Preference-Based Belief Dynamics

Extended Abstract

Eric Pacuit* Olivier Roy†
Institute of Logic, Language and Computation
University of Amsterdam

June 15, 2006

1 Introduction

Building on the work of [20] and [6], Savage showed that any agent with a preference ordering satisfying certain intuitive axioms can be represented as an expected utility maximizer [21]. The idea behind Savage's result is to take as primitive an agent's (state-based) preference over a set of prizes and *define* the agent's beliefs and utilities from its preference. Thus properties of an agent's beliefs, represented as subjective probability distributions, are *derived* from properties of the agent's preferences. See, for example, Chapter 1 of [17] for a discussion of the literature on the axiomatic foundations of decision theory. Building on Savage's work and the fundamental contribution by Anscombe and Aumann [1], a number of different belief operators have been proposed in the literature. Ansheim and Sovik provide an excellent survey of these contributions [3].

Other models of beliefs that have been studied in the epistemic logic and economics literature assume that beliefs are defined from an "epistemic possibility" relation. Starting with [13], the idea is to say that an agent believes some proposition E provided E is true in *all* situations the agent currently considers "possible". A notable exception to this approach can be found in a recent paper of Parikh [18] where a notion of a plan is taken as primitive and knowledge is defined relative to the execution of a plan. Another exception can be found in the recent work of Artemov [2] and Fitting [8] where a notion of *evidence* is added to the traditional possible worlds approach. Certainly, this list is not exhaustive; and the reader is referred to [7] for a discussion of alternative models of beliefs and knowledge. Typically these models are presented in the context of the so-called "logical omniscience problem."[1] See [15] for an approach to the logical omniscience problem in the context of decision-theory.

*epacuit@staff.science.uva.nl

†oroy@science.uva.nl

[1]Assuming an agent believes a proposition E provided E is true in all situations the agent considers possible, it is easy to deduce the fact that the agent must also believe all logical consequences of its beliefs and all valid facts (such as, for example, all true statements of arithmetic). The reader is referred to [7] and

In this paper, we study an approach first put forward by Stephen Morris in [16] where an agent's (non-probabilistic) beliefs are *defined* from its preferences. The main idea is to apply the logic of the Savage approach and *deduce* logical properties of an agent's beliefs (such as positive introspection) from properties of the agent's preferences. What interests us for this paper is the relationship between preference change and belief change. Already in [16], Morris showed how a specific kind of belief change can be related to a consistency requirement on preferences in sequential decision making. Indeed a number of authors have focused on the relationship between so-called "coherent" choice and (probabilistic) belief revision operators. See [3] and [19] for a discussion. In this paper we use an approach similar to Morris' in order to characterize various operations on beliefs in terms of preference changes.

One contribution of this paper is to analyze the beliefs operators defined from preferences from a modal logic point of view. Some steps in this direction are already present in [16] and [3], but much more can be said. In particular, recent work on dynamic epistemic logic [26] and dynamic doxastic logic [22] may prove relevant in this context. For example, there are many parallels between Bayesian updating on one hand and epistemic updates on the other hand (cf. [24]).

The paper in organized as follows. In section 2 we present Morris' result connecting static properties of preferences and beliefs, for which we give a new proof. In section 2.2 we turn to preferences dynamics, and generalize Morris' result to characterize belief changes in terms of preference changes. Finally, the last sketch concludes and discusses future work.

2 Defining Beliefs from Preferences

In this section, we describe the basic framework inspired by the work of Morris [16]. We first present a static model, then show how to extend it to a dynamic setting.

2.1 Static Decision Making

Decision theory is often referred to as the study of *rational* decision making under *risk* or *uncertainty*. Typically, "uncertainty" is thought of as *subjective*, that is, referring to the partiality of the decision-maker's information. Whereas "risk" is *objective*, that is, about to the outcome of certain random events such as a lottery. In this paper we are interested in what the agent believes or considers possible. Thus focus on uncertainty leaving aside risk.

Let Ω be a set of states. Typically we assume Ω is finite, although this is not important for our analysis. Any subset of Ω is called an event. Given an event $E \subseteq \Omega$, we say E is true at state $w \in \Omega$ provided $w \in E$. An agent's beliefs is represented by a set-valued **belief function**.

Definition 1 *Let Ω be a set of states. A* belief *operator is a function $B : 2^\Omega \to 2^\Omega$. A belief operator is said to be* normal *iff for each $E, F \subseteq \Omega$*

the two papers mentioned in the paragraph for a discussion.

- $B(E \cap F) = B(E) \cap B(F)$, *and*
- $B(\Omega) = \Omega$.

For any event, $E \subseteq \Omega$, $w \in B(E)$ is intended to mean the "agent believes E at state w." Note that we have opted to represent beliefs purely semantically. Thus we are making essentially use of the fact that the object of beliefs are propositions which can be identified with a set of states. This approach to representing beliefs was first proposed by Aumann [4] in the economics literature and is closely related to the use of Kripke structures for representing beliefs. See [9] for the exact connection between the two approaches.

In this paper, we are interested in how an agent's "doxastic state" changes over time. By a doxastic state, we simply mean the list of events that the agent currently believes. More formally, define a function $\mathsf{b} : \Omega \to 2^{2^\Omega}$ which for each state $w \in \Omega$ returns set of events the agent currently believes. Any function from Ω to 2^{2^Ω} is called a **neighborhood function** (see [5] for a discussion of the use of neighborhood functions in modal logic).

Definition 2 *Given a belief operator B, the neighborhood function based on B $\mathsf{b} : \Omega \to 2^{2^\Omega}$ is defined as follows: for each $w \in \Omega$, $\mathsf{b}(w) = \{E \mid w \in B(E)\}$.*

Conversely, given a neighborhood function $\mathsf{b} : \Omega \to 2^{2^\Omega}$ we can define an agent's belief function $B : 2^\Omega \to 2^\Omega$ as follows: $B(E) = \{w \mid E \in \mathsf{b}(w)\}$. The following property of neighborhood functions will be relevant for our study.

Definition 3 *A neighborhood function b is said to be a* filter *provided for each $w \in \Omega$, $\mathsf{b}(w)$ is a filter. That is for each $w \in \Omega$, $\mathsf{b}(w)$ is*

- *closed under supersets, i.e., if $X \in \mathsf{b}(w)$ and $X \subseteq Y$ then $Y \in \mathsf{b}(w)$,*
- *closed under (finite) intersections, i.e., if $X, Y \in \mathsf{b}(w)$ then $X \cap Y \in \mathsf{b}(w)$, and*
- *contains the unit, i.e., $\Omega \in \mathsf{b}(w)$.*

It is a well-now result, see for example [5], that a belief function is normal iff the corresponding neighborhood function is a filter. This fact will be used below.

Let X be a (finite or infinite) set of outcomes or prizes. A *prize function* is any function $x : \Omega \to X$. We think of a prize function x as an Ω-length vector of elements of X. Thus we write x_w instead of $x(w)$ to refer to the w^{th} component of x, with the intended meaning that the agent receives prize x in state w. Let X^Ω denote the set of all prize functions. At each state, the agent is assumed to have a preference over X^Ω.

Definition 4 *A* state-based preference *is a function $\succeq$: $\Omega \to X^\Omega \times X^\Omega$. We write $\succeq_w$ for $\succeq(w)$ to refer to the decision-maker's preference ordering over X^Ω relative to w. For each $w \in \Omega$, a state-based preference relation $\succeq_w$ is said to be* rational *iff it is*

- reflexive *i.e. for all $x \in X^\Omega$, $x \succeq_w x$.*

- complete*, that is, for all* $x, y \in X^\Omega$*, either* $x \succeq_w y$ *or* $y \succeq_w x$ *and*
- transitive *i.e. for all* $x, y, z \in X^\Omega$*, if* $x \succeq_w y$ *and* $y \succeq_w z$ *then* $x \succeq_w z$*.*

The agent is said to have rational preference relations if all his state-based preference relations are rational.

Intuitively, a prize vector is a complete description of which prizes are awarded in which state. The state-based preference relations order these *whole* prizes vector, and not only their state-components. So, for a prize functions $x, y \in X^\Omega$, $x \succeq_w y$ is intended to mean that the agent prefers the prize distribution of vector x over the one of vector y at state w. Indifference (denoted $x \sim_w y$) and strict preference (denoted $x \succ_w y$) are respectively defined as $x \succeq_w y$ & $y \succeq_w x$ and $x \succeq_w y$ & $\neg y \succeq_w x$.

For two prize functions (vectors) x and y and an event $E \subseteq \Omega$, we define the prize function (x_E, y_{-E}) as follows

$$(x_E, y_{-E})(w) = \begin{cases} x_w & w \in E \\ y_w & w \notin E \end{cases}$$

In what follows, we generalize the above notation to an arbitrary collection of sets that partition Ω. That is, suppose that $\{E_1, \ldots, E_n\}$ is a pairwise disjoint collection of sets and $E_1 \cup E_2 \cup \cdots \cup E_n = \Omega$. Then the prize function $(x^1_{E_1}, \ldots, x^n_{E_n})$ is $x^1(w)$ for $w \in E_1$, $x^2(w)$ for $w \in E_2$,... and $x^n(w)$ for $w \in E_n$. The basic idea behind Savage's approach of defining beliefs from preferences is to say that an agent believes that the event E occurred if eventualities outside of E have no influence on his preferences — he does not care what prize he would receive if E was not the case. The following definition, taken from [16], is intended to capture this idea.

Definition 5 *A belief operator* B *is said to* reflect *the agents preferences iff*

$$B(E) = \{w \mid \textit{for each prize function } x, y \textit{ and } z,\ (x_E, y_{-E}) \sim_w (x_E, z_{-E})\}$$

Intuitively, if the agent has the preference $(x_E, y_{-E}) \sim_w (x_E, z_{-E})$ then either the agent is indifferent between prizes y and z or the agent is sure that the actual state is in E and so the prizes offered outside of E do not matter. Now for a fixed E, if the agent has this preference *for all* prize functions x, y and z, then it is natural to say that the agent believes E.

Note that being a normal belief operator and a rational preference ordering are both *static* features — they do not place any constraints on how beliefs or preferences can change over time. The first result of [16] connects these two properties. Our proof is similar to Morris' proof, but adapted to our setting.

Theorem 1 *If the preferences of an agent are rational then any belief operator that reflects these preferences is normal.*

Proof Let $\succeq$ be a rational preference and suppose that the belief function B reflects $\succeq$. We show that the neighborhood function generated from B must be a filter. Suppose that $E \in \mathsf{b}(w)$ and $E \subseteq F$. We must show $F \in \mathsf{b}(w)$. We first note that given any two prize function x and y, since $E \subseteq F$, $(x_F, y_{-F}) = (x_E, x_{F-E}, y_{-F})$. Suppose that $F \notin \mathsf{b}(w)$. Then since B reflects $\succeq$, there are prizes a, b and c such that $(a_F, b_{-F}) \not\sim_w (a_F, c_{-F})$. Since $\succeq$ is complete, either $(a_F, b_{-F}) \succ_w (a_F, c_{-F})$ or $(a_F, c_{-F}) \succ_w (a_F, b_{-F})$. Without loss of generality, say that $(a_F, b_{-F}) \succ_w (a_F, c_{-F})$. As noted above, this is equivalent to saying that $(a_E, a_{F-E}, b_{-F}) \succ_w (a_E, a_{F-E}, c_{-F})$. Choose an arbitrary prize function x and define b' to be the prize (x_E, a_{F-E}, b_{-F}) and c' to be the prize (x_E, a_{F-E}, c_{-F}). Then $(a_E, b'_{-E}) \succ_w (a_E, c'_{-E})$. But this contradicts the fact that $E \in \mathsf{b}(w)$. So, $\mathsf{b}(w)$ is closed under supersets.

Suppose that $E \in \mathsf{b}(w)$ and $F \in \mathsf{b}(w)$. We must show $E \cap F \in \mathsf{b}(w)$. Let x, y and z be arbitrary prize functions. First note that $(x_{E\cap F}, y_{-(E\cap F)})$ can be written as $(x_{E\cap F}, y_{E-F}, y_{-E})$. Then, since $E \in \mathsf{b}(w)$, $(x_{E\cap F}, y_{E-F}, y_{-E}) \sim_w (x_{E\cap F}, y_{E-F}, z_{-E})$. Similarly, $(x_{E\cap F}, y_{E-F}, z_{-E})$ can be written as $(x_{E\cap F}, z_{F-E}, y_{E-F}, z_{-(E\cup F)})$. Since $F \in \mathsf{b}(w)$, $(x_{E\cap F}, z_{F-E}, y_{E-F}, z_{-(E\cup F)}) \sim_w (x_{E\cap F}, z_{F-E}, z_{E-F}, z_{-(E\cup F)})$. Since $\succeq_w$ is transitive and hence $\sim_w$ is transitive,

$$(x_{E\cap F}, y_{-(E\cap F)}) = (x_{E\cap F}, y_{E-F}, y_{-E}) \sim_w (x_{E\cap F}, y_{E-F}, z_{-E})$$

$$= (x_{E\cap F}, z_{F-E}, y_{E-F}, z_{-(E\cup F)}) \sim_w (x_{E\cap F}, z_{F-E}, z_{E-F}, z_{-(E\cup F)}) = (x_{E\cap F}, z_{-(E\cap F)})$$

Hence, by transitivity of $\sim_w$, $(x_{E\cap F}, y_{-(E\cap F)}) \sim_w (x_{E\cap F}, z_{-(E\cap F)})$ and so $E \cap F \in \mathsf{b}(w)$.

Finally, we note that by completeness, for each prize function x, $x \sim_w x$. Hence for all x, y and z, $(x_\Omega, y_{-\Omega}) \sim_w (x_\Omega, z_{-\Omega})$. Therefore, for all w, $\Omega \in \mathsf{b}(w)$.

□

Morris extends this analysis to other properties of belief operators such as the knowledge property (for all E, $B(E) \subseteq E$) and positive introspection (for all events E, $B(E) \subseteq B(B(E))$). We will discuss these and related results in the full version of the paper. Instead we turn our attention to preference dynamics.

2.2 Sequential choices and preference dynamics

In *dynamic* choice contexts, an agent makes a number of different decisions. It might well be that the agent's preferences *change*, in the face of new information, for example. An agent might prefer red over white wine as long as he is uncertain about the dinner menu, but change his preference when he learn that fish will be served. One important issue in dynamic choice theory, e.g. in [23] and [11], is to identify the properties of preference change that lead to coherent behavior.

The preference-based perspective on beliefs sketched above adds a new twist to these issues. As preferences change during a sequential decision making situation, so will the corresponding beliefs of the agent. This raises some interesting questions. First, suppose an agent has preferences that lead to coherent behavior, what sort of belief change reflects

these preferences? Second, and perhaps more interesting, given a particular theory of belief change, such as the AGM paradigm, what kind of properties does this theory impose on how preferences may change?

The are a number of approaches one can take in order to add dynamics to the above models. Perhaps the most obvious is to add time stamps and study how the agent's preferences (and hence beliefs) change over time in the presence of new information. A second approach is to fix the model and study how the agent's preferences are related between states. This approach will be sketched below, although in the full version of the paper both will be discussed.

A decision problem for an agent is a finite set of prizes $D \subseteq X^{\Omega}$. The **choice** set of the agent in state w is the set $C_w[D] = \{x \in D \mid \forall y \in D \;\; x \succeq_w y\}$. Intuitively, if the agent is faced with a decision problem D, then $C_w[D]$ is the set of prize functions that the agent would choose in state w. What we are interested in is the choice set of a fixed decision problem D as the agent's preferences change. In this setting, a change in preference is simply represented by a movement from a state $w \in \Omega$ to another state $w' \in \Omega$. A **dynamic decision problem** is a finite set D of prizes and a relation $R \subseteq \Omega \times \Omega$. Since the agent's preferences may be different in different state w and w', the corresponding choice sets $C_w[D]$ and $C_{w'}[D]$ may be different.

Definition 6 *A preference change wRw' is* $\uparrow$-coherent *if $\succeq_w \subseteq \succeq_{w'}$, and* $\downarrow$-coherent *if $\succeq_{w'} \subseteq \succeq_w$.*

Theorem 2 *For all $w \in \Omega$:*

1. *If R is $\uparrow$-coherent then for each $w, w' \in \Omega$ with wRw' we have $\mathsf{b}(w) \subseteq \mathsf{b}(w')$.*
2. *If R is $\downarrow$-coherent then for each $w, w' \in \Omega$ with wRw' we have $\mathsf{b}(w') \subseteq \mathsf{b}(w)$.*

Proof The proof of both items follows the same lines. We will only prove the first. Assume R is $\uparrow$-coherent and take $w, w' \in \Omega$ with wRw' and suppose $E \in \mathsf{b}(w)$. That means that for all x, y and z, $(x_E, y_{-E}) \sim_w (x_E, z_{-E})$. But since R is $\uparrow$-coherent, we know that $\succeq_w \subseteq \succeq_{S(w)}$, which means that for all x, y and z, $(x_E, y_{-E}) \sim_{w'} (x_E, z_{-E})$, as required.

□

This means that $\uparrow$-coherency restricts belief change to expansion, while $\downarrow$-coherency restricts change to contraction. The full version of the paper will extend this analysis to belief *revision*. That is we are interested in which properties of R correspond to a revision of beliefs.

3 Conclusion and Future Work

This paper is concerned with characterizing static and dynamic properties of beliefs in terms of preferences. This approach of doxastic modeling has its root in decision theory, but our main source was the (non-probabilistic) work of [16]. Although he didn't mentioned it

explicitly, Morris has shown that there is a tight connection between some kind of coherency of belief change and what is called "belief extension" in AGM. Here we have pushed this work one step further by showing which kind of preference changes characterize revision and contraction of beliefs. Still, many questions remain open for future work. Let us mention two.

Can we provide such a characterization under *partial* preference orderings? It seems very natural to us to assume that some options are *incomparable* for an agent. For example, someone might be unable to choose between political career and family life. But the reader might have noticed that completeness of the preference ordering was a *crucial* characteristic in all the characterization arguments given above. Works on representation of incomplete preference ordering exist in decision theory (e.g. [14]), but the connection with belief changes still have to be done, as far as we know.

What is the *logic* of preference-based belief changes? Our work is at the level of models. But most of the work in belief revision theory deals with beliefs states viewed as *syntactic theories*, see for example [22]. Also, work in *preference* (for example, see [12], [10], and [25]) and *epistemic* logic have a strong syntactic side. One should be able to use tools from these communities to axiomatize the logic of preference-based belief change.

References

[1] Anscombe, F. J., and Aumann, R. A definition of subjective probability. *Annals Math. Stat. 34* (1963), 199–205.

[2] Artemov, S., and Nogina, E. On epistemic logic with justification. In *Theoretical Aspects of Rationality and Knowledge. Proceedings of the Tenth Conference (TARK 2005)* (2005), R. van der Meyden, Ed.

[3] Asheim, G., and Sovik, Y. Preferenc-based belief operators. *Mathematical Social Sciences 50* (2005), 61 – 82.

[4] Aumann, R. Interactive epistemology I: Knowledge. *International Journal of Game Theory 28* (1999), 263–300.

[5] Chellas, B. F. *Modal Logic: An Introduction.* Cambridge University Press, Cambridge, 1980.

[6] de Finetti, B. La prevision: Ses lois logiques, ses sources subjectives. In *Annales de l'Institut Henri Poincare 7*. Paris, 1937, pp. 1–68. Translated into English by Henry E. Kyburg Jr., Foresight: Its Logical Laws, its Subjective Sources. In Henry E. Kyburg Jr. and Howard E. Smokler (1964, Eds.), Studies in Subjective Probability, 53-118, Wiley, New York.

[7] Fagin, R., Halpern, J., Moses, Y., and Vardi, M. *Reasoning about Knowledge.* The MIT Press, Boston, 1995.

[8] Fitting, M. A logic of explict knowledge. In *The Logica Yearbook 2004*, L. Behounek and M. Bilkova, Eds. 2005, pp. 11–22.

[9] Halpern, J. Set-theoretic completeness for epistemic and conditional logic. *Annals of Mathematics and Artificial Intelligence* (1999).

[10] Halpern, J. Y. Defining relative likelihood in partially-ordered preferential structure. *Journal of Artificial Intelligence Research 7* (1997), 1–24.

[11] Hammond, P. J. Changing tastes and coherent dynamic choice. *The Review of Economic Studies 43*, 1 (Feb. 1976), 159–173.

[12] Hansson, S. O. Preference logic. In *Handbook of Philosophical Logic (Second Edition)*, D. Gabbay and F. Guenthner, Eds., vol. 4. Kluwer, 2001, ch. 4, pp. 319–393.

[13] Hintikka, J. *Knowledge and Belief: An Introduction to the Logic of Two Notions*. Cornell University Press, Ithaca, N.Y., 1962.

[14] Levi, I. *Hard Choices: Decision Making Under Unresolved Conflict*. Cambridge University Press, 1986.

[15] Lipman, B. Decision theory without logical omniscience: Toward an axiomatic framework for bounded rationality. *Review of Economic Studies* (1999).

[16] Morris, S. The logic of belief and belief change: A decision theoretic approach. *Journal of Economic Theory 29*, 1 (April 1996), 1–23.

[17] Myerson, R. B. *Game Theory: Analysis of Conflict*. Harvard University Press, 1991.

[18] Parikh, R. What do we know, and what do we know. In *Theoretical Aspects of Rationality and Knowledge. Proceedings of the Tenth Conference (TARK 2005)* (2005), R. van der Meyden, Ed.

[19] Perea, A. A model of minimal probabilistic belief revision. Available at the authors website, 2006.

[20] Ramsey, F. P. Truth and probability. In *The Foundations of Mathematics and other Logical Essays*, R. B. Braithwaite, Ed. 1926.

[21] Savage, L. *The Foundations of Statistics*. J. Wiley, New York, 1954. second revised edition, 1972.

[22] Segerberg, K. Belief revision from the point of view of doxastic logic. *Bulletin of the IGPL 3*, 4 (1995), 535–553.

[23] Strotz, R. H. Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies 23*, 3 (1955 - 1956), 165–180.

[24] van Benthem, J. Conditional probability meets update logic. *Journal of Logic, Language and Information 12*, 4 (2003), 409 – 421.

[25] van Benthem, J., and Liu, F. The dynamics of preference upgrade. *Journal of Applied Non-Classical Logics* (2006).

[26] van Ditmarsch, H., van der Hoek, W., and Kooi, B. *Dynamic Epistemic Logic*. Cambridge University Press, forthcoming.