space for an ideal agent in such a way that Savage's (1954) representation theorem gives us a continuous utility function (unique up to positive linear transformation) and a corresponding unique probabilistic belief function for an agent. We cannot assume that an agent with such extraordinary discerning powers actually exists. It is impossible for us ordinary mortals to entertain a complete preference ordering over the infinitely rich option space that Savage's theory requires. Further, many of the options in the option space described by the "rectangular field condition" will not be physically possible. And not just due to any deficiencies us non-ideal agents might have, but because the actual world constrains the set of actions that any agent, ideal or otherwise, is able to carry out. Just because we can conceive of an abstract map from states to some combination of outcomes doesn't mean that the act in question is, will be, or ever was, a viable possibility in the actual world. To top it off, many state-outcome combinations will not merely be impossible for an agent to achieve, but will be outright contradictory. Schervish et al. (1990, p. 842) ask how it is possible, for instance, for the outcome "walk to work in the rain" to occur in a state such as "bright sunny morning". Surely if the weather is fine at my location I cannot be walking in the rain. So there is more than one way in which the acts in Savage's assumed outcome space are fictitious, whether or not we want to further introduce risk sentiments. Thus it is not clear why Broome, in his statement above, invokes causal impossibility as an unassailable obstacle for incorporating risk sentiments, in particular, into outcomes.

Given the ideal nature of Savage's "rectangular field condition", I think any attempt to draw from it concrete conclusions about the contents of act outcomes is

questionable. I do not think Savage's theory, in particular, rules out risk/regret sentiments from featuring in the description of act outcomes. Moreover, there is even less reason to think that Jeffrey's theory constrains outcomes in this way. There is no "rectangular field assumption" in Jeffrey's theory, and it is made explicit that outcomes include properties of the act—for Jeffrey, an outcome is the conjunction of an act and a state. In fact, Broome and Weirich agree with me on this point; they think it is a virtue of Jeffrey's theory that it can incorporate risk/regret sentiments in outcomes. But not only do I think that Jeffrey's theory shouldn't be singled out in this respect, in my opinion, we should be very cautious about how expansive (or "comprehensive") we want to make the description of outcomes.

4.7 A vacuous decision theory?

I think there are significant reasons to be concerned about an SEU model that allows risk/regret sentiments in outcomes, even if what is at stake is an independence-compatible explanation of the Allais-choices. The move threatens to trivialise SEU theory (whichever representation theorem is assumed as its justification). Simply altering the representation of preferences or redescribing outcomes (such that winning $1 million in the comfort of it being a certainty and winning $1 million in a gamble are different outcomes) to force compliance with the SEU model seems ad hoc or overly permissive. If we allow such moves then it is questionable whether we can produce counter-examples against SEU theory, which would make it essentially vacuous.

Consider Broome's analogy between what it means for an agent to comply with the independence axiom and what it means for the agent to comply with the (perhaps less controversial) axiom of transitivity.[90] The transitivity axiom would be vacuous if

[90] Broome (1991) eventually concludes that risk/regret sentiments *are* a legitimate aspect of outcomes, but he discusses at length the problem of being too permissive with respect to defining act outcomes.

every time it seemed as if an agent were violating it, e.g. their preferences over prospects *A*, *B* and *C* appeared to have the following cyclical form—

1. $A \succ B$ and $B \succ C$ but also $C \succ A$

(recall that "$\succ$" denotes the strict preference relation)

—the preferences could simply be redescribed (in the following way for instance) such that adherence to transitivity is really an assumption rather than a result to be tested:

2. $A \succ B1$ and $B2 \succ C$ and $C \succ A$

(where *B1* and *B2* are two other basic outcomes)

(Nothing is said in the second set of preferences above about the relationship between *B2* and *A*, so there is no violation of transitivity.) The same kind of argument applies to the independence axiom, which is what is at stake in Allais's problem, and in discussions about risk-sensitivity in general. The axiom is essentially vacuous if it acts as an assumption in our model of an agent's preferences/choice behaviour rather than as a result to be tested.

Let me note that some might be quite comfortable with decision theory being trivial in the sense alluded to above. Axioms such as transitivity and independence might be considered logical relations that an agent cannot but satisfy. An agent's preferences, then, are simply assumed to be rational, and the task is to determine the agent's probability (credence) and utility (value/desire) functions such that he/she can be represented as an expected utility maximiser. I do not agree with this account, however, because I think SEU theory is intended to provide a "thicker" notion of rationality than this. In other words, I think SEU theory is intended to provide us with some substantial guidance when it comes to our preferences amongst acts/outcomes. This means that it should be possible for an agent to *fail to comply* with the theory. And there should be a way to distinguish rational from irrational preferences, at least in principal. The debate about risk-sensitivity indicates that we can only make such distinctions if there are restrictions on what can count as an outcome.

Of course, we do not want to constrain the content of outcomes in a way that privileges particular theories of value. SEU theory is intended to be value-neutral. It is not supposed to be the final word on practical choice. The theory is silent on substantive matters of value—what it demands is just that one's preferences are consistent, and this may well be satisfied by an agent who prefers genocide to a walk in the hills. But we could always tighten up our model of choice by supplementing SEU theory with whatever ethical constraints seem most reasonable (Colyvan *et al*. to appear). When determining whether an agent is merely rational or coherent, we do not want to take controversial stands on what aspects of the world should be valued, and in what way. We are looking for more minimalist constraints on the content of outcomes.

My particular interest here, of course, is whether risk/regret attitudes can legitimately feature in outcomes. But it is useful to first consider general proposals for determining how outcomes should be described, and whether such proposals shed light on the risk/regret issue. Broome (1991), for instance, claims that we seek a principle for determining whether it is reasonable to have a preference between two outcomes, given that they differ in a specified way. Perhaps the differences between some outcomes are so inconsequential that it is irrational to be anything but indifferent between them. For example, it would seem irrational for me to prefer the state of affairs where I am reading the paper at position *A* to the state of affairs where I am reading the paper at position *B*, where *B* is one millimetre to the left of *A*, and all other things are equal. But even here we are taking a stand on values; in the case described, why is such a small difference in spatial location unimportant?

Instead of trying to make objective claims about what properties discriminate one outcome from another, some have suggested that what matters, rather, is whether an agent's preferences are in some sense consistent. This is the approach that I favour. Pettit (1991) provides a good way of thinking about this issue. He argues that we owe an explanation for why one outcome (or prospect) is preferred to another, and this explanation hinges on what properties of the outcomes in question affect our evaluations. If a particular kind of property matters to an agent in one situation, then

discrimination between outcomes on this basis is legitimate to the extent that the property always affects the agent's evaluations. So an agent must demonstrate consistency with respect to how he/she distinguishes and evaluates outcomes in terms of the properties that the outcomes exhibit. Of course, the story about properties will be complicated, as we will still need to make judgments about what can be suitably called a "property"—it must be a sufficiently salient or natural grouping.[91] There will be a further problem associated with interactions between properties and how such interactions influence evaluations. I will assume that such problems are surmountable, and thus take "property-consistency" as outlined to be the right kind of criterion for determining the proper content of outcomes for a particular agent.

The above criterion facilitates a thicker sense of rationality, meaning that it will be possible to determine instances when an agent violates rational principles such as transitivity. So what does the "property-consistency" constraint say about risk/regret attitudes, and determining what counts as a violation of independence? For starters, it is still unclear as to whether there are any kinds of risk/regret properties that can be associated with individual outcomes, or whether any such properties will be "global" properties of an act. If we are not talking about properties of individual outcomes, then it doesn't look like any kind of "property-consistency" criterion can be sensibly applied, and there is no hope for reconciling risk-sensitive choice behaviour with SEU theory. There is reason to think that risk/regret attitudes can be properties of individual outcomes, however. An obvious move is to claim that risk/regret attitudes are simply a certain type of emotional response, either anxiety or excitement about the perceived riskiness of an act, or regret (whether positive or negative) that is experienced when some outcome rather than another actually eventuates. It is likely that expected emotional responses consistently matter to an agent. Most plausibly, the agent will always value excitement and elation favourably, and anxiety and regret unfavourably. In such case, it is quite legitimate for the presence of such emotions to be cause for distinguishing between what would otherwise be identical outcomes.

---

[91] Some kind of "naturalness" condition is necessary, because otherwise properties could be any kind of gerrymandered grouping and would not serve to constrain the content of outcomes at all.

Jeffrey (1982) discusses an example problem where an apparent violation of independence can be explained away by taking into account obvious emotional responses that distinguish between otherwise identical outcomes.[92] Sen (1985) gives a similar example that goes roughly like this: Mr Smith may receive a letter, and its contents will be one of two things, either notification of a large cash win in Case 1, or else a court summons for some semi-serious traffic law infringement in Case 2. For both cases, if he does not receive the letter, then Mr Smith will either clean up the house (act *A*) or have some champagne and cheese (act *B*). What this example is supposed to illustrate is that it would be very reasonable for Mr Smith to prefer act *A* over *B* (cleaning the house over drinking champagne) in the event that he fails to receive a letter that might have contained a large cash prize, and at the same time reasonable for Mr Smith to prefer *B* over *A* in the event that he does not receive a letter that might have contained a court summons. On the face of it this seems like a violation of independence (because the alternative outcome of receiving a cash win/fine is identical for each option in both Cases 1 and 2), but most of us think that the choice situations are not symmetrical because drinking champagne having avoided a court summons is quite a different outcome from drinking champagne having failed to win a prize. We can well imagine in this situation that there will be strong emotions at play—disappointment at not having won the prize versus relief at having avoided the fine. To ignore such obvious emotions in the decision model just seems wrong.

We have here then a way to explain the Allais-choices. If an agent has some amount of anxiety when faced with uncertain options as compared to certain options, and anxiety is a property that consistently affects the agent's evaluation of an outcome, then further distinctions can be made amongst the outcomes in Allais's problem such that there is no violation of independence. (Indeed this is the kind of "redescribing outcomes" explanation of Allais's problem that is commonly put forward.) It is a

[92] Jeffrey in fact takes his example from Machina (1982), who does not incorporate risk/regret sentiments in outcomes and formulates the example to illustrate the deficiencies of the independence axiom.

reasonable account because it is very plausible that the qualitative difference between certain and uncertain gambles will be linked to a noticeably different emotional reaction or sentiment. Moreover, even if the parameters in Allais's problem are slightly modified, so that there is no qualitative certain/uncertain distinction, an appeal to emotional response can still be compelling. It is not inconceivable that an agent may be highly emotionally sensitive to the varying amount of riskiness associated with Allais-type options, whatever probabilities and monetary amounts are at play. In general, SEU theory can accommodate Allais-type choices to the extent that there are tangible risk-related emotions involved, and such a move does not make the theory vacuous.

4.8 Can we take risk/regret sensitivity further?

An explanation of the Allais-choices is all well and good, but Broome and Weirich seek a broader and more robust account of how choice may be affected by risk/regret considerations. They both hold that risk/regret sensitivity need not be linked to a tangible emotional reaction. And incidentally, while I have indicated that it is by no means impossible, I think the case for detectable emotional responses is not so strong in the Allais case as it is for the Mr Smith-letter case, particularly when the original probabilities are adjusted in such a way that there is not the qualitative certainty/uncertainty distinction. To the extent that Weirich and Broome try to marry more sophisticated risk/regret sensitivity with the SEU framework, I think their accounts run into problems.

Broome appeals to counterfactual properties of outcomes as a way to accommodate sophisticated risk/regret attitudes. In this way, risk/regret attitudes are effectively linked to "global properties" of an option, but importantly, these "global properties" are present in individual outcomes in the form of counterfactuals. In the Mr Smith case, for instance, there will be two different outcomes "drink champagne given that a money prize would otherwise have been received" and "drink champagne given that a

fine would otherwise have been received". We can distinguish these two outcomes without recourse to emotional response. (Appeal to emotions gives us the outcomes "drink champagne feeling disappointed" and "drink champagne feeling elated".)

My concern about the counterfactual tactic amounts to the old problem that it is too permissive with respect to the way outcomes are described. Recall the suggestion that outcomes may only be distinguished in terms of properties that consistently affect the agent's evaluations. The question is whether there can be any salient groupings of counterfactual statements that could serve as this kind of property. What group or property should "drink champagne given that a fine would otherwise have been received" belong to? We might be able to make some coarse-grained distinctions—for instance, we could group outcomes that involve the counterfactual statement that something much better or much worse would otherwise have eventuated. But if we were to treat outcomes differently depending on the precise combination of other possible outcomes associated with the same act, then we are talking about a large number of very specific outcome properties, and would thus be running the risk of having a vacuous decision theory. Any set of preferences could be defended as rational.

Weirich's (1986) account of how SEU theory can handle risk/regret could also be interpreted as resting on counterfactual properties of outcomes.[93] Weirich allows an outcome to be sensitive to the precise distribution of other possible outcomes associated with the act. Again this seems overly permissive, but Weirich proposes that the contribution of "global properties" to the value of outcomes be constrained by a specific algorithm or decision rule. He experiments with a rule whereby the value of an outcome is modified by a function of the variance of the act's overall outcome distribution. (This is supposedly consistent with SEU theory because once outcomes have been modified to include a variance factor, the best act is the one that maximises expected utility.) I think there are reasons to think that a variance factor is not a good

[93] Weirich himself does not appeal to counterfactual properties. But I think his account would be strengthened if he did present it in this way.

choice of rule for modifying outcomes.[94] But whichever rule is decided upon, I think it is suspect to categorise this response to risk/regret sensitivity as a mere "redescription of outcomes". Effectively an alternative decision rule is being used to decide upon the ordering of acts, but the details of the rule are hidden away in the evaluation of individual outcomes, so that on face value, consistency with SEU theory is maintained. And again, I do not think that specific distributions of outcomes are the right sort of things to count as properties of individual outcomes.

4.9 Risk/regret conclusions

So is redescribing outcomes to include global-property-inspired risk/regret sentiments a legitimate way to accommodate these sentiments within SEU theory, and given our problem of interest, a legitimate way to explain the Allais-choices? I have argued that a possible danger of this move is that it threatens to make SEU theory vacuous. Those who think that SEU theory merely elucidates inevitably logical relations among preferences may be unconcerned about this. I am not of this opinion, however. I think outcomes must be constrained in some way if SEU theory is to have any normative content. But I don't think we can find these constraints amongst the assumptions of Savage's, or any other, representation theorem. The idea that outcomes may be distinguished only with respect to properties that consistently affect the agent's evaluations of states of affairs is I think the right way to limit what counts as a new or different outcome. It is very plausible that an agent would consistently regard anxious sentiments as having disutility, and feelings of relief as having positive utility. In the Allais case, for instance, an extra property of relief or confidence might distinguish the outcomes associated with the certain prize in problem *A*. In this way, we will be able to explain a lot of apparent violations of independence. All that is required is a plausible story about the risk-related emotional state of the agent, and how such a state consistently affects the agent's evaluations of outcomes.

---

[94] Explicit modifications of SEU theory that involve the subtraction of a variance term (from the expected utility of the act) do not respect first-order stochastic dominance (Machina 1989, p. 1631).

But some, including Broome and Weirich, seek a more sturdy account of risk/regret sensitivity, one that does not appeal to emotions but rather to more objective properties of outcomes. Both Broome and Weirich can be interpreted as appealing to counterfactual properties of outcomes, where the counterfactuals refer to what alternative outcomes might otherwise have eventuated from the act in question. Counterfactuals offer a way to depict what appear to be "global properties" of an act as properties of individual outcomes. I find this account problematic, however, because such counterfactual propositions do not lend themselves to natural property groupings. Therefore, the move effectively makes SEU theory vacuous.

The upshot of all this, I think, is that if we want to accommodate any kind of systematic risk/regret sensitivity, then we should face up to the fact that we are talking about a violation of independence. This might be a rather predictable conclusion, given my starting position that we should be able to distinguish between SEU theory and alternative accounts of choice under risk/uncertainty. One could say that it is precisely because I do not think that SEU theory should be vacuous that I come to the conclusion that not all kinds of risk/regret sensitivity can be accommodated by the theory. For now, I will leave open the obvious next question as to whether a relaxation of independence should be considered rationally permissible. At stake is the normative status of theories like cumulative prospect theory. Such theories are much more flexible than SEU theory when it comes to determining whether an agent has rational preferences. There is essentially an extra degree of freedom with respect to representing the agent's ordinal preferences—as well as the utility and credence functions, we have a subjective risk factor to play with. Of course, in such case, the agent's ordinal preferences may be consistent with more than one (coherent) credence and value function combination, which certainly complicates the usual representation theorem story. The Dutch book justification of probabilism (which I analysed in Chapter 2) would also be compromised by a relaxation of the independence axiom (because the move would undermine the similar value-additivity-of-bets principle that the DBA depends on).

In brief, relaxing the independence constraint on rational choice has some significant consequences. In Chapter 6 I address this core question; I consider whether the sequential-choice context provides some stronger arguments for retaining the independence axiom. In the meantime, however, I will turn to another example decision problem that is illustrative of a general challenge to SEU theory—Ellsberg's (1961) problem. There is much in common between the Allais and Ellsberg problems, but the latter raises an additional question—should our representation of belief make a distinction between known risk and uncertain risk, or, more accurately, between "sharp" and "vague" degrees of belief? This is a further dimension to the risk/regret challenge to SEU theory.

5 ELLSBERG'S PROBLEM AND THE ORDERING AXIOM

5.1 Introduction

The last chapter was concerned with a measure of "risk" that pertains to the spread of outcome utilities for a given act. The question was whether, and to what extent, sensitivity to this kind of risk can be reconciled with SEU theory. I concluded that risk emotions can feature in an SEU model, but any more systematic treatment of risk involves a violation of the independence axiom. I now want to consider a further dimension to the "risk" story. It is one thing to worry about the spread of outcomes for an act when we are reasonably confident about its expected utility. Sometimes, however, it seems that we cannot specify what the expected utility of an act is, in the first place. One reason for this is that our beliefs/values concerning the outcomes of a given act may be, to some extent, indeterminate. This is to say that our beliefs/values may not always (and arguably never) have the "sharpness" of a real-valued number. For instance, perhaps my belief in whether it will rain tomorrow is simply that rain is more probable than not. Another possibility is that the actual proposition about which I am expressing an opinion is something that I take to be inherently vague. For example, consider the proposition expressed by "the population of Sydney is greater than 4 million". My belief in this proposition may be sensitive to the fact that Sydney has no definite boundaries (due to urban sprawl) and it is not clear who should be considered a resident of the city.

The examples above are illustrative of indeterminate belief, rather than indeterminate value, and indeed, in this chapter I will concentrate on the belief side of the story.

Much of what I have to say, however, will also apply to the modelling of any indeterminacy of value that might be associated with a given act's outcomes. To begin with, it is reasonable to think that both indeterminate belief and value can be well represented by numerical intervals—indeterminate belief by a probability interval and indeterminate value by a utility interval. (In Section 5.3, I point out that it is more accurate and informative to employ sets of probability/utility distributions to represent indeterminacy of belief/value, but intervals nonetheless provide a very convenient representation of the indeterminacy associated with individual outcomes.) Consider again my examples of indeterminate belief above. In the first case, my belief in rain seems best represented as Pr(rain) = (0.5, 1]. For the second case, my belief in the population of Sydney being greater than 4 million might plausibly correspond to the interval [0.3, 0.9]. The 0.3 lower bound is intended to correspond to the most stringent definition of the land-area of Sydney and who counts as a resident, while the 0.9 upper probability corresponds to a more widely-encompassing definition of Sydney's population.

It is worth emphasizing from the outset that I am considering the indeterminacy issue from a *subjective* standpoint. As per standard SEU theory, any probabilities within the decision model represent an agent's degrees of belief. I draw attention to this issue because it is tempting to start thinking in terms of objective probabilities when probability intervals (or sets of probability distributions) are introduced into the decision model. It is very natural to conceive of a probability interval as an estimate of the *true* probability of some event or proposition. For example, it might be thought that a probability interval for an outcome describes the narrowest range of probability values within which the agent is certain, or close to certain, that the precise objective probability of the outcome lies.[95] This use of probability intervals may have a place elsewhere, but it does not sit well with the subjective decision theoretic approach that I have been upholding in this thesis. Of course, objective probabilities have a role to play when it comes to specifying beliefs, whether these beliefs are determinate or indeterminate. Indeed, Lewis's (1980) Principal Principle holds that an agent's degree

[95] I am simply assuming here that there is some suitable account of the objective probability of an outcome.

of belief in the truth of a proposition should equate to their expectation of the relevant objective chance. Moreover, it is reasonable to think that an agent's beliefs will become "sharper", the more relevant frequency data, or other information about the objective chances, they come to learn. It is just that, at bottom, I am interested in probability intervals that represent actual states of belief, rather than confidence intervals for some sort of objective probabilities. The distinction is subtle but important, and I hope it will become even clearer as the discussion progresses.

Even when we restrict our attention to subjective interpretations of probability intervals/sets, there are different ways to conceive of how these intervals/sets match up with an agent's beliefs. My interest here is genuinely indeterminate belief (as per the examples I have given thus far), but we might alternatively employ probability intervals/sets to model an agent's uncertainty about their own (precise) state of belief. Levi (1985) refers to the latter kind of probability interval/set as an "imprecise probability". In Section 5.4, I discuss the importance of this distinction when it comes to determining whether an agent has consistent preferences. For the most part, however, I will be concerned solely with "indeterminate probabilities", i.e. probability intervals/sets that represent the relative vagueness of an agent's beliefs, rather than how well the agent knows their own (precise) states of belief.

It might be noted that I have thus far been concentrating on the intuitive reasons for modelling indeterminacy—reasons that appeal to the psychology of real persons. Importantly, in Section 5.3, I consider normative motivations for this sort of modification to the standard SEU decision model. To give focus to the discussion, however, I first introduce Ellsberg's (1961) well-known decision problem. The decision scenario that Ellsberg describes presents a good case for incorporating indeterminacy in a decision model. Furthermore, the problem focuses attention on whether it is rational for decision-makers to be sensitive to the "risk" associated with indeterminate belief. Indeed, when presented with Ellsberg's problem, many people make choices at odds with SEU theory. We cannot easily brush such findings aside as

demonstrations of human limitation and error,[96] and claim that apparently rational people all tend to make similar mistakes in Ellsberg-type cases. As per my discussion of Allais's problem, the important point is that in these special decision circumstances, apparently rational people uphold their "mistaken" choices (which I will refer to as the "Ellsberg choices"), even after the source of their "mistake" is exposed to them.

The "mistake" in question appears to be a violation of the independence axiom.[97] For this reason, the Allais and Ellsberg problems are often considered to be of a kind. But a number of theorists have already pursued the modelling of indeterminate belief/value in relation to Ellsberg's problem. And nearly all of these theorists, which includes Levi (1986 & 1997), Seidenfeld (1988a), Gärdinfors and Sahlin (1982), Weirich (2001) and Bandyopadhayay (1994),[98] have argued that a decision theory permitting indeterminacy of belief/value allows us to rationalise the Ellsberg-choices without bringing into question the independence axiom. The main part of this chapter (from Section 5.4 onwards) focuses on this very issue—whether we can rationalise the Ellsberg-choices by appeal to indeterminacy.

5.2 Ellsberg's Problem

Before proceeding further I will outline Ellsberg's problem: By the standard account,[99] you are told that an urn contains 30 red balls and 60 balls that are either black or yellow in some unspecified proportion. In problem *A*, you are offered two

---

[96] Although some theorists, such as Savage (1954) and Raiffa (1968, pp. 80–86), are in fact happy to assert that most people choose irrationally when presented with the Allais and Ellsberg problems.

[97] Independence is formally defined in the previous chapter (Section 4.2).

[98] I have left out Ellsberg himself (1961) because he seems to accept that the Ellsberg-choices show a violation of independence. Joyce (1999, pp. 101–102) also comments on Ellsberg's problem and the issue of indeterminate belief/preference, but he does not try to justify the Ellsberg-choices.

[99] Ellsberg himself (1961, pp. 653–654) describes the problem in this way.

options, I and II, and in problem *B* you are also offered two options, III and IV. *A* and *B* are distinct choice situations—you should not expect to play both games at once or in succession. In each case a ball is drawn at random from the urn and you receive specific rewards depending on the ball's colour and your chosen option. The payoff table is as follows:

Figure 5-1

	Red	Black	Yellow
Problem *A*			
I	$100	$0	$0
II	$0	$100	$0
Problem *B*			
III	$100	$0	$100
IV	$0	$100	$100

According to empirical tests, a significant majority of people choose option I in problem *A* and option IV in problem *B* (what I refer to as the "Ellsberg-choices").[100] More importantly, a number of reasonable-seeming people, after due reflection, purport to uphold this combination of choices. The paradox for the SEU theorist, or what I will refer to as the strict or precise Bayesian, is that whatever *precise* subjective probability an agent assigns to drawing a black, it is impossible that EU(I) > EU(II) and EU(III) < EU(IV) (where EU(#) stands for the expected utility of option #). According to this analysis, the only rational responses are to choose III and I, or IV and II, or to be indifferent between the two options in both cases.

Savage makes this case more striking by analyzing the situation in a manner that

[100] A number of empirical studies have shown that the majority of people respond to the Ellsberg problem in a way that *appears* to violate the independence axiom. See Ellsberg (1961, p. 654), MacCrimmon and Larsson (1979, pp. 372–6).

makes explicit the demands of the independence axiom: he notes that the winnings for a yellow ball being chosen are the same for both options in problem *A* and in problem *B*. Therefore, as per Allais's problem, we can ignore this column of the table for each game. But then it is very clear that the two games, *A* and *B*, are identical with respect to the remaining winnings. So your choice in problem *A* should be consistent with your choice in problem *B* (i.e. you should choose I and III, or II and IV or be indifferent in both cases).

Now a number of people have argued that we need not assign a *precise* subjective probability to a black ball being removed from the urn. And this changes the Ellsberg model, such that the Ellsberg-choices do not involve such an obvious violation of independence. I will come back to this reformulation of the problem in Section 5.4. For now, I will leave Ellsberg and consider the broader motivations for modelling indeterminacy of belief.

5.3 Indeterminacy and normative models

In Section 5.1, I illustrated the case for indeterminacy of belief/value by appeal to the psychology or real people. What is more important to this general discussion, however, is that it hardly seems a *requirement of rationality* that an agent have either an infinitely rich set of preferences, or associated partial beliefs with the determinacy of real numbers. Note that the various representation theorems justifying SEU theory appeal to a similar set of ordinal preference axioms, one of these being that an agent has a well-ordered set of preferences (implying that it must be a complete order, satisfying transitivity).[101] While a set of preferences that fails transitivity seems, at

[101] The "completeness" component of the ordering axiom is as follows (recall that "$\succ$" represents strict preference and "$\approx$" represents indifference):

> The decision maker has a definite preference with regard to every pair of propositions *X, Y* in the set of outcomes and acts, so that $X \succ Y$ or $Y \succ X$ or $X \approx Y$.

The "transitivity" component of the ordering axiom can be stated as follows:

first glance at least, to be irrational, the same cannot be said for a set of preferences that does not respect the completeness component of the ordering axiom.[102] Why is it such a bad thing to have no preference at all between two options? (It might seem that I have suddenly switched to talking about incomplete preferences rather than indeterminate beliefs, but the two are related via the representation theorems.[103] It is the completeness requirement on ordinal preference that prohibits indeterminacy in both the belief and value functions.) There doesn't seem to be anything wrong with absence of preference judgment. Completeness or real number precision seems to be a constraint that facilitates elegant mathematical modelling, rather than a feature of rational preferences.

Recall from Section 5.1 that I have been employing probability and utility *intervals* to represent indeterminate belief and value respectively, the width of the interval representing the amount of indeterminacy. Intervals are really just shorthand, however, for sets of precise probability and utility *distributions*.[104] Provided the set of distributions satisfies some reasonable constraints,[105] the probability for a singular

---

For any propositions/prospects X, Y and Z in the set of outcomes and acts,

if $X \succ Y$ and $Y \succ Z$, then $X \succ Z$,

if $X \succ Y$ and $X \approx Z$, then $Z \succ Y$,

if $X \succ Y$ and $Y \approx Z$, then $X \succ Z$,

and if $X \approx Y$ and $Y \approx Z$, then $X \approx Z$.

Note that these ordering axioms are very similar to those given in the presentation of von Neumann and Morgenstern's expected utility theorem in the Introduction, the difference being that the von Neumann-Morgenstern axioms of preference apply to "lotteries", rather than propositions. (Joyce 1999, p. 84) gives a similar presentation of "completeness" and "transitivity" (although he refers to both strict and weak preference relations).

[102] Joyce (1999, p. 45) makes a similar point about completeness. In fact, the decision theory I go on to focus on here (Levi's E-admissibility) relaxes transitivity as well as completeness with respect to ordinal preferences. (I do not consider the violations of transitivity entailed by the theory to be cause for complaint however.)

[103] Refer to Savage's (1954) representation theorem, for instance.

[104] Weatherson (2002, p. 3) notes that the standard treatment of non-precise probabilities involves sets of precise probability distributions; it is the common ingredient of prominent theories of non-precise probabilities, such as that of Levi (1980) and Walley (1991).

[105] The set must be convex. Note that set C is said to be convex if, for all x and y in C and all t in the interval [0, 1], the point $[(1 - t)\, x + t\, y]$ is in C.

state or outcome can then be expressed as an interval. The situation is analogous for utilities: the agent's preferences may be properly characterised by a set of utility distributions, and for a single outcome this yields a utility interval. The important thing to note is that if an agent has an incomplete set of preferences (that otherwise conform to SEU theory), then this will be consistent with a number of pairs of probability and utility functions; complete preferences, by contrast, are consistent with just one probability-utility pair (up to positive linear transformation of the utility function).

Some have argued that relaxing ordering, and in particular the completeness requirement, is unhelpful because there will still be a lurking problem of vagueness in belief/preference at the level of interval (set) boundaries (Howson and Urbach, 1989, pp. 68–70). Walley (1991, pp. 250–1) refers to this as the "two precise numbers instead of one" complaint. The worry here seems more applicable, however, to the use of intervals/sets to model an agent's uncertainty about their own beliefs/preferences. Indeterminacy is quite a different phenomenon; it was never intended to "solve" the problems associated with actually eliciting beliefs and preferences. In practice people may well find if difficult to determine what their state of preference is with respect to two options. The possibility of having no preference at all might be thought to complicate the issue—it means that we have to decide whether we prefer prospect *X* to *Y*, *Y* to *X*, are indifferent between *X* and *Y or have no preference between them*. Even in practical settings, these kinds of concerns about being in touch with one's own preferences are not always applicable. Indeed, the Ellsberg scenario is a case where the obvious choice of interval (set) boundaries for the probabilities of yellow and black are precise ones. An agent with beliefs of this character presumably has unambiguous preferences with respect to the kinds of bets that they would be prepared to buy and sell at the boundaries.[106] In any case, even if there are lurking problems when it comes to eliciting preference, the argument remains that an ideally-rational agent need not have a complete preference ordering. No preference at all between options is permissible. And in such cases, real-number

[106] If an agent's beliefs were represented by a probability set with vague boundaries, then their betting preferences would be vague at these boundaries. I discuss the relationship between indeterminate belief and betting preferences in Section 5.7.

values will not provide accurate representations of an agent's beliefs and desires.

There are some other more positive reasons for appealing to belief indeterminacy in the context of a normative decision model. I note that Levi (1974) thinks it is essential to the revision of partial beliefs that they take the form of non-precise probabilities.[107] I will not explore this idea further here, because there is another argument more pertinent to decision modelling. (Again it focuses on the belief side of things rather than the desire side.) Walley (1991, pp. 207–219) effectively argues that our partial beliefs *should* vary in both strength and "sureness", with the sureness depending on the amount of evidence on which the belief is based. In Walley's (1991, p. 207) own words, the degree of imprecision in probabilities should be directly related to the "amount of statistical information on which they are based, and to the degree of conflict between statistical and prior information".

Consider the following two scenarios, in which you are asked to represent your degree of belief that horse *A* will beat horse *B* in the Melbourne Cup: in the first instance, you have spent much time at the tracks watching the horses race and even have considerable frequency information regarding their form, you have researched their training regimes, have received tests on their respective physical conditions just prior to the race, etc. etc., but in fact none of this information distinguishes the two horses. Contrast this with the second situation, in which you know absolutely nothing about the two horses, indeed you haven't even seen them, you do not know their names, their trainers (or even if they have trainers) or their history. Not only are you likely to have different sorts of beliefs about whether *A* will win the race in these two situations, we might argue that you indeed *should* have different sorts of beliefs. The fact is, in the first case you have a lot of information to the effect that the match is even, and in the second you really have no basis for any particular amount of confidence in *A*.

[107] Levi thinks non-precise probabilities are necessary if we want to be able to update our beliefs due to a change of mind, and not in response to any new evidence. Note that this is a contentious account of how beliefs might change.

Acknowledging indeterminacy of belief, moreover, allows us to avoid appeals to the Principle of Insufficient Reason.[108] Such an appeal goes like this: given no information, you decide that the probability of *A* winning is 1/2, because this corresponds to the flat (symmetrical) distribution, or in other words represents maximum uncertainty.[109] Walley (1991, p. 234) dismisses this insufficient reason rationale (with some rhetorical flourish!) as the "nonsense of non-informative priors". Indeed, many have criticised the Principle of Insufficient Reason for being arbitrary, in that it supports an infinite number of probability distributions, depending on how the agent cares to partition the outcome space.[110] The principle is supposedly a means for assigning probabilities under conditions of complete ignorance, but it appears to presuppose prior knowledge about how the outcome space should be partitioned into equi-probable states! Now I do not want to get into the business here of making judgments about what are and what are not good partial beliefs—judgments that go beyond the criteria for consistency. I think it is a feature of the non-precise model, however, that it does not require an agent to adopt somewhat arbitrary precise beliefs when it seems entirely rational for them to be in a state of (at least partial) ignorance.

5.4 Back to Ellsberg

As mentioned, a number of people have remodelled the Ellsberg problem using sets of candidate probability distributions in lieu of precise probabilities. And most of

---

[108] The principle is otherwise known as the "Principle of Indifference". Gillies (2000, p. 35) gives Keynes' preliminary statement of the principle:

> The Principle of Indifference asserts that if there is no *known* reason for predicating of our subject one rather than another of several alternatives, then relative to such knowledge the assertions of each of these alternatives have equal probability.

[109] Walley (1991, p. 211) states that the degree of uncertainty expressed by a probability distribution amounts to the variability or dispersion of the distribution, which is commonly measured by its entropy, variance or standard deviation.

[110] This problem for the "Principle of Insufficient Reason" is well illustrated by Bertrand's paradox. See Hájek (2003, pp. 187–188).

these people have also argued that reformulation of the problem in this way can explain the Ellsberg choices.[111] While I am not so confident about this latter claim (and will go on to discuss this), I think the Ellsberg scenario is just the kind of case where it seems reasonable to represent the strength *and* (in)determinacy of partial belief. Indeed there are no clues whatsoever about the relative proportions of black and yellow balls; from the problem description, we know only that the proportion of each could be anything between 0 and 2/3. In more realistic decision scenarios, we generally have a stock of background knowledge that informs our beliefs in subtle ways. But given our complete lack of discerning information in this situation (amongst the values between 0 and 2/3) it is very natural for an agent to have a less-than-precise degree of belief about whether a yellow or black ball will be drawn from the urn. In particular, it would be very natural to have degrees of belief corresponding to Pr(black) = Pr(yellow) = [0, 2/3]. I am representing the probabilities as intervals, but recall that it is more accurate to think in terms of sets of candidate probability distributions over the possible states. In our Ellsberg case one obvious candidate distribution for <Pr(red), Pr(black), Pr(yellow)> is <1/3, 0, 2/3>, another is <1/3, 1/3, 1/3>, another is <1/3, 2/3, 0>, and so on. The convex set of all such candidate probability distributions yields the interval [0, 2/3] for the probability of both yellow and black.

Using sets of probability distributions to model the Ellsberg problem gets us corresponding expected utility sets for each option. In particular, if Pr(black) = Pr(yellow) = [0, 2/3], we get the following expected utilities (where U is the agent's utility function, and with expected utility expressed shorthand as intervals rather than as functions over the candidate probability distributions):

[111] Recall my comment in the introduction to this chapter about those theorists who model Ellsberg's problem with non-precise probabilities, and who find that a model of this sort allows a justification of the Ellsberg-choices. Note that these theorists nonetheless rationalise the Ellsberg-choices in slightly different ways. Moreover, there are others who suggest that a robust justification of the Ellsberg-choices would require more significant changes to SEU theory than the incorporation of indeterminacy. (I go on to argue for this latter position.)

Figure 5-2

	Expected Utility
Problem *A*	
Option I	1/3 × U($100)
Option II	[0, 2/3 × U($100)]
Problem *B*	
Option III	[1/3 × U($100), U($100)]
Option IV	2/3 × U($100)

So does this model legitimate the Ellsberg-choices: a preference for I over II, and for IV over III? It seems so. Options II and III do not have point expected utilities; rather they have expected utilities that span intervals, and this suggests that there is room to move in terms of comparing the desirability of the pairs of options in both problems *A* and *B*. In particular, given that neither option is clearly preferred in each case there seems to be scope for an agent to choose I in problem *A* and IV in problem *B*.

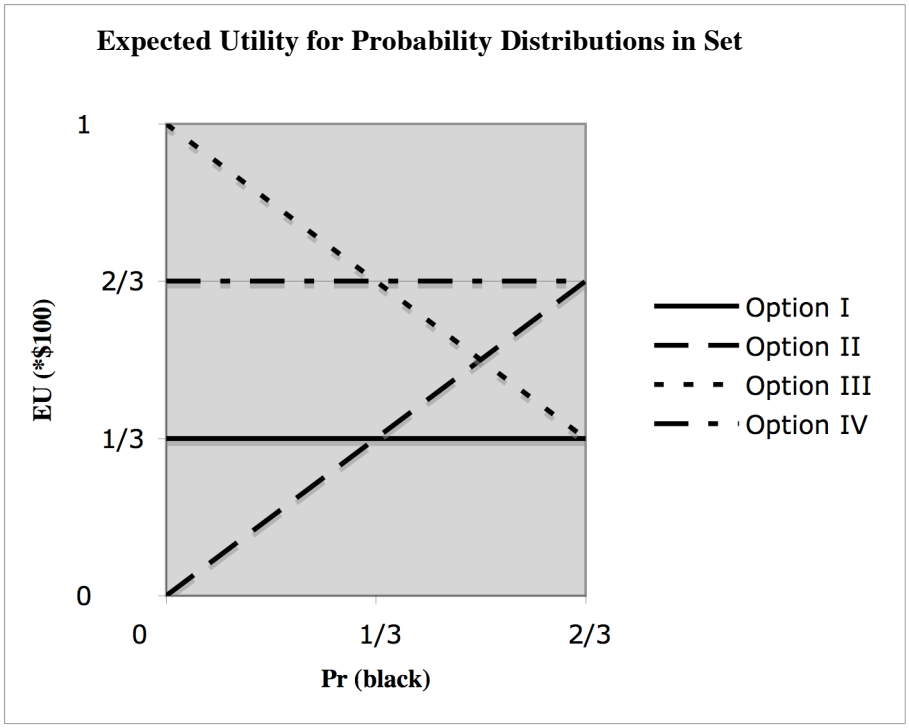
To begin with, however, whether the introduction of probability sets makes the Ellsberg-choices admissible depends on how these sets are interpreted. It is all very well to say that the requirement for agents to have degrees of belief represented by precise probabilities seems unnecessary, especially in scenarios like Ellsberg's problem. In Section 5.1, I drew attention to two possible interpretations of sets of probabilities, even when we start with the assumption that they should represent subjective belief. The first interpretation is more continuous with the strict (precise) Bayesian view—it is the view that probability sets represent an agent's ignorance about their own state of subjective belief (and related preferences). Here we model the agent as being unable to specify the correct representation of their belief state, but any candidate belief state amounts to a precise probability distribution. Levi (1985) uses the term "imprecise probability" to refer to this interpretation of the probability set. (I

will employ Levi's terminology here, but note that others use the term "imprecise probability" differently.) The second interpretation of probability sets is better suited to the idea of genuine indeterminacy of belief, and indeed, this has been my chief target in this chapter. It is the view that beliefs can be properly represented as probability sets, where the distributions within the set *together* represent the agent's degree of belief. Levi (1985, p. 392) describes the difference as follows:

> …a set of probability functions may be used to characterise the credal state as a set of permissible probability distributions and not as a set of possibly true hypotheses concerning the unknown uniquely permissible distribution.

Figure 5-3 serves to illustrate the consequences that the two different interpretations of probability sets have with respect to what are admissible choices in the Ellsberg scenario.

Figure 5-3



If probability sets are "imprecise probabilities" in the sense referred to above then the

agent is committed to choosing consistently with respect to at least one precise probability distribution within the set. For the Ellsberg case, once the agent commits to a particular option in situation *A* then this constrains how they should choose in situation *B*.[112] Note from the graph that for each individual probability distribution, if I is chosen then III should be chosen, and if II is chosen then IV should be chosen. (Where the line for option I is above the line for option II, the line for III is above the line for IV, and so on.) There is also the case corresponding to Pr(black) = Pr(yellow) = 1/3, where the agent is indifferent in both cases. In fact, only this latter probability distribution even *permits* the agent to choose options I and IV, let alone have a considered *preference* for these options.

If probability sets are thought rather to represent genuinely indeterminate belief, then determining what are the admissible options in each case is a very different matter. Unlike the "imprecise" model, there is no requirement of consistency with respect to any particular probability distribution within the set. A number of suggestions have been made regarding what is the appropriate decision rule here. Given that the overwhelming response by people is to choose the "safe" option in each Ellsberg scenario, some have suggested formalizing a "risk-averse" (or security-conscious) rule for dealing with indeterminate expected utilities. Gärdinfors and Sahlin (1982), for example, have recommended what can be referred to as a "lexi-min rule", which essentially recommends choosing the option with the greatest minimum expected utility. It is an attractive rule in terms of the Ellsberg problem because it actually prescribes the Ellsberg-choices: option I in problem *A* and option IV in problem *B*. Others propose more minimalist decision rules. Kyburg (1983), for instance, recommends a rule mandating preferences between options with interval expected utilities only in the clear cut cases: for any two options, if the minimum expected utility of one is greater than the maximum expected utility of the other, then there is a strict preference in favour of the former; in other cases the two options are simply incommensurable. Bandyopadhayay (1994) proposes a slight variation of Kyburg's

[112] The requirement that the agent choose consistently does not rest on the fact that the Ellsberg bets (*A* and *B*) are made at the same time or else one after the other. (Such a betting scenario is quite different from the Ellsberg problem and yields different expected utilities.) The requirement is rather about consistency in belief (given any particular precise probability distribution in the set).

rule that he calls "weak dominance". In the minimalist spirit, Bandyopadhayay claims that it is pointless to aim for a rule that always selects a single option (or a number of indifferent options) as optimal.[113]

There has been compelling criticism of all the above-mentioned rules. Levi (1986, pp. 137–138) argues that "lexi-min" is not a proper extension of SEU theory because the rule can recommend a choice that does not maximise expected utility under *any* probability function.[114] Seidenfeld (1983) further claims that a choice rule is rational if it deems admissible all and only those options that maximise expected utility for *some* permissible probability and utility pair. In fact, neither Kyburg's rule nor Bandyopadhayay's "weak dominance" (in addition to "lexi-min") fulfil this condition. Seidenfeld et al. (2004) show that if the condition is not satisfied, chosen options may be dominated if the option set is augmented to include all mixed options.[115] This is a compelling argument, because it is always possible for an option set to include mixtures, and so we do not want our chosen option to be otherwise dominated by a mixed option. I will not here go into further details regarding what rationality dictates in terms of comparing options with indeterminate expected utilities. In brief, I find the Levi/Seidenfeld analysis compelling. According to this rule, both Ellsberg options are admissible in scenarios *A* and *B*, because in each case, both maximise expected utility for at least some permissible probability distribution.

---

[113] Kyburg and Bandyopadhayay both suggest that further discriminations can be made amongst the rationally permissible options according to secondary security considerations (that may be specific to the agent). In this respect their approaches are similar to Levi's. Indeed, it is this aspect of Levi's account (the appeal to secondary criteria) that I investigate in the following sections.

[114] The assumption here is that we seek a good extension of SEU theory or strict Bayesian decision theory. Levi (1986, pp. 137–138) also notes that the "lexi-min" rule clearly violates the independence axiom. Furthermore, Levi is reluctant to commit to a decision rule that *stipulates* risk aversion.

[115] In fact, even without mixed options, Kyburg's rule can deem an option admissible, even if it is strictly dominated by another option with respect to every pair of probability-utility distributions in the set (see Seidenfeld 1983).

5.5 A place for "risk" attitudes?

If our model caters for indeterminate belief, it is thus permissible to choose options I and IV in the Ellsberg scenario. In the end, one option must be chosen in each of situations *A* and *B*, and it might as well be these ones. It looks like we have thereby rationalised the Ellsberg choices (in a more robust sense than by appeal to indifference between the options, which relies on the agent having the specific belief function that gives Pr(black) = Pr(yellow) = 1/3). Not so fast, however. The Ellsberg choices are considered paradoxical because they stem from an actual preference for option I over II and for IV over III. There would be no paradox to speak of if these choices amounted to a mere arbitrary selection of one or other of the options in each case. The point is that Ellsberg choosers would *not* be happy with just any option in each of situations *A* and *B*. Ellsberg choosers specifically want option I in situation *A* and option IV in situation *B*.

Arguably, if an agent is indifferent between two options, then, by definition, the agent does not have a preference between the options. But incommensurability of options might be another kind of matter altogether, and may well allow some kind of final preference amongst the admissible options. Now there is clearly a difference between saying that two options are incommensurable and saying that we are indifferent between them. Following Walley (1991, p. 209), we can say that for options that we are indifferent between, increasing either by any small amount of utility will result in that option being preferred. For incommensurable options on the other hand, more comprehensive changes in utility are needed for one option to be definitely preferred to the other. Another way of describing the distinction is in terms of happiness or satisfaction. If I am indifferent between two options then I expect them both to bring me equivalent amounts of happiness (since they both have the same utility). I expect no such thing with incommensurable options; in this case I simply cannot decide which will bring me the greatest amount of happiness (Joyce 1999, p. 101).[116]

[116] Furthermore, if we were to consider incommensurability as a kind of preference relation, it would not even obey the same constraints as a relation of indifference. For instance, the indifference relation (represented by "≈") is transitive: for prospects *X*, *Y* and *Z*, if $X \approx Y$ and *Y*

Despite the ways in which the two relations are disanalogous, I think that there is reason to think that incommensurability and indifference have the same implications when it comes to a choice situation. Either way, if we must make a choice *now* we are faced with a dilemma. Incommensurability and indifference are simply different ways of not being able to choose between options. The problem is essentially the same in both cases: we are forced to choose a single option but we do not actually prefer one option to the other. Either way we are faced with indecision, and any choice of a singular option will be arbitrary. We might argue that the choice of just one option *should* be arbitrary in both types of case, because if it were based on any further evaluations of goodness/desirability, then we have made a mistake in our description of the decision problem—our preferences and thus our utility function should already take into account everything that we care about. If there is no room for differing attitudes towards "risk" in the case of options that we are indifferent between, then neither should there be scope for differing attitudes towards "risk" (security) when comparing incommensurable options.

At this point I note that Weirich (2001 & 2004, pp. 77–9) appeals to "comprehensive outcomes" in order to explain Ellsberg's paradox. Weirich argues that it is possible to rationalise the Ellsberg response if the outcomes are specified in a way that captures everything the agent cares about in the decision situation. If the agent has indeterminate degrees of belief with respect to whether a black or yellow ball will be drawn from the urn, then this indeterminacy may well affect how the agent evaluates any outcome associated with drawing one of these colours. Given that the Ellsberg chooser does seem to avoid choices resting on indeterminate beliefs, then it is likely the case that the outcomes associated with yellow or black are all some amount less than the utility of $100, due to the extra angst associated with gambling on "risky" probabilities.[117] So Weirich models the Ellsberg problem with indeterminate

---

$\approx Z$ then $X \approx Z$. According to the choice rules I outlined in the last section, this is not necessarily the case with incommensurability (here represented as "$\approx_C$"). We might have the case where $X \approx_C Y$ and $Y \approx_C Z$, but it is NOT the case that $X \approx_C Z$.

[117] In the context of choice, Weirich refers to non-precise belief functions as "risky" probabilities.

probabilities and utilities, but he does not rationalise the Ellsberg choices by appealing to incommensurability between the options in both cases. Weirich thinks the Ellsberg choices are rational, rather, if the outcomes in *A* are specified so that option I is better for all probability distributions in the set, and the outcomes in *B* are specified so that option IV is better for all probability distributions in the set.

I agree with Weirich that it is possible to explain the Ellsberg paradox by appealing to "comprehensive outcomes". If the agent has a clear psychological reaction (in this case negative) towards gambling on "risky" probabilities, then it seems correct to include this emotion within the relevant outcomes. I support such appeals to "risk" (security) attitudes affecting outcomes only in the limited sense, however, where there is a clear emotional reaction involved. I have argued in the previous chapter that more methodical inclusion of risk attitudes in outcomes is counterproductive to SEU theory (whether precise or non-precise) because it makes the theory practically indistinguishable from competitors. In any case, my focus in this chapter is not Ellsberg's problem *per se*. There may well be alternative ways to rationalise the Ellsberg choices. Here I am interested in whether a model that caters for indeterminate belief *but has outcomes defined in the usual way* can yield the Ellsberg *preferences* for option I over II and for option IV over III.

5.6 Levi's distinction between levels of preference

Levi (1986, pp. 122–140) appeals to secondary or tie-breaking preferences *both* for cases of incommensurability and for cases of indifference. Briefly, he claims that we may be unable to choose between options (whether because we are indifferent between them or because we find them incommensurable) with respect to our primary preference ranking, but other considerations may be called upon to break ties. As mentioned, I support Levi's choice rule for comparing options with indeterminate expected utilities. According to this rule, options are admissible if and only if they maximise expected utility for some permissible probability-utility distribution pair.

Levi refers to the options that fulfil this criterion as the "E-admissible" options. I have worries about the second-step in Levi's formal framework, however, where secondary preferences enter in: he claims that amongst the "E-admissible" set of options an agent may choose a final "V-admissible" set according to their security preferences. In other words, amongst the rationally *permissible* options in a decision problem an agent can determine further preferences on the basis of security. While Levi (1986, p. 127) recommends maximising security amongst E-admissible options (selecting those options with the greatest minimum expected utility), he points out that any attitude towards security is permissible—one's sentiments about security cannot be stipulated by rules because this is rather a substantive value consideration.

Levi does not make a special case for *incommensurability* allowing scope for secondary judgments about risk or security; he claims that such secondary judgments are pertinent in cases of indifference as well. In other words, according to Levi, when more than one option maximises expected utility we can use some optional measure of risk (such as utility of the worst-case outcome or the variance in utility) to discriminate between the tied options. Levi argues (1986, p. 131) that this does not contradict the independence axiom, because independence applies to preferences based on the initial expected utility calculations. In fact, Levi (1986, pp. 130–131) claims that the Ellsberg choices can be rationalised in this way without even appealing to indeterminate belief. He does not find it a very satisfactory response to the paradox, however, because the account is peculiar to the special case where the agent's subjective probabilities for the coloured balls are Pr(red) = Pr(black) = Pr(yellow) = 1/3. In this case, the expected utilities for the pairs of options in both problems are equivalent. But while we have a relation of indifference in each case with respect to expected utility, Levi argues that the agent may then turn to risk considerations in order to make a final choice amongst the admissible options. (I do not see how this works, because if the probability of each colour is taken to be 1/3, not only do both options in each Ellsberg scenario maximise expected utility, the distribution of outcome utilities and thus the riskiness of the options in each case is also identical.) In any case the main point here is that Levi thinks tie-breaking preferences are operable in cases of indifference as well as cases of incommensurability.

Levi himself prefers the indeterminacy response to the Ellsberg paradox. When such a model takes the probability and outcome utility values referred to earlier, then both options are incommensurable and thus rationally permissible in each of the choice situations *A* and *B*. This result does not hinge on the agent having a particular belief function that seems somewhat unmotivated given the lack of discriminating information in the problem description. According to Levi, further distinctions between admissible options can be made on the basis of secondary security considerations. In this way, the Ellsberg choices can be deemed actual preferences without their violating further axioms of SEU theory besides ordering, such as the independence axiom. But the proviso here is that we must apply axioms of preference only to the first-stage or primary preference ordering and not to secondary tie-breaking preferences. It is this proviso that I challenge.

5.7 Applying axioms of preference

I find Levi's lexical decision rule intuitively very reasonable.[118] When it comes to Ellsberg's problem, the final selection of options I and IV via a process that involves first selecting the admissible options with respect to expected utility considerations and then maximising security amongst these options seems plausible. What I question here is whether we should regard such a preference structure as being compliant with the independence axiom.[119] Does it make sense to separate different tiers of preference relations, and apply axioms of preference just to the first or primary preference ordering? I question whether the account is compatible with representation-theorem

[118] A choice rule can be described as lexical if it recommends comparing options along an ordered set of dimensions, the preference relation between the options being established according to the first dimension in which the options are not considered indifferent. (For Levi's rule, E-admissibility is the first criterion, and E-admissible options may be further discriminated according to security considerations, or V-admissibility.)

[119] Note that while Levi's lexical choice rule seems very plausible, if it is deemed to violate the independence as well as the ordering axioms of SEU theory, then some further investigation of the import of such violations is called for. I pursue this issue in Chapter 6.

justifications of decision rules, or the gambling odds methods for hypothetically eliciting belief that presuppose such decision rules.

Levi (1986, p. 131) discusses the potential conflict between his lexical choice rule and the independence axiom. He notes that if preference is defined as revealed preference resulting from application of his lexical choice rule, then independence is violated, at least in those cases where the agent is not neutral with respect to security. But if preference is understood as the primary preference associated with expected utility calculations alone (as per E-admissibility), then independence remains intact (and only the ordering axiom of SEU theory fails). I am sceptical, however, about whether this distinction between basic and revealed preference is justified. Of course, in practical settings there are many reasons for distinguishing between actual preference and preference revealed through choice. People often make choices without due reflection, and yet when experimentally eliciting preferences we generally assume that choices are founded on considered preferences. Moreover, in cases where a person simply picks randomly, because for whatever reason they cannot decide which option is better, it appears that the chosen option is preferred when in fact it isn't. These are common criticisms of behaviourism as a method for eliciting preferences. In the experimental context, it is certainly useful to distinguish between revealed preference (what is elicited) and actual preference (what we are trying to elicit). But in the idealised theoretical context, there are no such problems with elicitation. If the agent's considered choice rule is a lexical rule whereby first expected utility and then security considerations are taken into account, then discriminations between options stemming from this choice rule surely represent the agent's actual preferences, not their rough-and-ready revealed preferences.

Levi draws a distinction, not between considered preference and somewhat sloppy actual choice behaviour, but between two underlying preference orderings that are both well considered—the primary preference ordering and the final preferences resulting from his lexical choice rule. But arguably the two should be in sync, especially if one wants to uphold the representation theorem rationale for choice rules. The representation theorems (e.g. Savage's (1954) or Jeffrey's (1983) theorems) rely

on a pragmatic understanding of preference. These theorems take us from an ordinal preference ranking, which is supposedly something we can get a handle on *given its pragmatic implications*, to an expected utility representation that many consider to be somewhat fictitious when it comes to the psychology of actual agents. The potential problem with introducing different tiers of ordinal preference in the indeterminate case is that it threatens the neat connection between preference and its pragmatic implications. According to the usual story, the agent is either disposed to choose option *A* over *B*, or they are disposed to choose *B* over *A*, or else they are genuinely ambivalent about which option is selected. It seems to me that an agent cannot have at the one time multiple choice-dispositions with respect to the options *A* and *B*. In which case, some would argue that the agent cannot entertain different tiers of ordinal preference, one corresponding to E-admissibility alone, and another corresponding to V-admissibility, or all-things-considered preference.

My arguments in Chapter 2 (particularly Sections 2.6 and 2.11) should suggest that I am not in fact committed to the view that belief and preference are comprehensible only in terms of their implications for choice. In fact, I am sympathetic to the position expressed by Shafer (1986), that we can *construct* ordinal preferences from independently established belief and utility functions. I say a bit more on this issue in the conclusion of this thesis. At this point, however, what I want to emphasise is that the incorporation of indeterminacy in a decision model need not amount to such a large departure from established decision theory wisdom. As mentioned, when we introduce indeterminate belief/preference we give up the completeness axiom of SEU theory. But we need not appeal to different tiers of preference, and thus give up the representation theorem emphasis on the pragmatic character of belief and desire.

I further appeal to the gambling odds method for quantifying belief because it too relies on the fact that an agent's (hypothetical) choice behaviour reflects a single ordinal preference ranking. The gambling odds account of precise partial belief does not sit very well with the idea of there being different tiers of preference ordering. Degrees of belief are determined according to expected utility considerations, and these are the *only* considerations that are assumed to affect an agent's preferences or

disposition to choose amongst gambles. If this rationale is extended to indeterminate partial belief, then here again the gambles an agent is willing to buy or sell should accord with a single preference ordering. (Although, when it comes to indeterminate belief, there is scope for interpreting the buying and selling prices of gambles in different ways.)

According to the Bayesian story an agent's personal probability that a particular proposition *Q* is true depends on what gambles they would be prepared to make. More precisely, the agent is asked to consider a bet that pays $1 (or one unit of utility) if the proposition *Q* is true and $0 if *Q* is false. The agent's buying price is the maximum price at which they would buy the bet, and the agent's selling price is the minimum price for which they would sell the bet. The strict (precise) Bayesian holds that these two values are equivalent—there is a "fair price" at which the agent is indifferent between buying and selling the bet. Note the importance of the indifference relation here—it is directly related to the agent's hypothetical betting behaviour and is critical to the story about how we determine an agent's degrees of belief. If the agent ultimately based their choices on secondary security considerations, then the gambling odds story would require revisions to this effect.

Where indeterminate belief is concerned, there is a distinction between the buying price and selling price of a gamble. I contrast two interpretations of what such a distinction amounts to. The first I refer to as the glutty interpretation and the second I refer to as the gappy interpretation. According to the former, the agent considers buying the gamble on proposition *Q* at a price below the lower bound as *uniquely* admissible, and selling the gamble at a price greater than the upper bound as *uniquely* admissible; at prices in between the upper and lower bounds, *both* buying and selling are admissible. This means that at prices below the lower bound you should definitely buy and for prices above the upper bound you should definitely sell, but at prices in between you can *either* buy *or* sell because both are admissible.[120] According to the

[120] Note that this doesn't necessarily make the agent vulnerable to a Dutch book, so long as the agent is committed to making rational *series* of bets, as well as rational individual bets. Weirich (2001, pp. 439–440) also makes this point.

gappy interpretation, by contrast, rather than being prepared to buy *or* sell at prices in between the upper and lower bounds, the agent is *neither* prepared to buy *nor* sell at these intermediate prices. In other words, the lower bound is the agent's maximum buying price for a gamble on Q and the upper bound is the agent's minimum selling price.[121]

As per the case of precise partial belief, neither the glutty nor the gappy gambling odds accounts of indeterminate belief countenance multiple preference orderings or choice dispositions. Admissible bets are those the agent finds acceptable in a given choice situation. It would confuse the issue if the agent actually had further preferences for particular bets amongst the admissible ones. Ellsberg problem *A* provides a good illustration. Let us assume that the agent assigns personal probability [0, 2/3] to Pr(black). This means that there is no (positive) price at which the agent would *definitely* buy the gamble that pays stake S (here \$100) if black is pulled from the urn and nothing otherwise. The agent would *definitely* sell the aforesaid gamble only for a price greater than $2/3S$. Recall that the alternative—option I—has value $1/3S$. According to the glutty interpretation, the agent therefore considers both buying the gamble on black at $1/3S$ and selling the gamble on black for $1/3S$ admissible, so effectively the agent cannot choose between the two options in Ellsberg problem *A*.

The gappy account of Ellsberg problem *A* provides an alternative story. According to Levi, this case corresponds to an agent who is concerned to maximise security. I am not convinced that this is correct. A security-conscious agent would always opt for the sure $1/3S$ because this option has the greater worst-case expected utility. According to the gappy gambling odds account, however, if the agent starts off with the choice of red valued at $1/3S$, then they will sit on this option because they are unwilling to buy the bet on black for any positive price. However, if the agent starts with the gamble on black, then they will sit on this option because they will be unwilling to sell the

[121] Both Walley (1991, p. 211, pp. 250–1) and Levi (1986, p. 122–3) countenance both the "glutty" and "gappy" gambling odds interpretations of indeterminate probabilities. Levi identifies the "glutty" interpretation with the preferences of an agent who chooses by E-admissibility alone, and the "gappy" gambling odds interpretation with the preferences of an agent whose final choices maximise security (the S-admissible choices).

gamble for anything worth less than 2/3*S*. In such case the agent's choice in the Ellsberg situation will depend on which option they consider the default or status quo option. So I think there is some discord between the choice behaviour of Levi's security-conscious agent, and the gappy gambling odds account. (And indeed I am not sure how to place the gappy account.) In any case, the gambling-odds story assumes that an agent's choice dispositions correspond to their preferences. In other words, there is no distinction between underlying preference and choice. Furthermore, the agent cannot have more than one kind of gambling disposition at the same time.

5.8 Conclusions

I draw attention to the role of the representation theorems and gambling odds accounts of partial belief not because I think a lexical choice rule of the kind proposed by Levi is unreasonable. I find Levi's rationalisation of the Ellsberg choices prima facie very plausible. To begin with, the Ellsberg problem is an excellent illustration of the reasons for depicting not just strength but also sureness of belief. In this example decision situation, it does not seem reprehensible for an agent to have beliefs of varying determinacy with respect to the colour of the ball that will be drawn from the urn. And sets of probabilities seem a good way to represent indeterminate beliefs. Provided these probability (and corresponding utility) sets are understood to depict genuinely indeterminate belief and desire, the Ellsberg choices can be shown to be rationally permissible. Indeed, when the decision model is adjusted to incorporate indeterminate belief (in the most obvious way), both options are admissible in each of the Ellsberg scenarios. But the Ellsberg problem is paradoxical to the extent that agents have an actual *preference* for option I in scenario *A* and option IV in scenario *B*. This is where Levi's secondary security considerations enter in; he claims that amongst the admissible options, an agent may form further preferences on the basis of security considerations.

While a lexical choice rule is attractive (where we have indeterminacy of

belief/value), I think Levi glosses over the significance of his distinction between primary and secondary preferences. In the last section, I emphasised the fact that this move does not sit well with the standard reading of the expected utility representation theorems, or with the gambling odds account of partial belief. The latter method for quantifying belief (whether determinate or indeterminate), in particular, assumes that an agent's choices amongst bets represent their preferences. One might downplay the identification of ordinal preference with choice dispositions, but I think this move still leaves the representation theorems on shaky ground. If we want to understand these theorems as a means of grounding numerical belief/desire functions in ordinal preference, then the possibility of different tiers of preference rankings surely confuses the issue—why should one preference ranking and not another correspond to an agent's beliefs and values?

Perhaps, in the end, these worries about the origins of numerical belief and value functions are not so important. Indeed, I am sympathetic to the view that we should be able to understand the properties of rational belief, if not desire, without recourse to ordinal preference. Moreover, it is reasonable to think we can at least *entertain* multiple preference orderings, even if our choices ultimately reflect our final all-things-considered preferences. But even granted these points, it is my opinion that a choice rule should be assessed in terms of its final rankings of options. If we want to know what properties or axioms of preference a choice rule upholds, it seems most reasonable (even for a lexical rule) to look to the final, all-things-considered preference rankings that are consistent with the rule.

Given these points, I suggest that Levi's lexical choice rule involves a more significant break with standard SEU theory than just relaxation of the ordering axiom. In particular, if the lexical rule is identified with a single preference ordering, and the agent in question is not security-neutral (the agent might make their final choices on the basis of, say, worst-case expected utility), then we have a violation of independence. While I have indicated that such a choice rule seems intuitively reasonable, it remains to be seen whether it can survive further scrutiny. In the next chapter, I will consider what are arguably the strongest arguments for upholding the

ordering and independence axioms. We have seen that there are already some good reasons for wanting to retain the independence axiom, but also some good reasons for wanting to relax it. In my opinion, a lexical choice rule that permits "risk-sensitivity" where indeterminacy of belief/value is involved provides the strongest case for relaxing both ordering *and* independence. (I am not so concerned about defending the "Allais-choices", but the "Ellsberg-choices" seem intuitively rational, irrespective of the agent having some emotional response to "risky probabilities".) For this reason, when I go on to assess the relevant arguments concerning preference axioms, my initial position (or prior belief set, if you like) tends towards some violations of ordering and independence being permissible.

6 ASSESSING DECISION RULES IN THE SEQUENTIAL-CHOICE CONTEXT

6.1 Introduction

We have seen that there are some compelling challenges to SEU theory. In this second part of the thesis I have drawn attention to various alternative decision theories that have been motivated by considerations of "risk-sensitivity". For instance, "cumulative prospect theory" can accommodate sensitivity to the specific distribution of outcome utilities associated with an act, while Levi's decision theory handles the "risk" that is associated with indeterminate belief and value. My preliminary investigations show these theories (and particularly the latter) to be very plausible, which means that they effectively challenge the presumption that the SEU axioms of preference define rational choice. But then we are left with a quandary—it is not clear how we should decide what *are* the minimal requirements for rational choice. People's intuitions clearly differ when it comes to what constitutes good decision-making, both in the field and in the lab, so to speak. Thus it looks like we have reached a standoff. There is a way forward in this debate, however, and that is to seek out more consequences associated with upholding/violating particular axioms of preference. It turns out that the sequential choice context is particularly fruitful in this respect. Indeed, this is my focus in this final chapter—in the tradition of Hammond (1976, 1977, 1988b, 1988c), McClennen (1990) and Seidenfeld (1988a), I look at what, if anything, an analysis of sequential decision-making tells us about the relative plausibility of choice rules.

The two core axioms of SEU theory—independence and ordering—are best investigated side-by-side in the sequential-choice context. It might be noted that I

have said more about the independence axiom in the previous couple of chapters, as compared to the ordering axiom. But both axioms are integral to SEU theory, and so any violation of ordering should be just as much cause for concern. Allow me to clarify here: it may not be such a big deal to relax just the completeness component of ordering, thus permitting an agent to have no preference at all between two particular options. But Levi's choice rule for handling indeterminate belief/preference, for instance, violates more than completeness, even without the secondary security considerations. Levi's choice rule also violates transitivity. I did not draw much attention to this fact in the previous chapter; I was more interested in whether the secondary security considerations involve a relaxation of independence. But there is good reason to worry about relaxing transitivity; indeed, if there is one axiom that seems like a self-evident constraint on rational preferences, surely it is this one. Things are not so straightforward, however. There are different shades to transitivity, and in the next section (6.2), I will present the axiom in a way that draws attention to this fact, and shows some violations of transitivity to be more defensible than others. We will also see that there are some significant similarities between ordering and independence, which might lead us to question whether either axiom can be singled out as the more inviolable preference axiom.

After considering the similarities between independence and ordering, I go on to analyse two key sequential-choice arguments concerning these axioms. Section 6.3 deals with Hammond's "consequentialist" argument, which is intended to be a defence of SEU theory in its entirety. While I do not agree with all of Hammond's assumptions, his approach provides inspiration for an alternative "diachronic-Dutch-book-style argument" against decision theories that relax the independence axiom. I formulate and assess such an argument in Section 6.4. I claim that, while diachronic-Dutch-book-style arguments have much merit, they are nonetheless open to challenge. In Sections 6.5–6.8, I go on to consider Seidenfeld's (1988a) rather involved argument that discriminates between different types of modifications to SEU theory. Seidenfeld claims that decision theories that relax the independence axiom have inconsistencies in the sequential-choice context that are not suffered by decision theories that relax ordering. I draw some conclusions about these various arguments, and thus what the sequential-choice setting contributes to the study of rational choice,

in Section 6.9.

6.2 Recasting ordering and independence

It is important to first recast the ordering axiom in a somewhat different light so that we can better understand its properties, including how it is related to independence. My examination of ordering will hinge on the relationship between an agent's choice rule/function and their preference ranking over options. (I draw heavily on the work of McClennen (1990, chaps. 2–3) in this section.) In the previous chapter (in Section 5.7), I argued that the two are effectively equivalent—an agent's choice rule should reflect their *all-things-considered* preferences, and conversely, we should be able to construct an agent's preferences from their choice rule. These representations nonetheless offer differing perspectives on rational choice. For instance, it couldn't be denied that the numerical expected utility representation sheds further light on the character of a preference ordering that obeys the SEU axioms. Importantly, we can also gain a better understanding of axioms of preference by translating them into constraints on a choice function. In fact, it is not immediately obvious how this should be done, and it has been the subject of some study (for a comprehensive treatment, see Sen 1979, chap. 1).

Let us take a moment to introduce some choice function terminology. I will refer to the options that are selected by a choice function C(.) as the "admissible set". For example, let's say we have a set S of options, then the admissible set C(S) is the subset of options in S that are not bettered by some other option in S.[122] (It is possible that C(S) = S, meaning that none of the options in the set is bettered by any other.) Consider then how we can construct an agent's preference ranking from their choice

[122] McClennen (1990, p. 35) follows Sen (1979, pp. 9–10) in referring to this set as the "maximal set", which can be contrasted with the "choice/optimal set". A choice/optimal set must be comprised of options that the agent is indifferent between, whereas a maximal set can include incommensurable options (those options that cannot be bettered by any others.) The "maximal set" is the more general notion; I refer to it simply as the "admissible set".

function. If we have a set S that comprises of only two options A and B, such that $S = \{A, B\}$, then $A \succ B$ just in case $C(S) = \{A\}$, $B \succ A$ just in case $C(S) = \{B\}$, and A is indifferent to B, or else the two are incommensurable, in the case that $C(S) = \{A, B\}$. This rationale can be extended to all pairs of options in some larger set X over which C(.) is defined, to determine the agent's entire preference ranking over the options in X.

Recall that my particular interest is what the ordering axiom looks like when it is construed in terms of constraints on a choice function. Sen (1979, chap. 1) provides the answer to this question; he shows that a choice function defined over some set X will yield preferences satisfying ordering if it obeys two constraints (referred to as a*lpha* and *beta*) with respect to every subset S in X. The constraints can be formalised as follows[123]:

> *alpha*: A choice function C(.) defined on X satisfies *alpha* just in case for all A_i in S, and all S^* such that S^* is a superset of S, if A_i is not in $C(S)$, then A_i is not in $C(S^*)$.

> *beta*: A choice function C(.) defined on X satisfies *beta* just in case for all A_i and A_j in S, and all S^* such that S^* is a superset of S, if both A_i and A_j are in $C(S)$, then either A_i and A_j are both in $C(S^*)$ or neither is in $C(S^*)$.

Transitivity is the core of the ordering axiom, so we could say that *alpha* and *beta* stipulate that an agent's preference ordering (over X) be transitive. It is worth noting that Levi's (1986) choice rule (which I discussed in the previous chapter) violates the *beta* condition, even if the agent is security-neutral. (According to Levi's rule, two options in a set might *both* maximise expected utility for some permissible probability-utility pair, but if we augment the set, it is possible that only one of these options continues to maximise expected utility for some permissible probability-

[123] I take this presentation of the *alpha* and *beta* constraints from McClennen (1990, p. 23).

utility pair.)

McClennen (1990, p. 23) refers to *alpha* and *beta* as "context-free" conditions on choice functions, since what they require is that

> …the choice-worthiness of various alternatives not be influenced in certain specific ways by the presence or absence of other alternatives. More specifically, what they require is that if an alternative is not choice-worthy, adding new options should not result in that option being transformed into one that is choice-worthy; and if two options are both choice-worthy, adding new options should not result in one being choice-worthy and one not.

This is certainly a different way of looking at the transitivity preference axiom. Transitivity seems like an inviolable constraint when we focus on an agent's preference ordering over an entire set *X* of options. But we might instead focus on the agent's choice function, and the options it selects from subsets of *X*. I think there is scope to question whether a choice function should obey the *alpha* and *beta* conditions outlined above. We might argue that context does matter. As per Levi's rule, two options may be choice-worthy in one scenario, but in the presence of further options one of these options may drop out of the admissible set. In other words, the preference relation between two options may change, depending on what other options are available. McClennen points out that such a *beta*-violating choice function may even be based on well-ordered preferences in any particular context or option set *S*; it is just that when we try to amalgamate the choice results from all the individual subsets of some larger set *X* over which the choice function is defined, we get an overall intransitive preference ordering. (This is indeed the case for Levi's choice rule.) In my opinion, a transitivity violation of this kind does not seem unreasonable. This leads me to conclude that ordering, like independence, will require some further defence if it is going to be a convincing constraint on rational choice.

It is also noteworthy that when ordering/transitivity is described in terms of constraints on a choice function, it looks to have more in common with the

independence axiom that what might ordinarily appear to be the case. In particular, there is a "context-free" aspect to both axioms. We have just seen that ordering requires the preference relationship between options to not depend on what other options are available to the agent. Recall (from Chapter 4) that independence also imposes a kind of context-freedom; in this case it pertains to the outcomes for an individual act. More specifically, independence requires that the preference relationship between options not depend on the nature of any outcomes that they have in common. This effectively means that the contribution an individual outcome makes to the value of an act should be independent of what other possible outcomes the act might yield.

Given that independence and ordering both stipulate a kind of context-freedom, it might be thought that violation of either axiom will have some similar consequences. In fact, Hammond shows this to be the case when it comes to dynamic or sequential decision-making, and he mounts a defence of SEU theory that turns on the similarities between the theory's core preference axioms. Ordering and independence stand together when it comes to Hammond's "consequentialist" argument—he shows that an agent who is concerned to maximise with respect to "consequences" in the sequential-choice context must subscribe to both these axioms. In the next section I will analyse Hammond's argument, paying particular attention to his definition of "consequentialism", and why it might be considered a requirement of rationality. The question is whether Hammond's argument provides the necessary extra defence of the SEU axioms, or whether there is yet scope to challenge the independence and ordering constraints on rational choice.

6.3 Hammond's argument

In Chapter 1, I defended the "sophisticated" approach to sequential choice, and I contrasted this approach with that of Hammond, which others have referred to as "naïve" or "myopic". Naïve/myopic choice is so-named because the agent is expected

to evaluate strategies from the point of view of their current beliefs and values, regardless of whether their future self can be counted on to carry out any such plans. I argued in Chapter 1 that Hammond's approach has some serious flaws if it is taken to be a proposal for evaluating any sequential-choice problem that an agent might possibly face. But I also pointed out that Hammond is best understood as responding to a different sort of issue. We might say that Hammond is concerned to model the temporally-extended ideal agent; he considers what are the characteristics of good sequential decision-making under circumstances in which the agent expects their future self to act precisely as planned. So, for instance, it is not that Hammond would recommend Ulysses take no precaution against the future perils associated with the island of the sirens, it is just that Hammond is not interested in decision scenarios like the one Ulysses faces. Hammond is not trying to provide an exhaustive decision-making framework; he is interested, rather, in what are rational decision-making plans, and what characteristics a sequential-choice approach should have, just in those cases where everything is expected to go according to plan.

Of particular interest is Hammond's stipulation that normal-form and extensive-form decision solutions should be identical. McClennen (1990, p. 115) refers to this condition as "normal-form/extensive-form coincidence" (NEC). It is the major underlying assumption behind the naïve approach to strategy assessment—naïve choice is simply inconsistent unless NEC holds. I argued in Chapter 1 that it is the NEC condition that is particularly dangerous if we seek a fully general decision-making approach—there is arguably always the possibility that an agent will have an uncontrollable change in belief or preference. In such case, only sophisticated choice is guaranteed to give consistent strategy advice (i.e. only sophisticated choice recommends strategies that the agent predicts they will be able to carry out), and NEC will not necessarily hold.[124] In fact, I argued in Chapter 1 that normal-form/extensive-form coincidence should only come into play after, as opposed to before, we determine the correct approach to sequential choice: static models should simply be brought into line with the sophisticated sequential analysis of a decision problem.

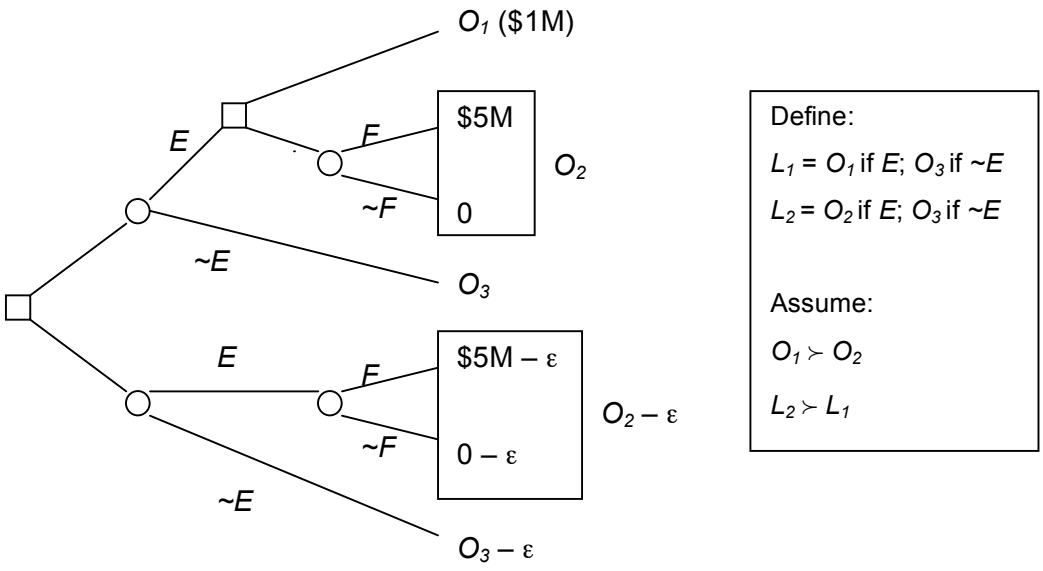
[124] Some defenders of "resolute" choice might argue that the agent can exercise resolute choice in such circumstances (and also uphold the NEC condition), but I have argued in Chapter 1 that the resolute approach is not plausible.

Consider again the problem of Ulysses and the sirens. There need not be any difference between the static and sequential representations of Ulysses' decision problem. We can simply specify that the strategy whereby Ulysses sails home directly to Ithaca, past the island of the sirens and without being tied to the mast, is not an available option, whether in the dynamic or the static context. The said strategy is shown to be infeasible when we consider the dynamics of Ulysses' problem, and so it should not be included as an option in the corresponding static decision model either.

The NEC condition may be important in another sense, however—perhaps a good choice rule should satisfy the condition just in those idealised cases in which the agent's beliefs and preferences *are* expected to remain stable through time. (Here we are assuming that the static-form available strategies are all the combinations of choices at choice nodes.) I will refer to this version of the condition as "idealised NEC". This is effectively the approach that Hammond takes towards the requirement that normal- and extensive-form decision solutions coincide. And Hammond (1976, 1977, 1988b, 1988c) proves that idealised NEC is only satisfied by choice rules that uphold both the independence and ordering axioms—i.e. SEU theory alone. We could say that it is the "context-free" aspect of both ordering and independence that brings them together when it comes to the idealised NEC condition. I will not present Hammond's proofs here. Instead, I will merely illustrate his result, by showing that a theory relaxing independence does not satisfy idealised NEC (Figure 6-1) and a theory relaxing ordering does not uphold the condition either (Figure 6-2).

Figure 6-1 depicts the same decision problem that I presented in the first chapter when discussing McClennen's concerns about sophisticated choice when it is coupled with an independence-violating choice rule. Note that the agent's preferences are given at the right of the diagram, and they show a violation of independence.

Figure 6-1

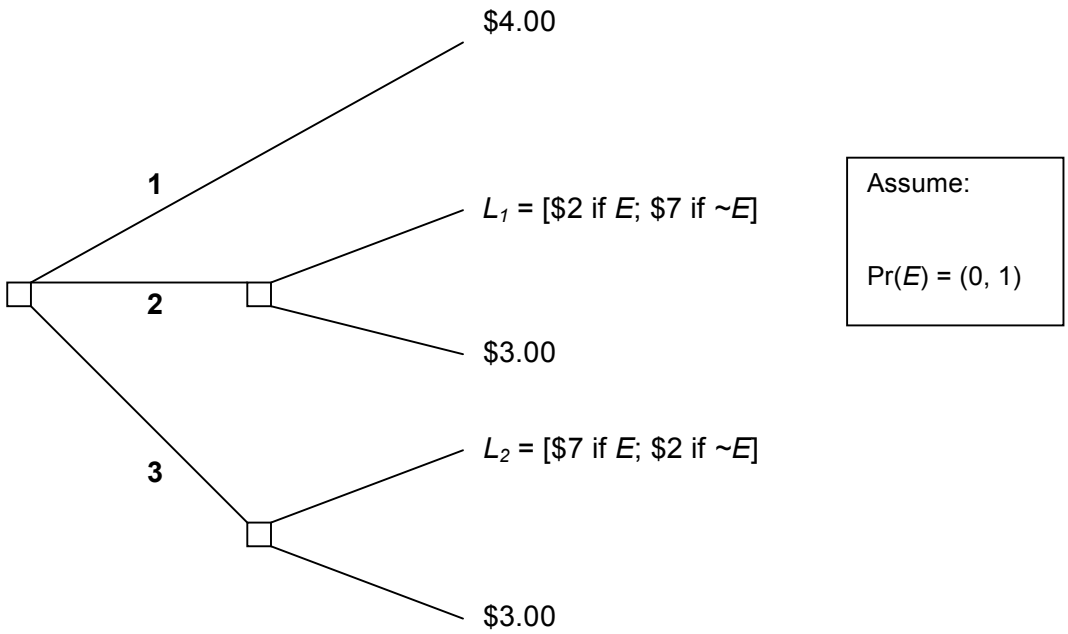


Given the specified preferences, the sophisticated agent will choose "down" at the initial choice node, which amounts to the lottery (L_2 - ε) (for some positive ε). Idealised NEC is not satisfied, because there is another strategy mapped out by this decision tree ("up" at the initial choice node, and then "down") that the agent would choose were it available to them. The fact that in this case the unavailable option strictly dominates the option that the sophisticated agent must pursue makes the case for idealised NEC more striking. Our independence-violating agent is guaranteed to get less than they might have because their very choice function places the more profitable option off-limits.

There is a similar story for decision theories that relax the ordering axiom of SEU theory. Refer to Figure 6-2, which depicts a decision problem confronting an agent who has an indeterminate belief in the proposition *E*. (We will assume that for this agent the utility of money is linear, and the agent distinguishes between incommensurable options on the basis of security considerations.[125])

[125] I argued in the previous chapter that such a choice function also violates the independence axiom.

Figure 6-2



The optimal strategy for the agent is Plan 1, because the other strategies will inevitably lead to a $3 prize (given that $3.00 is preferred to either L_1 or L_2). But from the perspective of the initial choice node, the only admissible options are in fact L_1 and L_2, because only these options have maximum expected utility for at least one permissible probability-utility pair. But these strategies are not possible. So again we have the situation where the most profitable strategies from the perspective of the initial choice node are not available to the agent, given their expected future choices.

According to Hammond (1976, 1977, 1988b, 1988c), it is simply a natural requirement on a choice function that it yield equivalent solutions whether a decision is analysed in normal or sequential form. Hammond claims that this constraint follows from a commitment to consequentialism, but he defines "consequentialism" in a very particular way. It is hard to dispute the following characterisation: that "acts be valued by their consequences" (Hammond 1988b, p. 25). But Hammond reads a lot into this statement when it comes to examining sequential decisions. He goes on to say, "consequentialist behaviour, by definition, reveals a consequence choice function independent of the structure of the decision tree." (Hammond 1988b, p. 25) In other

words, the value (and for Hammond, the admissibility) of an act should not depend on the particular decision tree in which it features; the same probabilistic distribution of terminal outcomes produced by any tree within the space of possible decision trees is effectively the same act, and should thus receive the same evaluation.

Idealised NEC might yet be a plausible criterion for choice rules, but it is not clear that it follows from a commitment to acts being valued by their consequences, as Hammond argues. Even if sequential and static decision solutions differ, we can still stipulate that the value of plans depend solely on the probability distribution of terminal outcomes, as the so-called "consequentialist" requires.[126] (Note that there is the further question of what is the proper content of outcomes, which I addressed in Chapter 4.) It may simply be the case that some normal-form plans turn out to be unfeasible in a particular sequential decision scenario, and so these plans shouldn't even be up for consideration. In this way the sequential tree may matter because it restricts the space of possible plans, not because it affects the subsequent evaluation of these plans. This is surely compatible with the sort of "consequentialism" that Hammond has in mind. Acts that have the same probabilistic distribution of outcomes will always have the same value, regardless of what particular path through a sequential decision tree they describe. It is just that the *availability* of acts varies between sequential-choice settings. Thus I think Hammond moves too quickly from his "consequentialist" assumptions to the conclusion that static and sequential decision solutions should be equivalent (in ideal circumstances).

While I disagree with Hammond that his position is just the result of a commitment to "consequentialism", we cannot ignore the fact that there is something to be said for a choice function that allows you to pursue the strategy you initially deem most attractive. If a particular combination of choices at nodes turns out to be the "best" plan, then it might seem that we would want our choice function to make this plan possible. The argument might go something like this: if your choice function identifies some combination of choices at nodes as optimal, then it should be the case

[126] "Consequentialism" need not mean the same thing here as it does in discussions of ethical theories.

that this same choice function makes the proposed plan possible, when it comes to the actual dynamics of the problem. Otherwise, it would seem that the choice function is in a sense self-defeating—it does not make possible the very plan that it deems optimal from the perspective of the initial node.

I do not think this argument for idealised NEC is sufficiently compelling, however. For starters, what are considered the "best" plans depends on the choice function, so consistency between "best" plans at the initial node and "best" plans at later nodes is only desirable if the choice function does indeed have merit. In other words, we might question whether satisfaction of idealised NEC really shows something positive about a choice function. One might insist that we have here a kind of internal consistency requirement. Of course, the choice function will need to be assessed in other ways, but it should at least be such that it is "self-reinforcing" in the sense just alluded to. But again, I think it is misleading to call the idealised NEC requirement a matter of internal consistency. It is not clear that there is anything self-refuting about a choice function that can lead to cases where some combination of choices at choice nodes must be disregarded because sequential analysis reveals the act in question to be impossible. In these cases the agent knows that at later choice nodes they will be equipped with more information upon which to make their choices, and a strategy that seemed competitive from the perspective of the initial choice node will not in fact be pursued.

6.4 A diachronic-Dutch-book-style argument

There may yet be circumstances in which a difference between the static- and extensive-form decision solutions seems troubling. My reasons above for rejecting idealised NEC as a condition on choice functions look more convincing when it comes to the example in Figure 6-2, as compared to the example in Figure 6-1. In the latter case, we can see that the agent stands to lose a *sure* amount on account of their chosen decision rule. The sure loss here is very significant; it makes it harder to

defend the agent's decision-making plans. It is not just that the agent in Figure 6-1 subscribes to a choice rule that could be described as self-refuting, as discussed above. Here we might go further and say that the agent's plans are shown to be *objectively* inconsistent, since they lead the agent to a strategy that is dominated by another (albeit unavailable) strategy. This provides a much stronger case for arguing that the agent in Figure 6-1 would be better off subscribing to a different choice function—to enable them to have access to a clearly better strategy.

So I do not think idealised NEC *per se* is a compelling constraint on decision rules. Rather, it is the possibility of suffering sure loss due to one's own decision rule that, arguably, demonstrates an inconsistency in that rule. It might be noted that this criterion for decision rules has much in common with the diachronic Dutch book argument (DBA) for conditionalisation as the rule for updating beliefs. The diachronic DBA, or the version of the argument that I think is plausible, at least, rests on the premise that it is irrational to have belief-updating plans that lead one to select a strategy that is dominated by another (albeit unavailable) strategy. (Recall my discussion of the diachronic DBA in Chapter 2.[127]) The general argument then is that an agent's decision-making plans, whether their belief-updating plans or their intended decision rule, should not prevent them from pursuing sure gains. I will refer to any such argument as a "diachronic-Dutch-book-style" argument. Shortly, I will present a general challenge to these arguments, but for the moment, let us consider what sort of decision rules are vulnerable to "Dutch-book sure losses".

The example in Figure 6-1 suggests that any decision theory relaxing the independence axiom will entail Dutch-book sure losses in at least some decision circumstances. In fact, we can easily generalise the decision problem in Figure 6-1 to show that this is indeed the case. In that particular example, specific monetary amounts were nominated to correspond to the outcomes O_1, O_2 and O_3 (and note that O_2 is a lottery or a non-basic outcome[128]). But O_1, O_2 and O_3 could be any three basic

[127] In fact, we can also liken Hammond's "consequentialist" argument to the naïve version of the diachronic DBA, which I argued to be unsuccessful.

[128] I am using the term "lottery" loosely here to refer to a probabilistic outcome, or in other

outcomes/lotteries, involving any sort of goods. The diachronic-Dutch-book-style argument will hold so long as there is some set of three basic outcomes/lotteries—O_1, O_2 and O_3—such that $O_1 \succ O_2$ but $(O_2 \text{ if } E;\ O_3 \text{ if } \sim E) \succ (O_1 \text{ if } E;\ O_3 \text{ if } \sim E)$. This will be true of any theory that relaxes the independence axiom.

While Figure 6-2 does not make this evident, Levi's choice rule (with secondary security considerations) can lead to Dutch-book sure losses. This is perhaps not surprising given that Levi's lexical choice rule violates independence, as argued in the previous chapter. For instance, consider the sequential decision problem in Figure 6-3 below. The parameters of the problem are inspired by Ellsberg's problem. Assume that the utility of money is linear. If the security-conscious agent were to find him/herself at the second choice node, then they would presumably choose "up", because the expected utility of this option is known to be 1/3 × U($100), whereas the other option has an interval expected utility with a lower worst-case value.[129] From the point of view of the initial node, then, there are only two strategies on offer: "up" then "up" again, or "down". Both options are admissible, but the security-conscious agent will choose the latter option, because it has an expected utility of 2/3 × U($100 – ε), while the other option has an interval expected utility with a lower worst-case value. (The value of ε should be sufficiently small so that this is indeed the case.) Importantly, the chosen option is strictly dominated by the strategy that is unavailable to the agent due to their own decision rule: the strategy that amounts to "up" then "down". Levi's security-conscious agent is here shown to suffer from a Dutch-book sure loss.[130]

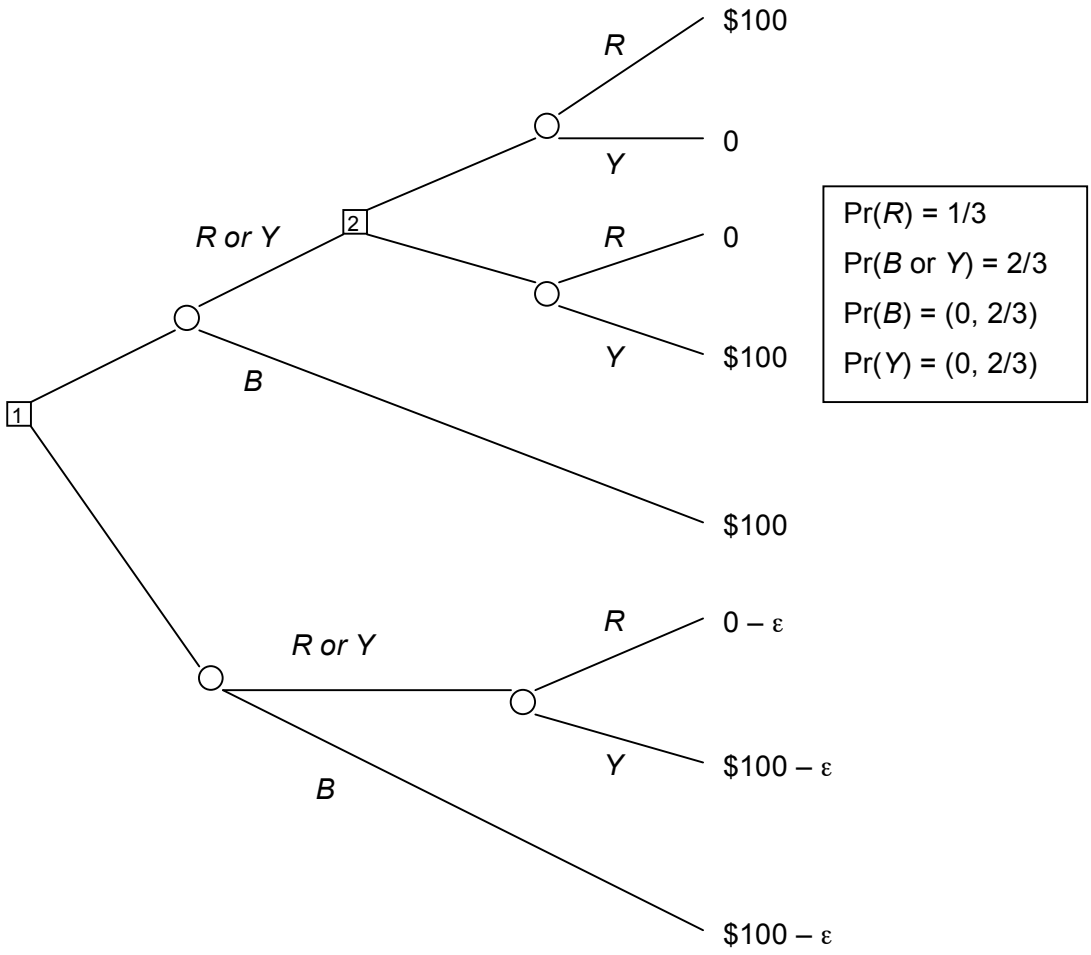
---

words, an outcome that can be decomposed into a probability distribution over more basic outcomes. The probabilities here need not be objective (contrary to the common usage of the term "lottery" in the decision theory literature.)

[129] This will presumably hold, even if we allow for the fact that the agent will update their probabilities after learning $\sim B$.

[130] Seidenfeld (1994, p. 459) presents another example in which Levi's lexical decision rule (with secondary security considerations) leads to a situation in which the chosen option is dominated by an unavailable or unfeasible option.

Figure 6-3



It seems clear then that the diachronic-Dutch-book-style argument rules out all decision theories that relax the independence axiom of preference, whether this be a "primary" or a "secondary" violation of independence (to use Levi's terminology).[131] On the other hand, I predict that decision theories relaxing *just* the ordering axiom will not, in general, be vulnerable to any Dutch-book sure losses. That would mean that the diachronic-Dutch-book-style argument does not rule out the security-neutral version of Levi's decision rule. I will not explore this possibility in any detail here, however. It would certainly be interesting if we could establish that decision theories relaxing ordering *alone* are not excluded by a diachronic-Dutch-book-style argument.

---

[131] I argued in Chapter 5 that Levi's lexical decision rule violates independence. It might still be useful to refer to this as a "secondary" violation of independence, meaning that the axiom is violated at the secondary stage of the decision rule, when security considerations enter in.

But in my opinion, it is not very useful to incorporate indeterminacy of belief/desire in a decision model, unless we allow secondary security considerations to play a role. Recall from the previous chapter, for instance, that the very plausible-seeming "Ellsberg-choices" cannot be rationalised by a decision theory that relaxes ordering alone;[132] the agent must be permitted to discriminate between "admissible" options on the basis of security. It could be said, then, that if we want to properly accommodate indeterminacy of belief/preference, we must challenge the diachronic-Dutch-book-style arguments.

In Chapter 2, I contended that the diachronic DBA for belief-update via conditionalisation, properly formulated, is a very plausible argument. But I also pointed out that the diachronic DBA is nonetheless open to challenge. Given that a diachronic-Dutch-book-style argument excludes what I consider to be the most defensible decision rule for handling indeterminacy, I will here take up that challenge. Recall my earlier point in Chapter 2 that the sure losses featured in the diachronic DBA are, in a sense, not genuine losses at all. It is not the case that the agent has several options available to them, and if they have faulty decision-making plans, they may end up choosing a strategy that is dominated by another *available* strategy. In the context of the diachronic-Dutch-book-style arguments, the dominating strategy against which we measure the agent's sure loss is not in fact a "dynamically feasible" option—the agent predicts that they would not make the requisite series of choices at the given choice nodes. It could well be argued that unavailable or unfeasible options are irrelevant, meaning that such options should not enter into any analysis of the decision problem and its solutions.[133] This is to say that we should concern ourselves solely with "dynamically feasible" options, whether we are using a decision rule to select the admissible options, or subsequently assessing the merits of that decision rule.

For the reasons just given, I do not think a decision rule that leads to Dutch-book sure

---

[132] More precisely, the Ellsberg choices cannot be rationalised by a relaxation of ordering alone unless we modify the description of the outcomes to include risk/regret emotions.

[133] This is the position that Seidenfeld (1988 & 1994) takes.

losses can be described as "inconsistent". On the other hand, I do not think it is quite right to say that dynamically unfeasible options are irrelevant to the assessment of decision plans. After all, in the diachronic-Dutch-book setting, it is the agent's very belief-updating plan or decision rule that makes the more profitable options unavailable to them. Surely this is a mark against any such decision plans, even if it does not completely rule them out as being irrational. We might conclude that, all other things being equal, it is preferable that a decision rule does not lead to Dutch-book sure losses. Note that this also commits us to the view that, while the diachronic DBA adjudicates in favour of conditionalisation as the rule for updating beliefs, it is no watertight defence of this rule.

There are others who have expressed concern about Dutch-book sure losses (to use my terminology; others do not speak of a generalised diachronic DBA), but who are unwilling to cast decision theories that are vulnerable to such losses as "irrational". Apart from Seidenfeld (1994), Rabinowicz (1995), for instance, has acknowledged the sure-loss problem described above for theories that relax the independence axiom, but he does not think it provides sufficient grounds for rejecting such theories. Granted this position—diachronic-Dutch-book-style arguments are somewhat inadequate—it does not look like any of the variant decision theories that I have entertained in the second part of this thesis can be discarded. At this stage, we cannot rule out decision theories that relax independence, or ordering, or both of these axioms. But that is not to say that the justificatory story with respect to preference axioms/choice-function constraints ends here. It remains to consider whether the sequential-choice framework yields any further arguments for/against particular kinds of decision rules.

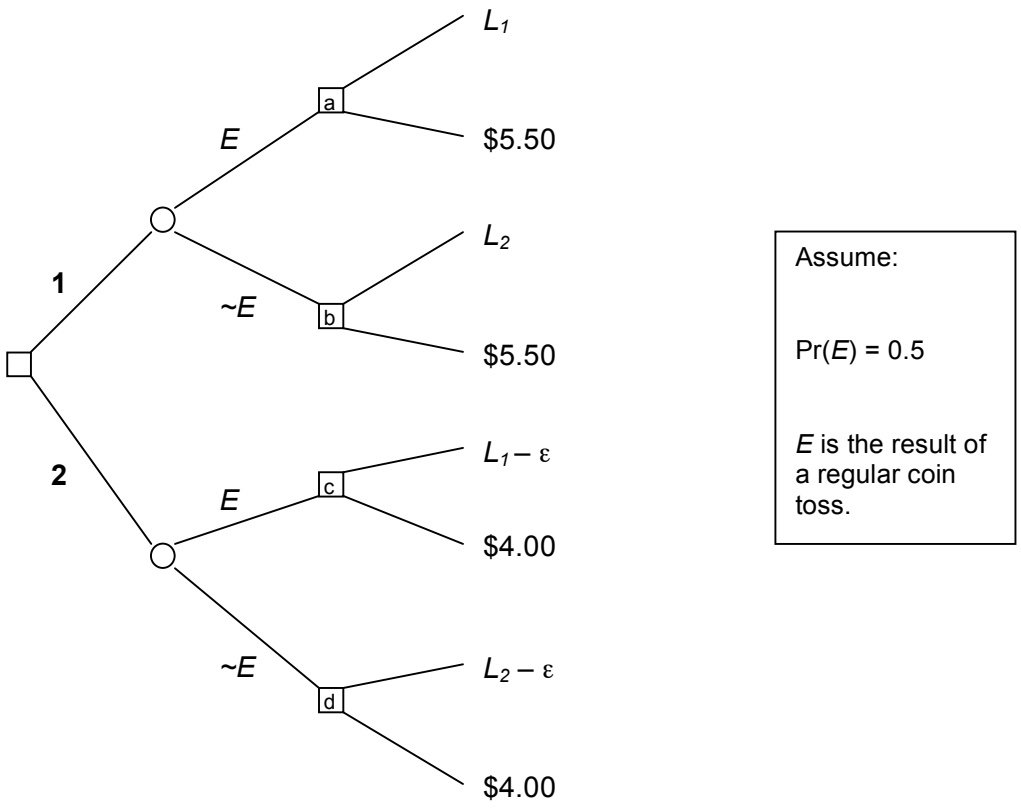
6.5 Seidenfeld's argument against relaxing independence

I turn now to Seidenfeld's (1988a) argument concerning decision rules in the sequential-choice context. Seidenfeld argues that there is an important distinction

between Levi's decision theory, for instance, and decision theories like cumulative prospect theory that relax just the independence axiom of preference. Only the latter kind of decision rule, he claims, leads to inconsistencies in the sequential-choice context. I will examine Seidenfeld's argument in detail, because it has a number of subtleties that must be spelled out carefully. Furthermore, the argument is not uncontroversial—it has already provoked some criticism, which I will draw attention to in the following sections. What is at issue is whether Seidenfeld provides any new argument against decision theories that relax just the independence axiom, and whether he is right to say that Levi's decision theory, for instance, is not vulnerable to the same problems.

It is useful to begin with the example decision problem Seidenfeld uses to illustrate his argument. Seidenfeld (1988a, pp. 281–283) goes on to give a generalised version of his result, but it is perhaps easier to analyse the specific example. The decision set-up is as per Figure 6-4 below:

Figure 6-4



Our agent has preferences that violate independence, but we will assume that she obeys ordering and first-order stochastic dominance.[134] The relevant part of the agent's preference ordering is as follows (with outcomes higher up the list preferred over those lower down the list and with indifferent options on the same line, separated by a comma)[135]:

$\$6.00, 0.5 \times L_1 + 0.5 \times L_2$

$\$5.75, 0.5 \times (L_1 - \varepsilon) + 0.5 \times (L_2 - \varepsilon)$

$\$5.00, L_1, L_2$

$(L_1 - \varepsilon), (L_2 - \varepsilon)$

$\$4.00$

Being a sophisticated chooser, the agent works backwards from the final choice nodes in order to determine and evaluate the possible decision strategies. So if she finds herself at either choice node (a) or (b), she recognises that in each case she will choose the sure \$5.50 over the lottery (whether it be L_1 or L_2) because the lotteries are worth only \$5.00. If, on the other hand, the agent were to find herself at choice node (c) or (d), she would in each case choose the lottery over the sure \$4.00. That is, she would choose $(L_1 - \varepsilon)$ at (c) and $(L_2 - \varepsilon)$ at (d).

We can now evaluate the two available plans (which I will refer to as "Plan 1a" and "Plan 2a"), bearing in mind that the result of the coin toss is not known at the initial node:

---

[134] I gave formal definitions of both of these axioms of preference in Chapter 4.

[135] Seidenfeld includes lotteries $(L_1, L_2, L_1 - \varepsilon, L_2 - \varepsilon)$ as outcomes. In the tradition of Anscombe and Aumann (1963), Seidenfeld is assuming that there are some objective probabilities that can be used to specify lotteries. The probability distribution over states in the decision problem is still subjective. Note, however, that Seidenfeld's argument can easily be redescribed without the use of lotteries that depend on objective probabilities.

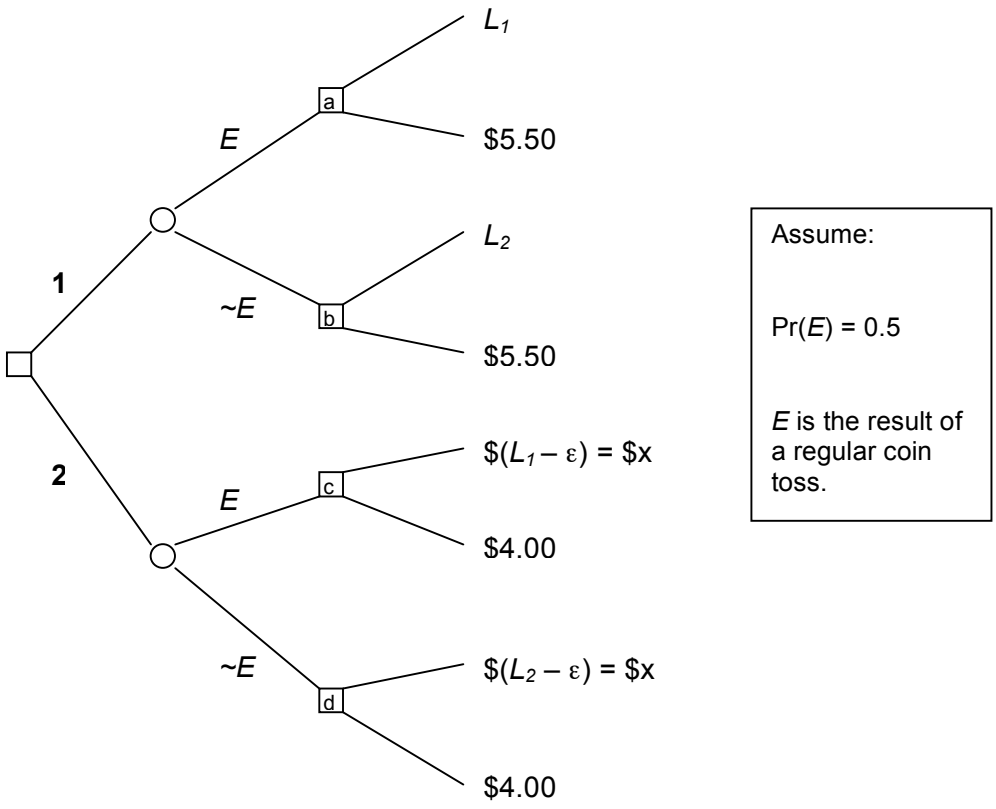
Plan 1a: $0.5 \times \$5.50 + 0.5 \times \$5.50 = \$5.50$

Plan 2a: $0.5 \times (L_1 - \varepsilon) + 0.5 \times (L_2 - \varepsilon) = \$5.75$

So the agent will choose Plan 2a over Plan 1a. But note that regardless of whether the coin shows up heads or tails in the case of Plan 2a, the agent will be left with a prize worth less than the $5.50 she could have received if she had followed Plan 1a. Note the peculiarity of this state of affairs—the agent prefers Plan 2a to Plan 1a when she can *foresee* that at the later choice node she will wish she had rather chosen Plan 1a. Furthermore, the agent suffers from a Dutch-book sure loss, as discussed in the previous section: there is another strategy which dominates the chosen Plan 2a, but this strategy is not available to the agent on account of their choice function. (The strategy in question is the one that equates to the lottery $0.5 \times L_1 + 0.5 \times L_2$. We might refer to this strategy as Plan 1b.)

As stated in the previous section, Seidenfeld does not think the above arguments against the agent's decision rule are decisive, and I agree. Seidenfeld pursues a different argument; he seeks to show that decision rules relaxing just the independence axiom are inconsistent in the sequential-choice context in a manner that does not apply to Levi's decision rule. To this end, Seidenfeld asks us to compare the decision problem in Figure 6-4 above with the following scenario that involves the lotteries $(L_1 - \varepsilon)$ and $(L_2 - \varepsilon)$ being swapped for their sure-dollar equivalents. Recall that $\$4.00 \prec \$(L_1 - \varepsilon) \approx \$(L_2 - \varepsilon) \approx \$x \prec \$5.00$. The revised problem is illustrated in Figure 6-5 below:

Figure 6-5



In this case, the plans will be evaluated as follows:

Plan 1a: $5.50 (as before)

Plan 2a: $0.5 \times \$x + 0.5 \times \$x = \$x$

Here Plan 1a will be chosen over Plan 2a, but the only change to the original decision problem is that some terminal lotteries were exchanged for their sure-dollar equivalents, or in other words, terminal outcomes were exchanged for "indifferents". Seidenfeld asserts that what we see here is a violation of "stochastic dominance in the sequential setting", or what we might call "dynamic SD". He claims that if, as in the second scenario above, Plan 1a stochastically dominates Plan 2a, then any plan that is effectively identical to Plan 2a, (the only difference being that at one or more choice points the continuation tree is exchanged for one that the agent is indifferent between), should also be less preferred than Plan 1a. Below is Seidenfeld's (1988a, p.

275) formal statement of this condition:

> *Dynamic SD: A decision rule is coherent if (i), admissible choices under the rule are stochastically un-dominated, and (ii), admissibility is preserved under substitution (at choice points) of "indifferent" options.*

Clearly our agent above does not meet this condition, and in fact Seidenfeld proves the general result—that no choice rule relaxing independence alone satisfies his dynamic SD condition.

6.6 Begging the question against independence-violators?

In his comments on Seidenfeld's (1988a) paper, McClennen (1988) argues that the dynamic SD condition essentially begs the question against independence-violators, and that in any case agents who do not satisfy ordering can be shown to be dynamically inconsistent according to a slightly modified dynamic SD condition. I will here argue in favour of McClennen's position. What McClennen successfully draws attention to, I think, is that the dynamic SD condition as it is stated above is unlikely to be a convincing rationality constraint in the eyes of someone who is already prepared to relax independence. (Just as a condition that depends in some uninteresting way on well-ordered preferences is unlikely to impress someone who wants to give up ordering.) This does not mean that there is nothing further that can be said to boost the case for dynamic SD. Seidenfeld (1988b, 2000a & 2000b) in fact pursues this path; he argues that a sophisticated chooser who violates independence, obeys ordering and wants the possibility of having stable preferences throughout a sequential decision, must recognise that this entails compliance with the dynamic SD condition. I examine this argument closely in Section 6.8. But first let me consider why an argument of this sort is needed; there are reasons to be unimpressed with the dynamic SD condition just as it is stands.

Consider again the revised example problem, as illustrated in Figure 6-5. It is clear that Plan 1a stochastically dominates Plan 2a, because whichever way the world turns out (whether the coin lands heads or tails), Plan 1a yields a basic outcome that is more desirable than the basic outcome received via Plan 2a. This is illustrated by the first two rows of the table in Figure 6-6 below:

Figure 6-6

	Heads up	Tails up
Plan 1a	$5.50	$5.50
Plan 2a	$4.00 < $x < $5.00	$4.00 < $x < $5.00
Plan 2a´	$L_1 - \varepsilon$	$L_2 - \varepsilon$

But it is unclear why exchanging the terminal outcomes in Plan 2a with "indifferent" lotteries ($L_1 - \varepsilon$) and ($L_2 - \varepsilon$), (so that we now have Plan 2a′ in the above table and are back to the original decision problem described in Figure 6-4), should preserve the preference relation between the plans. Indeed, the effect on preferences when changes to an act involve indifferent sub-lotteries as opposed to basic outcomes is precisely what is not stipulated by *first-order* stochastic dominance. (We can see from the table in Figure 6-6 that Plan 2a′ differs from Plan 2a by an exchange of indifferent sub-lotteries, as opposed to an exchange of indifferent basic outcomes.) It is the stronger independence axiom that requires preferences remain unchanged when exchanges of "indifferents" in acts involve sub-lotteries.

In a similar vein, McClennen notes that an agent who violates (just) independence will necessarily fail Seidenfeld's dynamic SD condition, and not by a series of unobvious steps but rather for very straightforward and uninteresting reasons. Independence-violators *by definition* relax the stipulation that preferences between plans remain unchanged when there is an exchange of indifferent sub-lotteries at chance nodes. As McClennen notes, the exchange of indifferents at *choice* nodes in a

dynamic decision scenario is a slightly different issue, but the two kinds of substitution are surely very much related. Indeed, given the assumption of sophisticated choice, any dynamic decision tree can be redescribed without the choice nodes (because the sophisticated chooser knows what they will choose at future nodes anyway). And in this way, every violation of Seidenfeld's dynamic SD condition can be represented as a violation of the original substitution of lotteries condition that is rejected by independence-violators. So as McClennen points out, reason to reject independence looks like reason enough to reject Seidenfeld's dynamic SD condition. In other words, if we think that there is good motivation for relaxing the independence axiom, then we should be unmoved by the fact that this means violating dynamic SD, because the latter condition looks very much like a mere restatement of independence.

McClennen (1988, p. 307) goes on to suggest a plausible variation of Seidenfeld's dynamic SD condition that is intended to catch-out those who violate ordering. The condition is as follows:

> *Dynamic Substitution with "Improved" Options (DYN-SUB*): In a dynamic choice context, if a particular plan is acceptable, and an agent substitutes for one of its component prospects another prospect judged even more acceptable (in pair-wise comparison), then he or she should judge the modified plan at least as acceptable as the original, unmodified one.*

The following figures (Figure 6-7 and Figure 6-8) demonstrate how an agent subscribing to Levi's theory and who, importantly, is risk-averse and so selects the admissible option with the greatest minimum expected utility, violates McClennen's DYN-SUB* condition stated above.

Figure 6-7

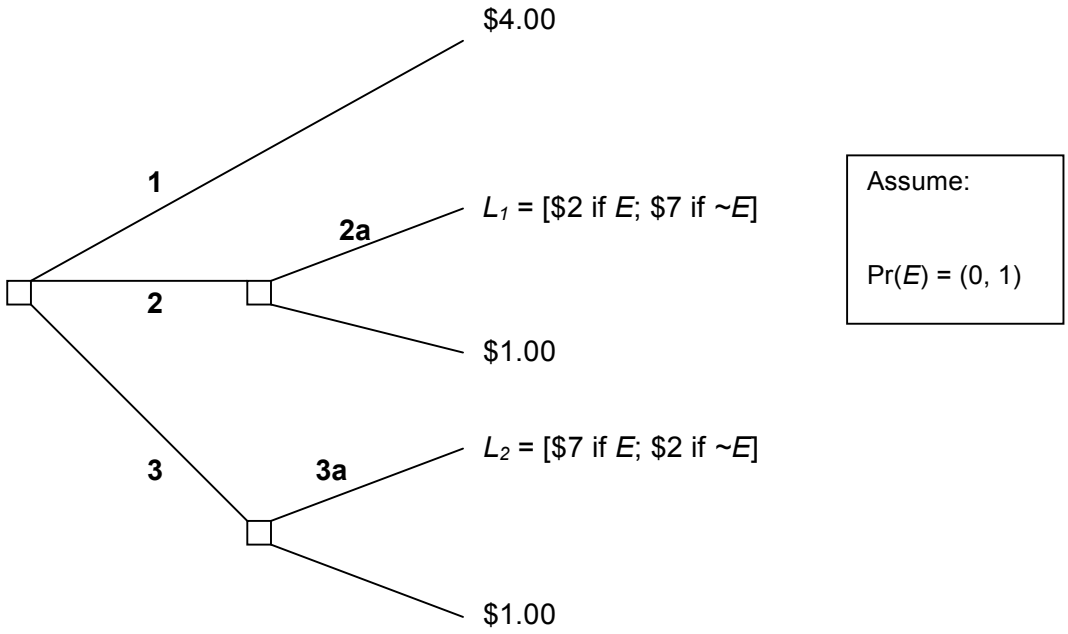
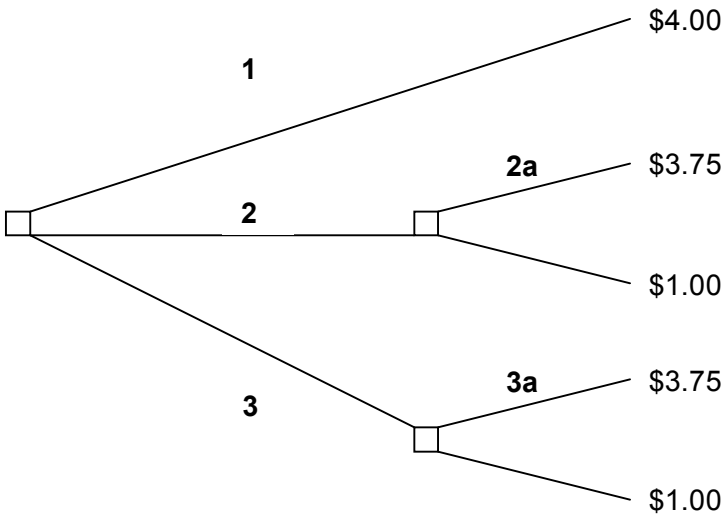


Figure 6-8



The agent in the choice situation in Figure 6-7 will choose either Plan 2a or Plan 3a, because these yield the lotteries L_1 = ($7.00 if E; $2.00 if ~$E$) and L_2 = ($2.00 if E; $7.00 if ~$E$) respectively. Plan 1 yields a sure $4.00, and in the presence of the other two lotteries this is ruled out because it doesn't maximise expected utility for any probability-utility function pair. However, when each of the lotteries above is

swapped with an outcome that is preferred according to a pair-wise comparison (as in Figure 6-8), the agent now chooses Plan 1, because it stochastically dominates the other two plans.

Now Seidenfeld (1988b, p. 312–313) argues that the above DYN-SUB* condition essentially begs the question against the security-conscious ordering-violator. He claims that DYN-SUB* is just a restatement of the transitivity condition, so an agent who subscribes to Levi's theory (and who appeals to greatest minimum expected utility to break ties) violates even the static version of the DYN-SUB* principle. (To see that such an agent violates transitivity in the static setting just consider the pair-wise preference relations between the three options \$4.00, \$3.75 and L_1 = (\$7.00 if E; \$2.00 if $\sim E$) when security criteria are used to select from the admissible set.) The stochastic dominance condition on the other hand, Seidenfeld asserts, is acceptable (in the static context) to both those who relax independence alone and those who relax ordering, and so the question as to whether these agents satisfy the condition in sequential contexts is not already prejudiced.

I agree that DYN-SUB* begs the question against ordering-violators, but as stated above, I also think that the dynamic SD condition begs the question against independence-violators. At core, I am not convinced by Seidenfeld's coupling of first-order SD with his dynamic SD condition, and the way this is contrasted with the static and dynamic "versions" of DYN-SUB*. The dynamic SD condition is not a mere extension of first-order SD; it is rather a stronger condition that essentially amounts to the independence axiom. So I think it is not quite right to say that dynamic SD is acceptable in the static context. Of course, we assume that everyone can agree to *first-order* stochastic dominance in the static context, but dynamic SD is something quite different altogether. Thus I don't see any real asymmetry between Seidenfeld's dynamic SD and the DYN-SUB* condition that McClennen toys with. Sure, DYN-SUB* subtly begs the question against Levi's agent by enforcing transitivity of preferences. But again, dynamic SD subtly begs the question against the independence-violator by essentially enforcing the substitution of lotteries condition.

6.7 Re-interpreting the dynamic stochastic dominance condition

Seidenfeld (1988a, pp. 284–288) outlines how his dynamic SD condition should be interpreted when it comes to theories like Levi's that relax ordering. (And Seidenfeld goes on to show that Levi's theory satisfies dynamic SD.) It seems to me that the independence-violator can follow suit, and offer an interpretation of dynamic SD that is more congenial to their starting position. Indeed, here I will try to mirror Seidenfeld's approach for the case of independence violation alone. In the next section, I will discuss why my suggestion here ultimately fails. It fails for a rather subtle reason, however. In fact, the problem for decision theories that violate only the independence axiom of SEU theory turns out to be related, but nonetheless somewhat tangential, to the dynamic SD condition. Thus I think it is worth continuing here with my criticisms of the dynamic SD condition on rational choice rules.

Recall that the dynamic SD condition requires that admissible choices under a rule be stochastically un-dominated, and importantly, that admissibility be preserved under substitution (at choice points) of "indifferent" options. It seems that the trick to applying this condition to Levi's theory rests on how we interpret "indifferent options". If ordering is relaxed, Seidenfeld asserts that options are indifferent "if and only if, whenever both are available either both are admissible or neither is". In other words, options are not indifferent just in case there is *some* situation (say the pair-wise comparison) where both are admissible. It must be the case that in all choice settings where one is admissible the other is admissible as well. This effectively limits the applicability of the dynamic SD condition when the ordering postulate is relaxed. For instance, the pair of choice scenarios illustrated in Figures 6-7 and 6-8 does not expose a violation of dynamic SD for Levi's theory because this is not a case of substituting "indifferents" at choice nodes, according to the definition just stated. Note that it is not the case that in every situation where both \$3.75 and, say, the lottery L_1 = (\$7.00 if E; \$2.00 if ~E) are available, both are admissible or neither is. (Just add the other lottery L_2 = (\$2.00 if E; \$7.00 if ~E) to the choice set, and we get a case where L_1 is admissible, but \$3.75 is not.) Seidenfeld (1988a, pp. 284–8) proves

the general result that in all cases Levi's theory satisfies dynamic SD, if it is interpreted in the way just described.

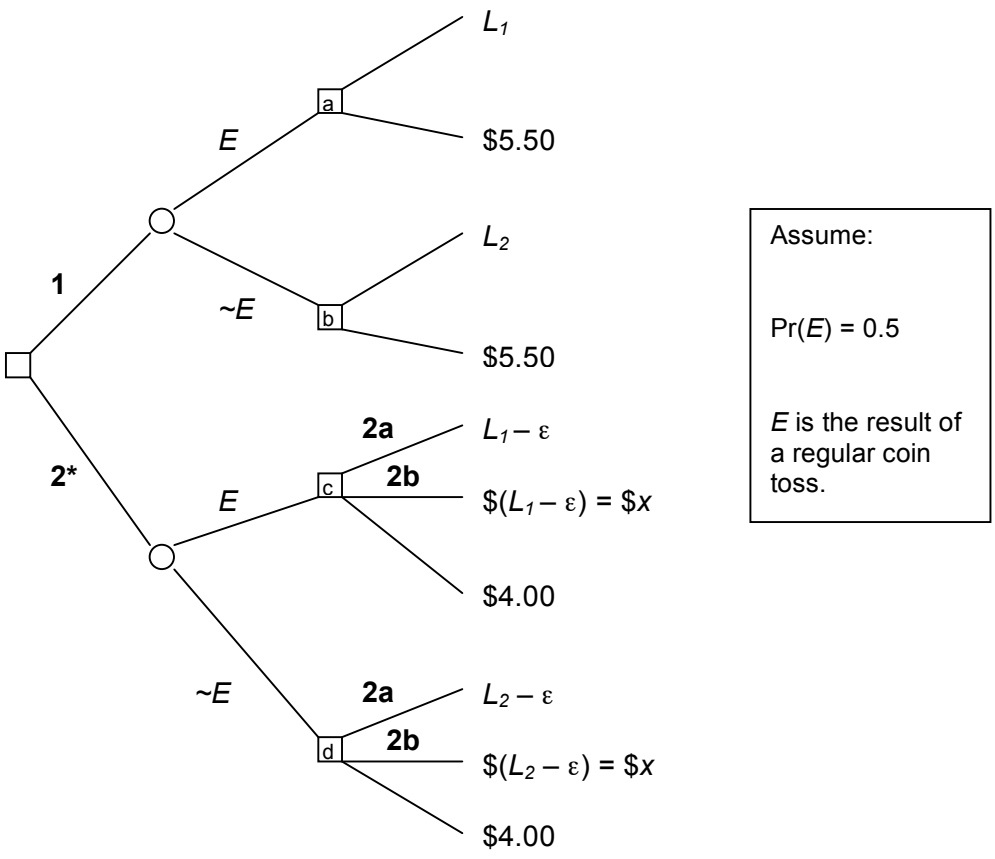
A similar re-interpretation of the dynamic SD condition seems appropriate to theories that relax independence alone. What counts as an exchange of "indifferents" at choice nodes, in particular, arguably needs to be more carefully specified in the domain of independence-violating theories. Admittedly, unlike the case where ordering is relaxed, "indifference" is already well defined here, at least in the static context. But it could be argued that when we are talking about the exchange of sub-trees or sub-lotteries, "indifference" requires more than just $L_1 \approx L_2$ when the two lotteries are compared in isolation. If we seek a condition for preserving preferences between plans, then sub-lotteries are "indifferent" if and only if it is the case that for any third lottery C and $0 \leq \alpha \leq 1$, $\alpha \times L_1 + (1 - \alpha) \times C \approx \alpha L_2 + (1 - \alpha) \times C$. Of course, this is effectively just the independence condition, restricted to the relation of indifference. It states that only in cases where lotteries are indifferent by "independence standards" can we expect their substitution to preserve preferences between plans. The dynamic SD condition is thus rendered superfluous. In general, I don't think a mere formal statement of any kind of dynamic SD condition is going to cut either way with respect to assessing decision theories that relax ordering versus theories that relax *just* independence. Dynamic SD might be interpreted such that it is satisfied, or in a way that leads to violations. Either way, one may well ask if anything new has been learnt about the decision theory in question.

6.8 Not so fast when it comes to defining "indifferents"

As forewarned, my suggestion above can be criticised, but for rather unobvious reasons. It can be shown that there is more at stake when it comes to defining what counts as a substitution of "indifferents" and how such a substitution should affect the admissibility of plans. In fact, I think Seidenfeld's (1988b, 2000a, 2000b) arguments

to this effect are much stronger than his original defence of the dynamic SD condition. He effectively changes his tack in response to the criticisms of dynamic SD that I discussed in the previous sections. The issue, Seidenfeld claims, revolves around how a sophisticated chooser should negotiate sequential decisions where they expect to be indifferent between two or more of the branches at some interior choice node. It turns out that such situations raise problems for the (desirable) assumption that it should always be possible for an agent to have dynamically stable preferences. The issue is best illustrated by considering a decision problem such as the one in Figure 6-9 below, which is effectively an amalgamation of the decision problems in Figures 6-4 and 6-5.

Figure 6-9



The original question was whether the lottery ($L_1 - \varepsilon$) and its cash equivalent $\$x$ (not to mention the lottery ($L_2 - \varepsilon$) and its cash equivalent, also $\$x$) should be considered

"indifferent" in the sense that they can be exchanged at choice nodes without affecting the admissibility-status of the plan. A different way of asking the question is whether Plans 2a and 2b in Figure 6-9 should both have the same admissibility status, i.e. whether it should be the case that either both are admissible or neither is. Seidenfeld argues that the latter is indeed a requirement in the dynamic-choice setting (if we want to uphold ordering), and that the condition cannot be reconciled with preferences that violate independence. For example, at the initial choice node in Figure 6-9, the Plans 2a and 2b are evaluated as follows:

$$\text{Plan 2a} = 0.5 \times (L_1 - \varepsilon) + 0.5 \times (L_2 - \varepsilon) = \$5.75$$

$$\text{Plan 2b} = 0.5 \times \$x + 0.5 \times \$x = \$x \text{ (where } \$4.00 < \$x < \$5.00)$$

Only Plan 2a is admissible. (And in Figure 6-9 even Plan 1 is preferred to Plan 2b.)

So why does Seidenfeld think there is, alas, a requirement for Plans 2a and 2b to both be either admissible or inadmissible? Well, we can see that at each of the choice nodes (c) and (d), the agent is "indifferent" between the lottery and its cash prize equivalent. When it comes to actually picking one of the options at either choice node, the agent will have to adopt some kind of tie-breaking rule. Importantly, the agent is, at the choice node, ambivalent about what tie-breaking rule they actually use. At this point, any rule will do. And so the two plans should stand or fall together.

We can see that the requirement for Plans 2a and 2b to both be admissible is going to cause some problems. It is at odds with the evaluation of the plans above, which shows that Plan 2a has greater utility from the perspective of the initial node. One might think that this preferred plan can surely be realised—it is a performable plan, and all that is required at the second choice node is that the tie-breaking rule goes in favour of the lottery. Some tie-breaking rules will indeed permit this, but the problem is that banking on a particular tie-breaking rule is not in the spirit of tie-breaking, and in any case, when the agent actually reaches the second choice node, they will be genuinely indifferent between the options, so at this point any tie-breaking rule will

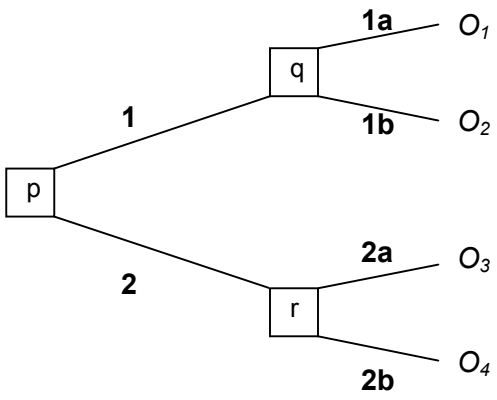
do. If the agent expects to have stable preferences through time (which should arguably always be a possibility), then it looks like we have an inconsistency. And such an inconsistency may provide the necessary motivation for Seidenfeld's dynamic SD condition.

I do think the above analysis reveals an inconsistency, but there are "ways out" of the problem that I think Seidenfeld does not attend to. What the above argument shows, I think, is that the sophisticated chooser cannot count as an available plan just any combination of choices that *might* end up being made within the course of a dynamic decision. Rabinowicz (1995, p. 619) refers to plans that *might* end up being performed as "weakly performable". This is to be contrasted with plans that are expected to definitely come about, if only the appropriate first move is made (and beliefs/preferences remain stable). The Plans 2a and 2b (from Figure 6-9) are only weakly performable—we are assuming that the agent expects to be indifferent about the options at the second choice node, whether it be (c) or (d), and so may at this point go either way. That is, the agent may intend to follow Plan 2a, but this plan could fall through because at the second choice node the agent may end up choosing according to Plan 2b. There is something of a problem here. Surely the set of available plans should only include those that the agent can be certain she will follow through. But we do not want to just exclude plans like 2a and 2b. After all, this is merely a case of the agent expecting at some stage to reach a choice node where she is indifferent between two or more available branches. So the question is: how should we account for the weakly performable plans?

Seidenfeld (1988b) provides a very good answer to this question where it applies to a sophisticated chooser who abides by Levi's decision theory. I presume that he thinks this style of response is not open to the independence-violator. I will suggest otherwise. But first let me examine the problem as it arises for Levi's agent.[136]

[136] Note that this example is taken from Hammond's (1988a) comments on Seidenfeld's result regarding the dynamic inconsistency of Independence-violators. Seidenfeld (1988b) gives an account of this example (as I will go on to outline) in his rejoinder to Hammond.

Figure 6-10



As per Hammond (1988a, pp. 294–295) and Seidenfeld (1988b, pp. 310–311), let us assume that the agent's preference structure is such that, from the outcome set $\{O_1, O_2, O_3, O_4\}$, only O_1 and O_3 are admissible. But if the available set is restricted to $\{O_1, O_2\}$ or $\{O_3, O_4\}$, both options in each case are admissible. Assume also that the agent is risk neutral, so security criteria are not used to further differentiate options in the admissible set. Now there seems to be the same kind of inconsistency lurking here as there was for the independence-violating agent in Figure 6-9. Working backwards, we see that the agent may choose either O_1 or O_2 at choice node (q), and either O_3 or O_4 at choice node (r). But when the agent assesses these plans at the initial choice node (p), she is not indifferent between Plans 1a and 1b, and Plans 2a and 2b. In fact only Plans 1a and 2a are admissible; it is desirable that she chooses "up" at the second choice node, whether it be (q) or (r). At this choice node, however, the agent may be disposed to choose "up" or "down", because both options are admissible at this point. So she cannot guarantee from the outset that Plan 1a or 2a will take place; they are only weakly performable.

Now Seidenfeld resolves this troubling situation in the following way. Given that the agent is uncertain about whether she will choose "up" or "down" at the second node, she should assign the maximally uncertain convex set of probability distributions to her choice at either of these nodes. It is as if the available plans are not 1a, 1b, 2a and 2b, but rather Plan 1* and Plan 2*, where 1* is a combination (via indeterminate

probabilities) of Plans 1a and 1b, and 2* is likewise a combination of Plans 2a and 2b. Of course, the use of indeterminate probabilities to represent uncertainty about the outcome of initially choosing "up" or "down" involves a relaxation of ordering, but this is already one of the assumptions of Levi's theory.

Rabinowicz (1995, p. 619 & 2000) suggests something very similar to Seidenfeld's proposal above, but in relation to choice rules that relax independence (as per the decision scenario in Figure 6-9). He claims that weakly performable plans like 2a and 2b (in Figure 6-9) can be counted as available, but they must be evaluated according to their expected *choice* value rather than their expected *outcome* value. The expected *choice* value is calculated similarly to the way Seidenfeld evaluates Plans 1* and 2* above, but Rabinowicz appeals to the principle of indifference, rather than indeterminate probabilities, to represent uncertainty about what the agent will choose when they are indifferent between tree branches. According to Rabinowicz, both Plans 2a and 2b (in Figure 6-9) have the same expected *choice* value, which is the average of their individual values: $0.5 \times \$5.75 + 0.5 \times \$x$. He notes that this makes the value of plans, and thus the admissibility of plans, dependent on context. If Plan 2a is present in isolation, it will have a value of $5.75, but when it is associated with Plan 2b it has a lesser value due to the fact that this less attractive plan might be the one that actually eventuates. The ordering postulate is thus violated. (And this is perhaps why Rabinowicz in fact favours an alternative account that I will not give any prominence here, because I do not think it is successful.[137])

In my opinion, neither Seidenfeld nor Rabinowicz sufficiently explores the avenues open to the independence-violator when it comes to the problem of ensuring the possibility of stable preferences. Seidenfeld does not try to apply his solution to the independence-violation case, and Rabinowicz, I think, needs to expand on his

[137] According to Rabinowicz's alternative account, Plan 2a (in Figure 6-8) has value $5.75 and Plan 2b has value $\$x$, meaning that only Plan 2a is admissible. The agent can be confident that she will stick to Plan 2a because her later self will deviate at a choice node only if there is a positive reason to change plans. Later indifference between 2a and 2b is not positive reason to deviate from the pre-selected plan. (I don't like this account because it has hints of "resolute" choice—it requires the agent to keep track of what they had previously regarded the admissible plans; she is not free to choose afresh at each new choice node.)

suggestion outlined above. Both acknowledge that, when evaluating strategies, an agent should be uncertain about how they will choose at any later node where they expect to be indifferent between branches. One employs indeterminate probabilities to represent this uncertainty, while the other relies on the flat distribution to do this job (as per the principle of indifference). I want to consider this latter proposal in a bit more detail. The flat distribution might seem an attractive way to represent uncertainty because it is arguably a more conservative approach than the adoption of indeterminate probabilities. On first appearances, it seems that we can avoid a violation of ordering if we use the flat distribution to model uncertainty about future choices (for the case of the independence-violator). But, as stated, Rabinowicz's expected-choice-value account does violate ordering; it makes the value of strategies dependent on what other strategies are potentially available. So the status of the ordering axiom does not really go in favour of the flat-distribution approach.

In fact, there are a number of problems with using the flat distribution to model uncertainty about how an agent will later choose between indifferent options. When the choice function that an agent adopts violates independence, then the use of any particular distribution to model uncertainty about a later selection amongst indifferent options has some worrying features.[138] It may well be the case that the uncertain mixture of options at the choice node in question does not have the same value as either of the indifferent options. For instance, an independence-violator may find that applying the flat distribution over two indifferent options S_1 and S_2 (where these options stem from some future choice node) effectively amounts to a mixture of the options that is worth more than either one of them. That is, it might be the case that:

$$0.5 \times S_1 + 0.5 \times S_2 \succ S_1, S_2$$

In such case, it seems that S_1 and S_2 will be overvalued from the perspective of an earlier choice node, and this will affect the overall value of the strategies that these options (or partial strategies) are a part of. When the agent actually reaches the node in question, their indifference between S_1 and S_2 does not translate to taking a mixture of the two options. At this point, one would think that the value of the agent's choice

[138] I thank Teddy Seidenfeld for bringing this issue to my attention.

should equate to the value of S_1 (or S_2). (Of course, the mixture of S_1 and S_2 may also be an available option, in which case we will not have a case of indifference because the mixture will be the most preferred option.)

Perhaps the independence-violator can live with the oddity of it being possible for an uncertain choice between two indifferent options to be worth more (or less) than either option on its own. But this puts a lot of pressure on selecting the correct distribution to represent uncertainty. Use of the flat distribution for this purpose is not uncontroversial. When an agent is indifferent between a number of options, it is questionable whether we should say that each one has an equal probability of being chosen. In fact, this issue takes us back to the general problems associated with the "principle of indifference" that I raised in the previous chapter. I argued that genuine uncertainty or ignorance is more accurately represented by indeterminate probabilities than by a single "maximally uncertain" probability distribution. And the situation that we are concerned with here—when an agent does not know what they will choose at some future choice node where they expect to be indifferent between options—is surely best considered a case of ignorance. This is why it seems most reasonable to represent the agent's uncertainty about their future choice between "indifferents" with the maximally uncertain convex set of probability distributions.

When it comes to modelling how they might later choose amongst "indifferents", then, appealing to indeterminate probabilities seems the most plausible way for the independence-violator to proceed. Admittedly, it would be no small task to generalise a decision theory like cumulative prospect theory in order to accommodate indeterminacy of belief/preference. Levi's decision theory, by comparison, is a rather modest modification of SEU theory, given that admissible options must maximise expected utility for at least some probability-utility pair. Notwithstanding this issue, I do not think Seidenfeld's argument serves as an unassailable obstacle for decision theories like cumulative prospect theory that relax the independence axiom. Of course, Seidenfeld is careful not to overstate his case: he claims merely that any choice rule relaxing independence *alone* will not permit the possibility of stable preferences in every conceivable choice scenario. This is to say that cumulative

prospect theory, just as it is, will lead to certain inconsistencies in the sequential-choice context when preferences are assumed to be stable. But if we make some modifications to such decision theories, i.e. employing indeterminate probabilities in the manner described above, or at least appealing to the flat distribution to account for uncertainty about future choices between "indifferents", then we can escape the inconsistencies that Seidenfeld draws attention to.

By way of a summary, I will retrace some of the steps in this rather involved discussion. Recall Seidenfeld's initial claim that the exchange of "indifferents" at choice nodes should preserve the admissibility of plans. I have been referring to this condition as "dynamic SD". Theories like cumulative prospect theory that relax *just* the independence axiom do not satisfy dynamic SD. But the obvious question is why dynamic SD should be a constraint on rational choice functions. Seidenfeld initially argued that this condition simply follows from a commitment to first-order stochastic dominance. Like McClennen, I do not think this is satisfactory; dynamic SD is not a simple extension of first-order stochastic dominance, and it would not appeal to someone who is already prepared to relax independence. This is not to say, however, that there is not some further reason to require the exchangeability of "indifferents" in the way that the dynamic SD condition stipulates. Indeed, in this last section I have analysed Seidenfeld's arguments to this effect. But I do not think he quite succeeds in bolstering the case for dynamic SD. When more than one indifferent option is available at a choice node, then the individual strategies that incorporate these options need not stand or fall together. It is simply the case that the agent is uncertain about what they will choose at the node in question, and we can employ either the flat distribution or indeterminate probabilities to represent this uncertainty.

6.9 Conclusion

What Seidenfeld does uncover with his dynamic SD argument, I think, are some deep issues for independence-violating theories when it comes to tie-breaking. The

sequential-choice scenario draws attention to these issues because here we have the possibility that an agent must model how they will choose at some future node where they are indifferent between options. When it comes to SEU theory, if an agent is indifferent between options, then they are ambivalent about which tie-breaking function might be used to finally select one option. This is because any mixture of indifferent options has the same value as the indifferent options themselves. But this is not always the case for independence-violating theories. Different tie-breaking functions may be worth different amounts, because the distributions of outcomes associated with the mixtures may not be identical. So we cannot say that indifference amounts to being ambivalent about which tie-breaking function will be used. The question then is how should we model and evaluate the choice between "indifferents" when it comes to a decision rule that relaxes the independence axiom. I argued in the previous section that the use of any particular distribution to model uncertainty about what an agent will choose puts a lot of pressure on the chosen distribution being the correct one. It seems more in keeping with the subjectivist approach to model such uncertainty with indeterminate probabilities, but, admittedly, such a move adds a lot of complexity to a decision rule that was originally intended to relax only the independence axiom.

We can say then that the sequential-choice framework does uncover a grave problem for decision theories like cumulative prospect theory that relax just the independence axiom of SEU theory. In effect, any such decision theory requires some added complexity if it is to be free of inconsistencies in the sequential-choice setting. The added complexity will involve a relaxation of ordering, whether we decide to model uncertainty about future choices with the flat distribution, or with indeterminate probabilities. When it comes to Seidenfeld's argument then, it is not independence-violation *per se* that causes problems in the sequential-choice context. If we are prepared to relax ordering as well as independence, then we can escape the sequential inconsistencies that Seidenfeld has drawn attention to. Note also that Levi's decision theory is free from these sequential inconsistencies, and I argued in the previous chapter that Levi's decision rule violates independence as well as ordering (if the agent is security-conscious).

It seems then that it is the specific form of a decision rule that matters, rather than simply whether or not the rule violates independence. The arguments that were intended as outright defences of specific SEU preference axioms—Hammond's consequentialist argument for SEU theory in its entirety, and the diachronic-Dutch-book-style argument for upholding independence—turned out to be, according to my analysis, unsuccessful. The latter of these arguments has much plausibility, but in the end, I do not think its featured sure losses provide a sufficiently strong case for deeming a decision rule "irrational". In particular, I do not think the possibility of Dutch-book sure losses provides compelling reason to reject Levi's lexical decision rule. Other decision rules that violate independence (and necessarily, ordering, in order to escape the dynamic inconsistencies that Seidenfeld exposes) would need to be assessed on a case-by-case basis. To be at all plausible, the merits of such theories (with respect to their treatment of risk) would have to outweigh the negative Dutch book result.

CONCLUSION

Let us reconsider the guiding questions for this thesis. Recall from the introduction that I set out to investigate whether the notion of precautionary reasoning presents a challenge to the dominant normative decision model—subjective expected utility (SEU) theory. Such an investigation naturally divides into two parts, corresponding to the two somewhat distinct senses of "precaution". In Part I of the thesis, I was concerned with the temporal dimension of choice—I considered what are rational expectations about the future, in particular, future beliefs and desires, and how such expectations should impact on decision-making. This involved thinking about the dynamic structure of a decision problem, something that is not well represented by the standard static decision model. It also required an investigation of the rule of conditionalisation for updating belief. My conclusions about the diachronic Dutch book argument (DBA) and planned changes in belief gave important perspective to my analysis of planned changes in desire, a topic that has been little discussed in the literature. Part II of the thesis dealt with challenges to SEU theory itself, in particular, the theory's treatment of uncertainty. I drew attention to the main types of modified decision theory that have arisen in the literature and which permit choice to be affected by differing measures of "risk". Two representatives of these alternative decision theories are cumulative prospect theory and Levi's decision theory for handling indeterminate beliefs/preferences. The task was to evaluate the motivations behind such theories, and to determine whether they provide a credible challenge to the monopoly that SEU theory has on rational choice. In the course of my analysis, I identified some important links between sequential-choice arguments concerning the assessment of decision rules, and my favoured version of the diachronic DBA for conditionalisation.

The sequential representation of a decision problem proved critical for investigating both risk- and future-oriented decision-modelling issues. I hope to have made clear,

however, that different projects or purposes call for differing sequential-choice assumptions. This is something that I do not think is sufficiently well appreciated in other sources. It is one thing to approach the sequential-choice model as a tool for assisting an agent with their on-the-ground decision-making. It is quite another matter to employ the sequential-choice framework for the purposes of justifying particular updating rules or choice-function constraints. In the latter context, it is reasonable to think that all variables (apart from the updating plan or decision rule that is being tested) should be held fixed. Typically, this means considering only those decision situations in which an agent expects their conditional beliefs/desires to be stable with time. This does not amount to ignoring the fact that beliefs/desires can change in unplanned ways. It is just that the justificatory, as compared to the actual decision-making use of sequential-choice models, warrants some extra idealisations. Moreover, it is quite legitimate to apply different criteria in these two separate contexts when it comes to analysing decision problem solutions.

When it comes to ordinary "on-the-ground" decision-making, I argued in Chapter 1 that the appropriate method for assessing sequential-choice strategies must be very permissive with respect to an agent's expectations about their future self. For whatever reason, a given agent may predict that they will hold radically different beliefs or preferences at some future point in time, and our sequential-choice method must be able to accommodate any such predictions. In fact, I argued that only the "sophisticated" approach to strategy assessment is up to this task. Of course, I do not want to suggest that a rational agent need not at least have respectable *plans* when it comes to updating their beliefs and preferences. It is just that the agent may not be confident that their best-laid plans will come to pass. In addition to arguing for sophisticated choice, I addressed the relationship between sequential and static decision models: my position is that the two should answer to the same problem, which effectively means that, in ordinary decision-making scenarios, static decision models must be "brought into line" with sophisticated findings.

While an agent's actual changes in belief need not conform to the rule of conditionalisation, their premeditated *plans* for updating their beliefs should conform

to this rule. It is this weaker interpretation of conditionalisation that is at least partially defended by the "sophisticated" diachronic DBA, which I argued to be the only plausible version of the argument. In fact, I contended that there is an even more convincing *non-pragmatic* case for conditionalisation, when it is weakly interpreted as a rule governing how an agent should merely plan to update their beliefs. For the desire/preference case, I argued in Chapter 3 that the situation is rather more complex. I showed that the diachronic DBA does not look so convincing as a means for defending any particular desire-updating rule. Moreover, I argued that while it is perfectly reasonable for an agent to plan on updating their desires in line with their current conditional desires, there is yet another way in which we might say that an agent *plans* a change in desire—they might have a "higher-order" preference for some alternative utility function, and so pursue a strategy that makes this change in desire more likely. I take this suggestion to be a novel one, which is why I explored the logic of acting on "higher-order" preferences in some detail in Chapter 3.

This talk of updating plans brings me to the second use of the sequential-choice model: as a tool for justifying decision methods. The aforementioned diachronic DBA for conditionalisation as the rule for updating beliefs is a prime example—the version of the argument that I defended in Chapter 2 revolves around the assessment of sequential decision strategies. Importantly, the argument concerns only ideal decision scenarios in which the value of monetary prizes is expected to be stable with time. It is shown that, in some cases, a non-conditionalising agent will forgo sure gains that would be available to the conditionalising agent, and there are no scenarios in which the conditionalising agent is vulnerable to the same sort of losses. As discussed, the diachronic DBA loses its bite when it comes to defending a rule for updating desire. In this case, we cannot assume that the value of money, or any other sort of outcome, will be constant with time, because that is precisely what is at issue! This means that, in the desire case, we cannot appeal to sure losses as the telltale sign of inconsistent updating plans. The diachronic DBA does have more general applicability, however, and this does not seem to have been recognised elsewhere. In Chapter 6, I showed that this style of argument is also pertinent to the assessment of decision rules. Here we focus only on decision scenarios in which both conditional beliefs and desires are expected to be stable with time. I showed that any theory violating the independence

axiom is vulnerable to what we might call "Dutch book sure losses", meaning that the agent's very decision rule can make a dominating strategy unavailable to them.

I indicated in Chapters 2 and 6 that I find the (sophisticated) diachronic-Dutch-book-style arguments rather compelling. Indeed, they are much more persuasive than a closely related sequential-choice argument—Hammond's so-called "consequentialist" argument for SEU theory in its entirety. Hammond requires that a rational choice rule make available, in ideal cases, all strategies mapped out by the sequential decision tree, or in other words, all combinations of choices at choice nodes. (In fact, Hammond's argument can be likened to the naïve version of the diachronic DBA, which I rejected in Chapter 2.) I argued that Hammond's criterion for rational choice rules is not sufficiently well motivated. The possibility of "Dutch-book sure losses", on the other hand, is, I think, a relevant consideration when it comes to assessing decision methods. But I do not think the possibility of such losses provides sufficient reason for labelling a decision method "irrational". Recall that Dutch-book sure losses do not rest on an agent choosing a strategy that is dominated by another *available* strategy; rather, the agent suffers sure loss relative to an unavailable or "dynamically unfeasible" strategy. This does not amount to an outright inconsistency on the part of the agent. I thus concluded that the diachronic-Dutch-book-style arguments provide some defence for conditionalisation as the rule for updating beliefs, and for upholding the independence axiom of preference, but they do not prove that it would be irrational to do otherwise.

There is another key argument concerning decision rules that appeals to the sequential-choice setting—Seidenfeld's sequential-choice argument, which is intended to call into question those theories that relax the independence axiom, and except theories that relax ordering. In Chapter 6, however, I drew attention to the fact that Seidenfeld's argument only targets certain types of independence-violating theories—he exposes problems for decision rules that relax *just* the independence axiom of SEU theory. For instance, Levi's decision rule, which violates both ordering and independence (or so I argued in Chapter 5), is free from the sequential-choice inconsistencies that Seidenfeld raises. In addition, we might modify an independence-

violating theory like cumulative prospect theory, so that it too can escape Seidenfeld's sequential-choice inconsistencies. (Such a modification would necessarily involve a relaxation of ordering.) I am led to conclude, then, that it is the specific combination of properties that a decision rule has, rather than whether or not it upholds (violates) particular SEU axioms, that renders it (in)consistent in the sequential-choice context.

So I do not think the sequential-choice context yields any knockdown arguments against violations of any particular SEU axiom of preference. There is scope to pursue decision theories that do not violate independence *alone*. In this respect, however, we should proceed with caution: I have argued that the possibility of Dutch-book sure losses is a mark against decision rules that violate independence, so it is reasonable to require that any "risk-sensitive" alternative to SEU theory be very well motivated. With this in mind, I am more inclined to defend Levi's decision theory than any modified cumulative prospect theory, or other theories of the same ilk. As discussed in Chapter 4 in relation to Allais's problem, the latter kind of decision theory permits the choice-worthiness of acts to be affected by their distribution of outcome utilities. While I did not express any strong views one way or the other on this issue, I do not think such decision rules are sufficiently compelling to outweigh the Dutch-book sure losses, not to mention the added complexity that would be necessary to overcome the sequential inconsistencies that Seidenfeld draws attention to. In fact, I argued that SEU theory can accommodate tangible risk/regret emotions, and such emotions might be all that drives the "Allais-choices" and similar behaviour. What concerns me more is the status of decision rules that permit sensitivity to the kind of risk associated with indeterminate belief/preference. In my opinion, it is reasonable for an agent's choices to be affected by the degree of (in)determinacy of the relevant beliefs and desires. As discussed in Chapter 5, I favour Levi's lexical decision rule for handling indeterminacy. According to this rule, admissible options must at least maximise expected utility for some probability-utility pair, but final choices can be made on the basis of "security" considerations. While I claimed that the rule violates ordering and independence and is thus vulnerable to Dutch-book sure losses, I think its merits are sufficient to outweigh this shortcoming.

In light of the sequential-choice arguments that I have identified, then, I hold that there is at least one plausible "precautionary" alternative to SEU theory—a decision theory (Levi's) that incorporates indeterminacy of belief/value, and associated risk-sensitivity. In practical decision-making terms, this is a significant result. It permits beliefs and values in different decision situations to have varying "sharpness", and this may affect the choice-worthiness of options. The level of sharpness or determinacy (for beliefs at least) will likely depend on how much relevant evidence is available. For instance, a given agent (whether a single person or an institution) may have relatively sharp/determinate beliefs about the effect that a 0.2% rise in interest rates will have on the national economy, as compared to their beliefs regarding the consequences of a 2 degree rise in average global temperature. I support the position that it is permissible for agents to have varying reactions towards the "risk" associated with indeterminacy. A security-conscious agent (which could potentially be a public institution) might prefer to base their decisions on the worst-case expected utility of acts. This has significant implications for public decision-making, and, in particular, our understanding of the Precautionary Principle, which I discussed in the introduction to this thesis. This is not to say, however, that it would be irrational (although it may be unethical) to take a more optimistic approach when decision-making in the presence of indeterminacy, and so compare acts according to their respective best-case expected utilities.

Finally, I want to consider the theoretical implications of regarding rivals to SEU theory as rationally permissible. In Chapter 2, in particular, I highlighted the prominent role that SEU theory plays in relation to the justification of epistemic norms. It is commonly thought that the SEU representation theorems provide the best defence of probabilism, which is the view that degrees of belief should conform to the probability calculus. As the representation-theorem story goes, a rational agent's preferences should satisfy a limited number of axioms (including independence and ordering), in which case, the agent can be represented as an expected utility maximiser with a unique probabilistic belief function and a value/desire function that is unique up to positive linear transformation. My findings indicate, however, that a rational agent need not satisfy all the SEU preference axioms. And in such case, it is not clear that there is going to be a unique belief-desire representation of the agent's

ordinal preferences. This means that we cannot establish that an agent with rational preferences necessarily has a probabilistic belief function. It might be contended that this just highlights the important role of the synchronic Dutch book argument (DBA). I argued in Chapter 2 that the synchronic DBA is a less ambitious defence of probabilism than the representation theorems, and this could well be to its advantage. Notably, the argument requires only that the value of monetary bets be additive in *some* betting domain.

As indicated in Chapters 2 and 6, I do have sympathy for the Dutch book arguments. But I also suggested that these pragmatic arguments for probabilism and conditionalisation are somewhat redundant. To begin with, I have been arguing that diachronic-Dutch-book-style arguments are not as strong as one might hope. Besides, conditionalisation does not seem to require a diachronic Dutch book defence; I argued in Chapter 2 that it is simply a matter of correctly interpreting conditional probabilities that we should plan to update our beliefs in accordance with these probabilities. Additionally, it could well be contended that the probabilist epistemic norms are more self-evident than any assumptions about betting behaviour that we might endorse in order to defend them. There are also other justificatory possibilities for these norms; we might appeal to alternative non-pragmatic arguments, such as Joyce's (1998) defence of probabilism. Indeed, while it is important to recognise that the epistemic and the pragmatic are intimately related, it is surely more productive, if not necessary, to pursue independent justifications for the relevant norms. Intuitively speaking, belief does not always seem connected to action, and so it is reasonable to think that norms for belief should be compelling in the absence of any pragmatic considerations. Furthermore, it is surely beneficial to the study of practical rationality that constraints on preference not be burdened with much broader representational and epistemic significance (as per Savage's or Jeffrey's representation theorems). It seems more constructive to begin with suitable numerical belief and desire functions, and then set about formulating various possible decision rules, and subsequently assessing their merits as models of rational choice.

APPENDIX 1

Savage's (1954) Expected Utility Theorem[139]

DEFINITIONS:

S is the set of states, X is the set of consequences (outcomes), and F is the set of all functions on S into X.

$A, B \subseteq S$; $x, y \in X$; $f, g \in F$

$<$ on F is the basic binary relation with $\approx$ and $\leq$ defined in the usual way: $f \approx g \Leftrightarrow$ (not $f < g$, not $g < f$), and $f \leq g \Leftrightarrow (f < g$ or $f \approx g)$.

$f = g$ on $A \Leftrightarrow f(s) = g(s)$ for all $s \in A$. $f = x$ on $A \Leftrightarrow f(s) = x$ for all $s \in A$.

A^c represents the complement of A.

A is null $\Leftrightarrow f \approx g$ whenever $f = g$ on A^c.

$x < y \Leftrightarrow f < g$ when $f = x$ and $g = y$ on S.

$x < f \Leftrightarrow g < f$ when $g = x$ on S. Similar definitions hold for $x \approx y$, $f \approx y$, $x \leq f$, and so forth.

Conditional preference is defined as follows: $f < g$ *given* $A \Leftrightarrow f' < g'$ whenever $f = f'$ and $g = g'$ on A, and $f' = g'$ on A^c. $\approx$ *given* A and $\leq$ *given* A are defined in the usual way. $x < g$ *given* A means that $f < g$ *given* A whenever $f = x$ on A.

THEOREM:

---

[139] This presentation of Savage's theorem is taken directly from Fishburn (1970, pp. 191–193).

Suppose that the following seven conditions hold for all $f, g, f', g' \in F$, $A, B \subseteq S$ and $x, y, x', y' \in X$:

P1. $<$ on F is a weak order;

P2. ($f = f'$ and $g = g'$ on A, $f = g$ and $f' = g'$ on A^c) $\Rightarrow$ ($f < g \Leftrightarrow f' < g'$);

P3. (A is not null, $f = x$ and $g = y$ on A) $\Rightarrow$ ($f < g$ *given* $A \Leftrightarrow x < y$);

P4. [($x < y$, $f = y$ on A, $f = x$ on A^c, $g = y$ on B, $g = x$ on B^c) and ($x' < y'$, $f' = y'$ on A, $f' = x'$ on A^c, $g' = y'$ on B, $g' = x'$ on B^c)] $\Rightarrow$ ($f < g \Leftrightarrow f' < g'$);

P5. $x < y$ for some $x, y \in X$;

P6. ($f < g$, $x \in X$) $\Rightarrow$ there is a finite partition of S such that, if A is any event in the partition, then ($f' = x$ on A, $f' = f$ on A^c) $\Rightarrow f' < g$, and ($g' = x$ on A, $g' = g$ on A^c) $\Rightarrow f < g'$;

P7. ($f < g(s)$ *given* A, for all $s \in A$) $\Rightarrow f \leq g$ *given* A. ($g(s) < f$ *given* A, for all $s \in A$) $\Rightarrow g \leq f$ *given* A.

Then, with $<^*$ defined on the set of all subsets of S by

$A <^* B \Leftrightarrow f < g$ whenever ($x < y$, $f = y$ on A, $f = x$ on A^c, $g = y$ on B, $g = x$ on B^c),

there is a unique probability measure P* on the set of all subsets of S that satisfies

$A <^* B \Leftrightarrow$ P*(A) < P*(B) for all $A, B \subseteq S$,

and P* has the property that

($B \subseteq S$, $0 \leq \rho \leq 1$) $\Rightarrow$ P*(C) = ρP*(B) for some $C \subseteq B$,

and, with P* as given, there is a real-valued function U on X for which

$f < g \Leftrightarrow$ E[U(f(s)), P*] < E[U(g(s)), P*] for all $f, g \in F$,

and when U satisfies this it is bounded and unique up to a positive linear transformation.

APPENDIX 2

Dutch book Theorem

The following table summarises the wins and losses for a bet on some proposition *Q*, as per my outline of the betting scenario in Chapter 2, Section 2.2:

Table A-1

Q's truth value	Payoff for *Q*	Payoff against *Q*
True	$(1 - p) \times S$	$-(1 - p) \times S$
False	$-p \times S$	$p \times S$

Here "*p*" represents the betting odds on *Q*, and *S* is some amount of money that serves as the "stakes" for the bet. Either the bettor or the bookie takes the bet "for *Q*" and the other of this pair takes the bet "against *Q*".

Recall from Chapter 2, Section 2.2 that a Dutch book is a set of bets, each of which the agent regards as fair or favourable, which collectively guarantees that they suffer a loss. The Dutch book theorem states that if an agent's fair betting quotients (credences) do not conform to the probability calculus, then there exists a Dutch book against them.

To prove the Dutch book theorem, we need to show that an agent whose fair betting quotients do not satisfy each of the key probability axioms listed below can have a Dutch book made against them:

1. Non-negativity: $\Pr(X) \geq 0$ for all X in **S**.
2. Normalisation: $\Pr(T) = 1$ for any tautology T in **S**.
3. Additivity: $\Pr(X \vee Y) = \Pr(X) + \Pr(Y)$ for all X, Y in **S** such that X is incompatible with Y.

(Here Pr is a probability function over a non-empty set of sentences **S** closed under negation and disjunction.)

Let us consider the probability axioms in turn. For axioms (1) and (2), p is the agent's "fair betting quotient" or credence for some proposition Q. The agent will be prepared to accept any bet that is fair or favourable by the lights of p.[140]

For 1:

Suppose $p < 0$. Then $-p > 0$, $(1 - p) > 1$, and both are positive. Now consider Table A-1. Since both the payoffs $-p \times S$ and $(1 - p) \times S$ are positive, the bookie can guarantee himself a net gain by betting for Q at any positive stake S. Whether Q is true or false, the bookie's payoffs are positive—a Dutch book is made against the agent.

Suppose $p > 1$. Then $1 - p < 0$ and both $-(1 - p)$ and p are positive. So the bookie can make a Dutch book against the agent by betting against Q.

For 2:

Let us suppose that Q is certain and show that if the agent's betting quotient for p is less than 1, a Dutch book can be made against the agent. (We have already shown that a Dutch book can be made against the agent if their betting quotient is greater than 1.)

---

[140] My proofs here, which show that an agent whose credences do not satisfy each of the probability axioms can be Dutch-booked, closely follow those of Resnik (1987).

Since Q is certain, we know that the bottom row of Table A-1 will never apply. So we need only consider the top row. If $p < 1$, then $(1 - p)$ is positive, and so the bookie can guarantee himself a sure gain by betting for Q.

For 3:

Here we assume that the agent's "fair betting quotients" for Q, R and (Q or R) are q, r, and p, respectively. Again, the agent is prepared to accept any bets that are fair or favourable by the lights of these betting quotients.

Table A-2

		Q		R		Q or R	
Q	R	For	Against	For	Against	For	Against
True	True	$1 - q$	$-(1 - q)$	$1 - r$	$-(1 - r)$	$1 - p$	$-(1 - p)$
True	False	$1 - q$	$-(1 - q)$	$-r$	r	$1 - p$	$-(1 - p)$
False	True	$-q$	q	$1 - r$	$-(1 - r)$	$1 - p$	$-(1 - p)$
False	False	$-q$	q	$-r$	r	$-p$	p

We must show that if Q and R are mutually exclusive and $p \neq q + r$, a Dutch book can be made against the agent. Let us assume that Q and R are mutually exclusive. This means that the first row of the betting table never applies and can be ignored. Also assume that $p \neq q + r$. Then either $p < q + r$ or $p > q + r$. Let us consider these two scenarios in turn.

Suppose $p < q + r$. Then $(q + r - p)$ is positive. If the bookie bets against Q and against R, but for (Q or R), his total payoffs for the last three rows of the table all

equal $q + r - p$. Thus by betting as indicated, the bookie can guarantee himself a positive gain no matter what the truth values of Q and R turn out to be.

Suppose $p > q + r$. Then $p - (q + r)$ is positive. If the bookie bets for Q and for R but against (Q or R), his total payoffs for the last three rows of the table all equal $p - q - r$. This value is positive, and so the bookie makes a Dutch book against the agent by betting as indicated.

REFERENCES

Allais, M. 1953. Fondements d'une théorie positive des choix comportant un risque et critique des postulats et axioms de l'école americaine. Econométrie 40:257–332.

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

Armendt, B. 1980. Is There a Dutch Book Argument for Probability Kinematics? Philosophy of Science 47:583–88.

———. 1986. A Foundation for Causal Decision Theory. Topoi 5:3–19.

———. 1992. Dutch Strategies for Diachronic Rules: When Believers See the Sure Loss Coming. PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1992 (Volume One: Contributed Papers):217–229.

———. 1993. Dutch Books, Additivity and Utility Theory. Philosophical Topics 21:1–20.

Arntzenius, F. 2003. Some Problems for Conditionalization and Reflection. Journal of Philosophy C (7):356–370.

Arntzenius, F., A. Elga, and J. Hawthorne. 2004. Bayesianism, Infinite Decisions, and Binding. Mind 113 (450):251–83.

Bandyopadhayay, P. S. 1994. In Search of a Pointless Decision Principle. PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1994 (Volume 1: Contributed Papers):260–9.

Bell, D. E. 1982. Regret in Decision Making Under Uncertainty. Operations Research 30 (5):961–981.

Bolle, F. 1983. On Sen's second-order preferences, morals, and decision theory. Erkenntnis 20 (2):195–206.

Bradley, R. 2005. Radical Probabilism and Bayesian conditioning. Philosophy of Science 72 (2):342–364.

———. 2006. The Kinematics of Belief and Desire.

Broome, J. 1991. Weighing Goods: Equality, Uncertainty and Time. Oxford; Cambridge, MA: Basil Blackwell Ltd.

Christensen, D. 1991. Clever Bookies and Coherent Beliefs. The Philosophical

Review, 229–247.

Colyvan, M. 2004. The Philosophical Significance of Cox's Theorem. International Journal of Approximate Reasoning 37 (1):71–85.

Colyvan, M., D. Cox, and K. Steele. to appear. Modelling the Moral Dimension of Decisions.

(COMEST), World Commission on the Ethics of Scientific Knowledge and Technology. 2005. The Precautionary Principle. Paris: UNESCO.

Cranor, C. F. 2001. Learning from the Law to Address Uncertainty in the Precautionary Principle. Science and Engineering Ethics 7 (3):313–326.

de Finetti, B. 1937. La Prévision: ses lois logiques, ses sources subjectives. Annales de l'Institut Henri Poincaré 7:1–68.

———. 1980 (1937). Foresight: its logical laws, its subjective sources. In Studies in Subjective Probability, edited by H. E. Kyburg and H. Smokler. Huntington, NY: Krieger.

Earman, J. 1992. Bayes or Bust?: A Critical Examination of Bayesian Confirmation Theory. Cambridge, MA; London: MIT Press.

Elga, A. 2000. Self-locating belief and the Sleeping Beauty problem. Analysis 60 (2):143–147.

Ellsberg, D. 1961. Risk, Ambiguity, and the Savage Axioms. The Quarterly Journal of Economics 75 (4):643–669.

Elster, J. 1979. Ulysses and the Sirens. Cambridge: Cambridge University Press.

———. 2000. Ulysses Unbound. Cambridge: Cambridge University Press.

European Commission (EC) Directorate-General for the Environment. 2000. White Paper on environmental liability. Luxembourg: European Commission.

Fishburn, P. 1970. Utility theory for decision making. New York: Wiley.

———. 1981. Subjective Expected Utility, a Review of Normative Theories. Theory and Decision 13:139–199.

Fitelson, B. forthcoming. Likelihoodism, Bayesianism, and Relational Confirmation. Synthese.

Friedman, M., and L. J. Savage. 1948. The Utility Analysis of Choices Involving Risk. The Journal of Political Economy 56 (4):279–304.

Gauthier, D. 1997. Resolute Choice and Rational Deliberation: A Critique and a Defense. Noûs 31 (1):1–25.

Gärdinfors, P., and N. E. Sahlin. 1982. Unreliable Probabilities, Risk Taking and Decision Making. Synthese 53:361–386.

Gillies, D. 2000. Philosophical Theories of Probability. London and New York: Routledge.

Goldstein, M. 1983. The Prevision of a Prevision. Journal of the American Statistical Association 78:817–819.

Hacking, I. 1967. Slightly More Realistic Personal Probability. Philosophy of Science 34:311–325.

Hammond, P. J. 1976. Changing Tastes and Coherent Dynamic Choice. The Review of Economic Studies 43 (1):159–173.

———. 1977. Dynamic Restrictions on Metastatic Choice. Economica 44 (176):337–350.

———. 1988a. Orderly Decision Theory: A Comment on Professor Seidenfeld. Economics and Philosophy 4:292–297.

———. 1988b. Consequentialism and the Independence Axiom. In Risk, Decision and Rationality, edited by B. R. Munier. Dordrecht; Boston: D. Reidel.

———. 1988c. Consequentialist Foundations for Expected Utility Theory. Theory and Decision 25:25–78.

Hájek, A. 2003. Conditional Probability Is the Very Guide of Life. In Probability Is the Very Guide of Life: The Philosophical Uses of Chance, edited by H. E. Kyburg, Jr. and M. Thalos. Chicago and La Salle, Ill.: Open Court.

———. 2005. Scotching Dutch Books? Philosophical Perspectives 19 (1):139–151.

Hájek, A., and L. Ericksson. forthcoming. What are Degrees of Belief? Studia Logica.

Howson, C., and P. Urbach. 1989. Scientific Reasoning: The Bayesian Approach. La Salle, III.: Open Court.

Jeffrey, R. 1965. The Logic of Decision. 1st ed. New York: McGraw-Hill.

———. 1974. Preferences Among Preferences. Journal of Philosophy 71:377-91.

———. 1982. The Sure Thing Principle. PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1982 (Volume Two: Symposia and Invited Papers):719–730.

———. 1983. The Logic of Decision. 2nd ed. Chicago: University of Chicago Press.

Joyce, J. M. 1998. A Nonpragmatic Vindication of Probabilism. Philosophy of Science 65:575–603.

———. 1999. The foundations of causal decision theory. Cambridge; New York: Cambridge University Press.

Kahneman, D., and A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. Econometrica 47 (2):263–292.

———. 1982. On the study of statistical intuitions. Cognition 11 (2):123–141.

Keeney, R. L., and D. von Winterfeldt. 2001. Appraising the precautionary principle—a decision analysis perspective. Journal of Risk Research 4 (2):191–202.

Kyburg, H. E. 1983. Rational Belief. The Behavioral and Brain Sciences 6:231–73.

Leach, J. 1968. Explanation and Value Neutrality. The British Journal for the Philosophy of Science 19 (2):93–108.

Levi, I. 1974. On Indeterminate Probabilities. Journal of Philosophy 71 (13):391–418.

———. 1980. The Enterprise of Knowledge. Cambridge, Mass.: MIT Press.

———. 1985. Imprecision and Indeterminacy in Probability Judgment. Philosophy of Science 52 (3):390–409.

———. 1986. Hard choices: decision making under unresolved conflict. Cambridge; New York: Cambridge University Press.

———. 1987. The Demons of Decision. The Monist 70:193–211.

———. 1991. Consequentialism and Sequential Choice. In Foundations of Decision Theory, edited by M. Bacharach and S. Hurley. Oxford and Cambridge, MA: Basil Blackwell.

———. 1997. The Covenant of Reason. Cambridge: Cambridge University Press.

Lewis, D. 1980. A Subjectivist's Guide to Objective Chance. In Studies in Inductive Logic and Probability, edited by R. Jeffrey. Berkeley: University of California Press.

———. 1981. Causal Decision Theory. Australasian Journal of Philosophy 59 (1):5–30.

Loomes, G., and R. Sugden. 1982. Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. The Economic Journal 92 (368):805–824.

MacCrimmon, K. R., and S. Larsson. 1979. Utility Theory: Axioms versus 'Paradoxes'. In Expected Utility and the Allais Paradox, edited by M. Allais and O. Hagen. Dordrecht; Boston: Reidel Publishing Company.

Machina, M. J. 1982. “Expected Utility” Analysis without the Independence Axiom.

Econometrica 50.

———. 1989. Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty. Journal of Economic Literature 27:1622–1668.

Maher, P. 1992. Diachronic Rationality. Philosophy of Science 59 (1):120–141.

Manson, N. A. 2002. Formulating the Precautionary Principle. Environmental Ethics 24 (3):263–274.

McClennen, E. F. 1988. Ordering and Independence: A Comment on Professor Seidenfeld. Economics and Philosophy 4:298–308.

———. 1990. Rationality and Dynamic Choice: Foundational Explorations. Cambridge: Cambridge University Press.

O'Riordan, T., and A. Jordan. 1995. The Precautionary Principle in Contemporary Environmental Politics. Environmental Values 4:191–212.

Pettit, P. 1991. Decision Theory and Folk Psychology. In Foundations of Decision Theory: Issues and Advances, edited by M. Bacharach and S. Hurley. Oxford: Blackwell.

Putnam, H. 1980. Models and Reality. Journal of Symbolic Logic 45:464-482.

Quiggin, J. 1982. A theory of anticipated utility. Journal of Economic Behavior and Organization 3 (4):323–43.

Rabinowicz, W. 1995. To Have One's Cake and Eat it Too: Sequential Choice and Expected-Utility Violations. Journal of Philosophy 92 (11):586–620.

———. 2000. Preference Stability and Substitution of Indifferents: A Rejoinder to Seidenfeld. Theory and Decision 48:311–318.

Raiffa, H. 1961. Risk, Ambiguity, and the Savage Axioms: Comment. The Quarterly Journal of Economics 75 (4):690–694.

———. 1968. Decision Analysis: Introductory Lectures on Choices under Uncertainty. Reading, Mass.: Addison-Wesley.

Ramsey, F. P. 1950 (1926). Truth and Probability. In Foundations of Mathematics, edited by R. B. Braithwaite. New York: Humanities Press.

Resnik, M. D. 1987. Choices: an introduction to decision theory. Minneapolis: University of Minnesota Press.

Resnik, D. B. 2003. Is the Precautionary Principle Unscientific? Studies in History and Philosophy of Biological and Biomedical Sciences 34C (2):329–344.

Sandin, P., M. Peterson, S. O. Hansson, C. Rudén, and A. Juthe. 2002. Five charges

against the precautionary principle. Journal of Risk Research 5 (4):287–299.

Savage, L. J. 1954. The Foundations of Statistics. New York: Wiley.

Schervish, M. J., Teddy Seidenfeld, and Joseph B. Kadane. 1990. State-Dependent Utilities. Journal of the American Statistical Association 85 (411):840–847.

Schick, F. 1986. Dutch Bookies and Money Pumps. Journal of Philosophy 83:112–119.

Schmeidler, D. 1989. Subjective Probability and Expected Utility Without Additivity. Econometrica 57 (3):571–587.

Seidenfeld, T. 1983. Decisions with Indeterminate Probabilities. The Behavioral and Brain Sciences 6:259–261.

———. 1988a. Decision Theory Without "Independence" or Without "Ordering". Economics and Philosophy 4:267–290.

———. 1988b. Rejoinder [to Hammond and McClennen]. Economics and Philosophy 4:309–315.

———. 1994. When Normal and Extensive Form Decisions Differ. Logic, Methodology and Philosophy of Science IX:451–463.

———. 2000a. Substitution of Indifferent Options at Choice Nodes and Admissibility: A Reply to Rabinowicz. Theory and Decision 48:305–310.

———. 2000b. The Independence Postulate, Hypothetical and Called-off Acts: A Further Reply to Rabinowicz. Theory and Decision 48:319–322.

Seidenfeld, T., J. B. Kadane, and M. J. Schervish. 2004. A Rubinesque theory of decision. IMS Lecture Notes Monograph 45:1–11.

Seidenfeld, T., M. J. Schervish, and J. B. Kadane. 1995. A Representation of Partially Ordered Preferences. The Annals of Statistics 23 (6):2168–2217.

Sen, A. 1977. Rational Fools: A Critique of the Behavioural Foundations of Economic Theory. Philosophy and Public Affairs 6:317–44.

———. 1979. Collective Choice and Social Welfare. Amsterdam; New York: North-Holland.

———. 1985. Rationality and Uncertainty. Theory and Decision 19:109–28.

Shafer, G. 1986. Savage Revisited. Statistical Science 1 (4):463–485.

Skyrms, B. 1993. A Mistake in Dynamic Coherence Arguments? Philosophy of Science 60 (2):320–328.

———. 1984. Pragmatics and Empiricism. New Haven: Yale University Press.

———. 1986. Choice and Chance: an introduction to inductive logic. 3rd ed. Belmont, California: Wadsworth Publishing Company.

———. 1987. Dynamic Coherence and Probability Kinematics. Philosophy of Science 54 (1):1–20.

Sobel, J. H. 1987. Self-Doubts and Dutch Strategies. Australasian Journal of Philosophy 65 (1):56–81.

Steele, K. 2006. The precautionary principle: A new approach to public decision-making? Law, Probability and Risk 5 (1):19–31.

———. 2007. Distinguishing indeterminate belief from 'risk-averse' preferences. Synthese 158 (2):189–205.

Stein, P. 2000. Are Decision-makers too cautious with the Precautionary Principle? Environmental Planning and Law Journal 17 (1):3–23.

Strotz, R. H. 1956. Myopia and Inconsistency in Dynamic Utility Maximisation. Review of Economic Studies 23:165–80.

Teller, P. 1973. Conditionalization and Observation. Synthese 26:218–58.

———. 1976. Conditionalization, Observation, and Change of Preference. In Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, edited by W. Harper and C. Hooker. Dordrecht: D. Reidel.

Tickner, J. A. 2003. Precautionary Assessment: A Framework for Integrating Science, Uncertainty, and Preventive Public Policy. In Precaution: Environmental Science and Preventive Public Policy, edited by J. A. Tickner. Washington DC: Island Press.

Tversky, A. 1975. A Critique of Expected Utility Theory: Descriptive and Normative Considerations. Erkenntnis 9:163–173.

Tversky, A., and D. Kahneman. 1986. Rational Choice and the Framing of Decisions. Journal of Business 59:S251–S278.

———. 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. Journal of Risk and Uncertainty 5:297–323.

van Fraassen, B. C. 1984. Belief and the Will. Journal of Philosophy 81:235–256.

von Neumann, J., and O. Morgenstern. 1944. Theory of Games and Economic Behaviour. Princeton: Princeton University Press.

Walley, P. 1991. Statistical Reasoning with Imprecise Probabilities. London:

Chapman & Hall.

Weatherson, B. 2002. Keynes, Uncertainty, and Interest Rates. Cambridge Journal of Economics 26:47–62.

Weirich, P. 1986. Expected Utility and Risk. British Journal of Philosophy of Science 37 (4):419–42.

———. 2001. Risk's Place in Decision Rules. Synthese 126:427–441.

———. 2004. Realistic Decision Theory. Oxford: Oxford University Press.

Weymark, J. A. 1981. Generalized Gini Inequality Indices. Mathematical Social Sciences 1:409–430.

Wolf, C. 2003. Intergenerational Justice. In A Companion to Applied Ethics, edited by R. G. Frey and C. H. Wellman. Malden, Oxford, Melbourne and Berlin: Blackwell Publishing.

Yaari, M. E. 1987. The Dual Theory of Choice Under Risk. Econometrica 55:95–115.