A Derivation of Quasi-Bayesian Theory

Fabio Cozman
Robotics Institute, Carnegie Mellon University
fgcozman@cs.cmu.edu — http://www.cs.cmu.edu/~fgcozman/home.html

CMU-RI-TR 97-37

The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

October 13, 1997

Abstract

This report presents a concise and complete theory of convex sets of distributions, which extends and unifies previous approaches. Lower expectations and convex sets of probability distributions are derived from axioms of preference; concepts of conditionalization, independence and conditional independence are defined based on convex sets of distributions.

This research is supported in part by NASA under Grant NAGW-1175. Fabio Cozman was supported under a scholarship from CNPq, Brazil.

1 Introduction

A variety of approaches for decision-making deal with interval-valued inferences. Researchers have investigated the properties of inner/outer measures [19, 23, 44, 57], and lower probability [4, 9, 14, 28, 56] for evaluating and selecting courses of action; Dempster-Shafer theory employs belief and plausibility functions [44, 49] to represent interval-valued "beliefs" in events.

Several authors advocate the use of *convex sets of distributions* as a flexible, meaningful, and realistic model for decision-making and inference [20, 32, 47, 58]. Unfortunately, results concerning convex set of distributions are scattered in a variety of fields, and are generally obscured by non-uniform terminology and notation. Even worse, several contradictory proposals exist for conditionalization and independence relations in convex sets of distributions.

This report presents a concise and complete theory of convex sets of probability distributions for finite models. The main objective is to update *Quasi-Bayesian theory*, originally proposed by Giron and Rios [18], by introducing appropriate concepts of irrelevance and independence. The derivation of main results is presented in a novel form, transitioning from preferences to lower expectations and then to convex sets of distributions; in this process, equivalence of Giron and Rios' axioms and Walley's almost-preference axioms is proved. Walley's generalized Bayes rule is also proved in a novel form, which emphasizes the concept of conditional preferences.

The result is theory of decision-making and inference that is simple and intuitive, and yet can handle the finite models that are commonplace in Artificial Intelligence.

2 Paper roadmap

Section 3 classifies and reviews modifications to Bayesian theory, providing background and discussing related work. A brief overview of Quasi-Bayesian theory is then presented in section 4.

Section 5 presents a complete derivation of Quasi-Bayesian theory, including irrelevance and independence relations. Some technical aspects of Quasi-Bayesian theory are presented in sections 6 (lower and upper values) and 7 (decision-making).

Appendix A contains the proof for theorem 1; all other proofs are collected in Appendix B.

3 Relaxing Bayesian theory

One of the most successful theories of decision-making is *Bayesian* theory [2, 13, 15, 39]. Consider a finite set of states $\{\theta_1, \theta_2, \ldots, \theta_n\}$. A *decision* is a function f that assigns a utility value for each possible state of the world. The idea is that, if decision f is selected and state θ obtains, the decision-maker gets $f(\theta)$.

Bayesian theory postulates that a decision f is evaluated through its *expected value* $E_p[f]$ with respect to a single probability distribution $p(\theta)$:

$$E_p[f] = \sum_j f(\theta_j)p(\theta_j). \tag{1}$$

In the real world we can rarely meet all the assumptions of a Bayesian model. First, we have to face imperfections in a decision-maker's beliefs, either because the decision-maker has no time, resources, patience, or confidence to provide exact probability values. Second, we may deal with a group of disagreeing experts, each specifying a particular distribution [32]. Third, we may be interested in abstracting away parts of a model and assessing the effects of this abstraction [8, 21]. For example, in the model of figure 1, a

decision-maker may want to assess the impact of the link between variables X_1 and X_2, or the impact of merging variables X_3 and X_4 into a single variable.

There are two types of models that violate the Bayesian model:

Resource-bounded decision-makers. The first type of model imposes constraints on the decision-maker. For example, the decision-maker may have to pay for the cost of computation [19, 29, 35, 46], or the decision-maker may be handicapped (e.g., have limited memory) [16, 25, 26]. The term *bounded* rationality was coined by H. Simon to represent this situation [52, 54]. Simon proposed that *heuristic* knowledge is the best way to cope with boundedness [55, 53]. Recently, researchers have tried to apply standard decision theory to the task of evaluating the cost of deliberation [24, 27, 38, 41, 42, 45].

Imperfect decision-makers. The second type of model relaxes the requirements of Bayesian theory; decision-makers are *imperfect* [3, 6, 17, 22, 32, 48, 51, 56]. For example, one may adopt interval representations of uncertainty [4, 37], or belief functions as in Dempster-Shafer theory [12, 50]. Complaints about the excessive degree of detail demanded by the elicitation of a single probability distribution have been common in Artificial Intelligence and Robotics (Cheeseman [5] discusses most of these criticisms).

Work reported in this paper focuses on imperfect decision-makers. Fundamentally,

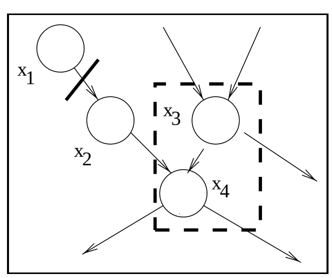

Figure 1: Abstraction in Bayesian networks: elimination of weak links and grouping of variables.

imperfect rationality gives additional freedom to a decision-maker, while bounded rationality subtracts some of the usual freedom by imposing extra costs. Whereas bounded rationality focuses on computational gains, imperfect rationality focuses on more realistic and interesting decision processes.

4 Overview of Quasi-Bayesian theory

Bayesian decision-makers can make an ordering for all possible decisions using expected values. A natural relaxation is to consider the possibility of a partial ordering for decisions.

Quasi-Bayesian theory is a generalization of Bayesian theory which starts from a partial order of preferences. A decision-maker reasons with a convex, closed set of distributions[1] and a single utility function. The main practical consequence of this model is that every event is associated with a probability interval, and every decision is associated with an expected utility interval.

A convex, closed set of distributions maintained by a decision-maker is called a *credal set*. To simplify terminology, use the term credal set *only* in connection to a set of distributions containing more than one element. The credal set for a variable X is denoted by $K(X)$.

Inference is performed by applying Bayes rule to each distribution in a joint credal set; the posterior credal set is the union of all such posterior distributions. To obtain maximum and minimum values of posterior probabilities, one must look only at the vertices of the posterior credal sets, which are obtained by applying Bayes rule to the vertices of a joint credal [58].

[1] Given a set of functions p_i, their convex combination is determined by $\sum_i \alpha_i p_i$, where all α_i are positive and $\sum \alpha_i = 1$. A convex set of probability distributions is a set of distributions that is closed under convex combinations.

Consider a variable Y with a set of values $\hat{Y}$, and denote by $\mathcal{Y}$ the set of all subsets of $\hat{Y}$. For any event A in $\mathcal{Y}$, take the conditional credal set $K(X|A)$ for a variable X. Denote the collection of all such $K(X|A)$ by $K(X|Y)$.

Giron and Rios do not discuss independence relations in their formulation; I propose to adopt Walley's definition of independence, which is based on the concept of irrelevance [58]. Variable X is *irrelevant* to variable Y when the sets $K(X|Y)$ and $K(X)$ contain the same distributions. Variables X and Y are independent when X is irrelevant to Y and vice-versa.

A credal set can be interpreted as an *incomplete* or as an *exhaustive* model of beliefs. In the first case, it is assumed that the decision-maker maintains a single probabilistic model, but resource constraints do not allow the decision-maker to fully recognize or use this probabilistic model; this is the "black-box" model by Good [19]. In the second case, it is assumed that a credal set is a complete model in itself, and the decision-maker has no commitment to any single underlying "true" distribution. In this paper, credal sets are taken as exhaustive models, with not requirement that a decision-maker adopt any single distribution in a credal set.

5 A derivation of Quasi-Bayesian theory

A number of axiomatizations of convex sets of distributions are available in different fields. The present derivation contributes Theorem 1, which proves the equivalence of Giron and Rios' approach and Walley's almost-equivalence axioms. This point was left open by Walley [58, note 3.7.7]. Theorems 2, 3, 4 and 7 are well-known results based on Walley's *correspondence theorem* [58, section 3.8.1]. Theorems 5 and 6 are also adaptations of known results to the terminology of Giron and Rios' axioms. Since Giron and Rios did not publish proofs for their theorems, proofs for theorems 2 to 6 are included in Appendix B. Finally, theorems 8 and 9 are fundamental results that come directly from Walley's book, and are reproduced without proof.

The ideas behind Quasi-Bayesian theory have been proposed by a number of authors, notably by Levi and Kyburg. Levi's considerations about probability are identical to Giron and Rios', but Levi also admits sets of utility functions [32]. Kyburg has proposed a variety of approaches to uncertainty, some of which employ convex sets of distributions [20]. The most comprehensive study of inferences based on partial orders has been conducted by Walley [58], but his axioms and techniques start from very different premises; Walley focuses on lower expectations as the main entity in inferences. The most general theory of decision for finite spaces is due to Seidenfeld et al. [47], who provide a complete and elegant axiomatization for sets of probabilities *and* utilities. Future research must consider the implications of sets of utilities.

The strategy here is to start from preferences, derive lower expectations and then obtain convex sets of distributions. Most of the derivation follows from results in Giron and Rios original work [18] and Walley's book [58]. The presentation is more complete than Giron and Rios' (they do not discuss expectations nor independence relations) and more intuitive than Walley's (he starts from lower expectations in very general spaces, and his basic coherence axioms are harder to motivate than Giron and Rios' preference axioms).

5.1 Conventions

Functions are real-valued, defined over a finite universe $\{\theta_1, \theta_2, \ldots, \theta_n\}$, and form a linear space. The *indicator* function for event A, denoted by $\delta_A(X)$, is equal to 1 if $X \in A$ and 0 otherwise. Constant functions, indicator functions and all necessary convex combinations of functions are included in the space of functions. Functions are denoted by $f, g, h, \ldots$ (possibly subscripted). Variables are denoted by $X, Y, Z, \ldots$ (possibly subscripted), and have a finite number of values. The set of values for variable X is denoted $\hat{X}$, and the set of all subsets of $\hat{X}$ is denoted $\mathcal{X}$. Sets and events are denoted $A, B, C, \ldots$ (possibly subscripted). Write $f > g$ when $f(\theta) > g(\theta)$ for all θ.

5.2 Preferences

Quasi-Bayesian theory, like Bayesian theory, assumes the existence of a utility function. A decision is a function f that assigns a utility value for each possible state of the world. The key problem is how to compare decisions. A preference pattern must be defined so that the decision-maker can compare functions.

The axioms below are valid for a preference relation $\succeq$ defined for pairs of functions. The statement $f \succeq g$ means f *is at least as preferred as* g. To simplify notation, define the strict preference relation $\succ$ by: $f \succ g$ if and only if $f \succeq g$ and not $g \succeq f$.

Giron and Rios' axioms are [18]:

1. If $f \succeq g$ and $g \succeq h$ then $f \succeq h$. [transitivity]
2. If $f > g$ then $f \succ g$. [dominance]
3. For $\lambda \in (0, 1]$, $f \succeq g$ if and only if $\lambda f + (1 - \lambda)h \succeq \lambda g + (1 - \lambda)h$. [convex combination]
4. If $f_i \to f$ and $g \succeq f_i \succeq h$ for all i, then $g \succeq f \succeq h$. [convergence]

These axioms are similar to axioms proposed by Walley [58, section 3.7.6]; Walley indicates that his axioms are "apparently equivalent" to Giron and Rios' axioms [58, note 3.7.7]. The following theorem proves the equivalence of the two systems.

Theorem 1. *Giron and Rios' axioms are equivalent to the following axioms:*

1. *If* $f = -1$ *and* $g = 0$, *it is not the case that* $f \succeq g$. *[sure gain]*
2. *If* $f \succeq g$ *and* $g \succeq h$ *then* $f \succeq h$. *[transitivity]*
3. *If* $f \geq g$ *then* $f \succeq g$. *[monotonicity]*
4. *If* $f \succeq g$ *and* $\lambda > 0$ *then* $\lambda f \succeq \lambda g$. *[positive homogeneity]*

5. *If* $f + \lambda \succeq g$ *for all* $\lambda > 0$ *then* $f \succeq g$. *[continuity]*

6. $f \succeq g$ *if and only if* $f \Leftrightarrow g \succeq 0$. *[cancellation]*

5.3 Lower expectations

To investigate the consequences of the axioms, define a functional $\underline{E}[f]$, called the *lower expectation* of function f:

$$\underline{E}[f] = \max_{\mu} [f \succeq \mu] . \tag{2}$$

Theorem 2. *Lower expectations have the following properties:*

1. $\underline{E}[f] \geq \inf f$.
2. $\underline{E}[\lambda f] = \lambda \underline{E}[f]$ *for* $\lambda > 0$. *[positive homogeneity]*
3. $\underline{E}[f + g] \geq \underline{E}[f] + \underline{E}[g]$. *[super-additivity]*

Lower expectations have a direct correspondence to preference patterns, as proved by the following theorem.

Theorem 3. $\underline{E}[f \Leftrightarrow g] \geq 0$ *if and only if* $f \succeq g$.

This result shows that preferences can be obtained from lower expectations [58, section 3.8.1]:

$$f \succeq g \text{ when } \underline{E}[f \Leftrightarrow g] \geq 0. \tag{3}$$

The concept of lower expectation is not just a derivative of Giron and Rios' axioms; properties in theorem 2 define all functionals that can possibly generate valid preference patterns.

Theorem 4. *Any functional satisfying the properties in theorem 2 generates a preference pattern (through expression (3)) satisfying Giron and Rios' axioms.*

5.4 Convex sets of distributions

The representation of preference patterns by lower expectation demands specification of a numeric value for every function of interest. A more compact representation is in terms of convex sets of distributions.

Call $E_p[f] = \sum_j f(\theta_j)p(\theta_j)$ the *expectation* of f with respect to distribution p (expression 1). An expectation E_p *dominates* a lower expectation $\underline{E}$ if $E[f] \geq \underline{E}[f]$ for all f.

Consider:

- For a lower expectation $\underline{E}$, the set K of all *dominating distributions*:

$$K = \{p(\theta) : \underline{E}[f] \geq E_p[f] \text{ for all } f\}.$$

- For a closed convex set of distributions K, the *dominated lower expectation* $\underline{E}'$:

$$\underline{E}'[f] = \min_{p \in K} E_p[f] \tag{4}$$

Suppose one starts with lower expectation $\underline{E}$, calculates the set of dominating distributions K for $\underline{E}$, and then calculates the dominated lower expectation $\underline{E}'$ for K. $\underline{E}'$ is always equal to $\underline{E}$; a lower expectation $\underline{E}$ is *represented* by its set of dominating distributions K [28].

Theorem 5. *A functional $\underline{E}$ is a lower expectation if and only if it is represented by its set K of dominating distributions.*

This result leads to the basic theorem of Quasi-Bayesian theory [18, theorem 3.2].

Theorem 6. *For a preference pattern satisfying Giron and Rios' axioms, there is a closed and convex set of probability distributions K such that:*

$$f \succeq g \Leftrightarrow E_p[f] \geq E_p[g] \textit{ for every } p \in K.$$

5.5 Conditional preferences

For any event A, the expression $f \succeq^A g$ means f *is at least as preferred as* g *when event* A *obtains*. The purpose of this section is to formalize this concept.

For any function f, define the auxiliary function f^A:

$$f^A(\theta_i) = \begin{cases} f(\theta_j) & \text{if} \quad \theta_j \in A, \\ \mu & \text{if} \quad \theta_j \notin A, \end{cases}$$

where μ is an arbitrary constant.

Giron and Rios take the conditional preference $f \succeq^A g$ to mean that $f^A \succeq g^A$; comparisons among decisions given A should only focus on the preferences restricted to states in A.

The conditional preference $f \succeq^A g$ is equivalent to the statement that $\delta_A f \succeq \delta_A g$:

$$f^A \succeq g^A \quad \Leftrightarrow \delta_A f + (1 \Leftrightarrow \delta_A)\mu \succeq \delta_A g + (1 \Leftrightarrow \delta_A)\mu \quad \Leftrightarrow \quad \delta_A f \succeq \delta_A g.$$

As a consequence, $\succeq^A$ satisfies all preference axioms for any A such that $\underline{E}[\delta_A] \geq 0$. An event A such that $\underline{E}[\delta_A] = 0$ is called a *null* event.

For a non-null event A, define the *conditional lower expectation given* A:

$$\underline{E}[f|A] = \max_{\mu} \left[f \succeq^A \mu \right],$$

which leads to:

$$\underline{E}[f|A] = \max_{\mu} \left[\delta_A f \succeq \delta_A \mu \right] = \max_{\mu} \left[(f \Leftrightarrow \mu)\delta_A \succeq 0 \right]. \tag{5}$$

I adopt the convention that the preference pattern $\succeq^A$ for a null event A is vacuous; when A is null, $\underline{E}[f|A] = \inf(f)$.

The functionals $\underline{E}[f|A]$ and $\underline{E}[f]$ are closely related by the *generalized Bayes rule* (first proposed by Walley [58, section 6.4.1]):

Theorem 7. *For a non-null event A, $\underline{E}\,[f|A]$ is the only solution of the equation:*

$$\underline{E}\,[(f \Leftrightarrow \lambda)\,\delta_A] = 0. \tag{6}$$

From the previous theorem, the following important result can be derived [58, section 6.4.2]:

Theorem 8. *For a non-null event A and for a lower expectation $\underline{E}$ with set of dominating distributions K:*

$$\underline{E}\,[f|A] = \min_{p \in K} E_p[X|A],$$

where $E_p[f|A]$ is the conditional expectation determined by Bayes rule:

$$E_p[f|A] = \frac{E_p[\delta_A f]}{E_p[\delta_A]}.$$

For each function f, $\underline{E}\,[f|A]$ is attained at one of the extreme points of the set K.

In many situations, a joint credal set is not directly specified; instead, marginal and conditional credal sets are provided. The following result is important in this situation.

Consider variables X and Y with sets of values $\hat{X}$ and $\hat{Y}$, and take $\mathcal{X}$ and $\mathcal{Y}$ to be the sets of subsets of $\hat{X}$ and $\hat{Y}$ respectively. Suppose a marginal credal set $K(X)$ and a collection of conditional credal sets $K(X|Y)$ are defined. Define the *lower likelihood* $L(A|Y) = \underline{p}\,(A|Y)$ and the *upper likelihood* $U(A|Y) = \overline{p}\,(A|Y)$.

The following result shows that the lower and upper likelihoods summarize all the information in $K(X|Y)$ [58, section 8.5.3].

Theorem 9. *Consider a function $f(Y)$ defined in $\hat{Y}$ and a non-null event A in $\mathcal{X}$. There is a number λ such that the value of $\underline{E}\,[f|A]$ is attained when $p(A|Y)$ is:*

$$p(A|Y) = \begin{cases} U(A|Y) & \text{if} \quad f(Y) < \lambda, \\ L(A|Y) & \text{if} \quad f(Y) \geq \lambda. \end{cases}$$

5.6 Independence relations

The standard definition of probabilistic independence[2] cannot be easily generalized to convex sets of distributions. Consider the following example.

Example 1. *Take binary variables X and Y, and the credal sets:*

$$p(X = 0) = 0.2\alpha + 0.8(1 \Leftrightarrow \alpha), \qquad p(Y = 0) = 0.3\beta + 0.7(1 \Leftrightarrow \beta),$$

for $\alpha, \beta \in [0, 1]$. Now generate the set of all joint distributions $K'(XY)$ such that $p(XY) = p(X)p(Y)$ for any $p(XY) \in K'(XY)$. The set $K'(XY)$ is not convex. For example, $p(X = 0, Y = 0) = 0.56 \Leftrightarrow 0.74\alpha + 0.24\alpha^2$ when $\alpha = \beta$. This function is equal to 0.56 when $\alpha = 0$ and 0.06 when $\alpha = 1$, but it is strictly smaller than the convex combination of extreme points $0.56\alpha + 0.06(1 \Leftrightarrow \alpha)$ for $\alpha \in (0, 1)$.

One solution is to drop convexity [30], which would violate the convex combination axiom. Another solution is to say that the joint credal set for two independent variables X and Y is the convex hull of all distributions $p(X)p(Y)$. This joint credal set is called the *type-1* extension[3], as suggested by Walley [58, section 9.4.5].

Type-1 extensions are direct generalizations of standard independence concepts, but they are not based on any convention regarding preferences, lower expectations or credal sets. The only justification for type-1 extensions is their apparent similarity to standard joint distributions. Despite this similarity, researchers have voiced concerns about type-1 extensions [1, 8, 11] because not all distributions in a type-1 credal set have independent marginals; the next example shows this fact.

Example 2. *In example 1, take the joint credal set K as the convex hull of K'. The following distribution is in K, but it does not have independent marginals:*

$$p(X = 0, Y = 0) = 0.3864, \qquad p(X = 0, Y = 1) = 0.2456,$$
$$p(X = 1, Y = 0) = 0.2256, \qquad p(X = 1, Y = 1) = 0.1424.$$

[2] In Bayesian theory, two variables X and Y are independent if $P(XY) = P(X)P(Y)$ [40].

[3] This convention is called *type-3* independence by Campos and Moral [11], who consider it the most interesting independence concept for sets of distributions.

I propose to adopt Walley's definition of independence [58, sections 9.1, 9.2, 9.3], which is based on conditional lower expectations and uses the concept of *irrelevance*. Variable X is *irrelevant* to variable Y when $\underline{E}[f(X)|Y] = \underline{E}[f(X)]$ for any function $f(X)$. Variables X and Y are independent when X is irrelevant to Y and vice-versa.

To develop a theory of convex sets of distributions, it is important to recast Walley's definition using credal sets.

Consider two sets of variables X and Y and the credal sets $K(X,Y)$, $K(X)$, $K(Y)$, $K(X|Y)$ and $K(Y|X)$. Note that distributions in $K(X)$ and $K(X|Y)$ are defined over the same algebra of events once Y is fixed; likewise, distributions in $K(Y)$ and $K(Y|X)$ are defined over the same algebra of events once X is fixed.

X is *irrelevant* to Y if $K(Y)$ is equal to $K(Y|X)$ regardless of the value of X. Likewise, Y is irrelevant to X if $K(X)$ is equal to $K(X|Y)$ regardless of the value of Y. X and Y are *independent* if X is irrelevant to Y and Y is irrelevant to X.

This concept of independence does not imply that joint credal sets contain only joint distributions with independent marginals; it does not even imply *uniqueness* for the joint credal set. The next example, taken from Walley [58, section 9.3.4], is the simplest way to illustrate this fact.

Example 3. *The credal sets for binary variables A and B contain all distributions such that $0.4 \leq p(1) \leq 0.5$. The largest set of joint distributions that have the correct marginal distributions for A and the correct conditional distributions for B is the convex hull of distributions in table 1. But the convex hull of distributions (a) and (b), or distributions (c) and (d), or distributions (a), (b), (c) and (d), also produces correct marginal and conditional distributions.*

The concept of independence can be generalized to include conditional independence. Given a collection of variables with a set of joint distributions K, and three groups of variables in this collection, $\tilde{X}$, $\tilde{Y}$ and $\tilde{Z}$, say that $\tilde{X}$ is independent of $\tilde{Y}$ given $\tilde{Z}$ if $K(\tilde{X}|\tilde{Z})$ is equal to $K(\tilde{X}|\tilde{Y},\tilde{Z})$, and $K(\tilde{Y}|\tilde{Z})$ is equal to $K(\tilde{Y}|\tilde{X},\tilde{Z})$, regardless of the

Distribution	$p(A=0,B=0)$	$p(A=0,B=1)$	$p(A=1,B=0)$	$p(A=1,B=1)$
a	0.25	0.25	0.25	0.25
b	0.16	0.24	0.24	0.36
c	0.2	0.2	0.3	0.3
d	0.2	0.3	0.2	0.3
e	$0.22222\underline{2}$	$0.22222\underline{2}$	$0.22222\underline{2}$	$0.33333\underline{3}$
f	$0.1818\underline{18}$	$0.2727\underline{27}$	$0.2727\underline{27}$	$0.2727\underline{27}$

Table 1: Joint distributions for example 3.

value of $\tilde{Z}$.

6 Lower and upper values

Suppose that instead of starting from preferences or lower expectations, one specifies a credal set K.

For any function f, obtain *lower and upper expectations*:

$$\underline{E}\left[f\right] = \min_{p \in K} E_p[f], \qquad \overline{E}\left[f\right] = \max_{p \in K} E_p[f]. \tag{7}$$

Upper expectations can be obtained from lower expectations, since $\overline{E}\left[f\right] = \Leftrightarrow \underline{E}\left[\Leftrightarrow f\right]$ for any function f.

For any event A, obtain *lower and upper envelopes*:

$$\underline{p}\left(A\right) = \min_{p \in K} p(A), \qquad \overline{p}\left(A\right) = \max_{p \in K} p(A). \tag{8}$$

Upper envelopes can be obtained from lower envelopes, since $\overline{p}\left(A\right) = 1 \Leftrightarrow \underline{p}\left(A^c\right)$ for any event A. To obtain upper and lower envelopes from lower and upper expectations, take the indicator function δ_A for any event A:

$$\overline{p}\left(A\right) = \underline{E}\left[\delta_A\right], \qquad \overline{p}\left(A\right) = \overline{E}\left[\delta_A\right].$$

	θ_1 (*rain*)	θ_2 (*no rain*)	$E[d_i]$ for $\alpha = 0$	$E[d_i]$ for $\alpha = 1$
d_1 (*park*)	-10	10	-4	4
d_2 (*market*)	5	-4	2.3	-1.3
d_3 (*home*)	0	0	0	0

Table 2: Utilities and expected utilities as a function of α

A credal set always generates unique lower and upper envelopes, but a lower envelope does not define a unique credal set [58, section 2.7]. The Quasi-Bayesian approach sidesteps this difficulty by taking convex sets as basic entities.

For any variable X, the variance of X is $V_p[X] = E_p[X^2] \Leftrightarrow (E_p[X])^2$. Lower and upper variances are defined [58]:

$$\underline{V}[X] = \min_{p \in K} V_p[X], \qquad \overline{V}[X] = \max_{p \in K} V_p[X].$$

7 Quasi-Bayesian decision-making

Since preferences are partially ordered for a Quasi-Bayesian decision-maker, not all decisions are comparable through expected utility.

Example 4. *Suppose a decision-maker has three alternatives, d_1 (*go to the park*), d_2 (*go to the market*) and d_3 (*stay home*). There are two states of nature, θ_1 (*rain*) and θ_2 (*no rain*). The decision-maker has a credal set defined by $p(\theta_1) = \alpha 0.3 + 0.7(1 \Leftrightarrow \alpha)$, $\alpha \in [0, 1]$. The utility function is given in Table 2. Notice that for $\alpha = 0$, $E[d_2] > E[d_3] > E[d_1]$; for $\alpha = 1$, $E[d_1] > E[d_3] > E[d_2]$.*

There is considerable debate about how a Quasi-Bayesian decision-maker actually makes a decision. The most direct strategy is to find a "best" decision for *each* distribution in the credal set, and leave the decision-maker with a set of "best" decisions

[4, 17, 19, 34]. The purpose of this section is not to advocate a particular decision-making stratey, but only to point out some current proposals.

There are several suggestions on how a *single* "best" decision should be selected by a Quasi-Bayesian decision-maker. The first strategy is to select a single distribution from the credal and find the "best" decision for this particular distribution. One possible choice is the maximum-entropy distribution [37, 36, 59]; another is the centroid of the credal set [30].

Levi's approach is to consider only courses of action maximizing expected utility for at least one distribution in the joint credal set [32, 33]. Levi provides additional methods to select one decision on what he calls *security* considerations.

Another approach is to select a decision with maximum upper expected utility. This approach has been advocated in the planning literature; it is appropriate when the upper bound is guaranteed to be attained using secondary actions.

Finally, another approach is to select a decision with maximum lower expected utility. This approach seems reasonable in the context of robustness analysis, and in fact variants of it have been used in robust Statistics [2, 28, 43].

Example 5. *In example 4, d_3 is never the best for any value of α, so it would never be selected according to Levi, but d_3 has maximum lower expected utility. d_1 has maximum upper expected utility.*

8 Conclusion

As originally presented by Giron and Rios, Quasi-Bayesian theory displays an attractive balance between generality and simplicity. But the theory is rather incomplete: they did not explore independence relations, and provided almost no discussion of conditionalization and decision-making.

This paper updates the original Quasi-Bayesian theory, while preserving the same balance between generality and simplicity. The main contribution is the adaptation of Walley's irrelevance and independence concepts to Giron and Rios' preferences and convex sets of distributions.

The derivation presented here flows from preferences to lower expectations, and then to convex sets of distributions. This contrasts with other approaches where convex sets of distributions are derived directly from preferences [32, 47]. The approach taken here simplifies definitions and theorems and gives a more intuitive interpretation for key concepts in the theory. Consider functions as commodities; the lower expectation for a function f is the higher "price" that is paid for this function. The "price" of f depends on the values of f and the beliefs regarding each state of nature. Such beliefs are represented by convex sets of distributions.

In the derivation, both Walley's almost-preference axioms and Walley's generalized Bayes rule are obtained from axioms of preference. First, Walley's almost-preference axioms are demonstrated to be identical to Giron and Rios' axioms. Second, theorem 7 derives Walley's generalized Bayes rule directly from conditional preferences. These new results highlight the conceptual similarity between Walley's theory of imprecise probabilities, which is based on lower expectations, and Quasi-Bayesian theory, which is based on preferences.

A Walley's almost-preference axioms

To prove theorem 1, start with some simple lemmas, which will lead to the proof of the main result.

Lemma 1. *If* $Z \succeq 0$*, then* $Z \succeq \frac{Z}{2}$ *and* $0 \succeq \Leftrightarrow\frac{Z}{2}$.

Proof. Use the convex combination axiom with Z and $\Leftrightarrow Z$. First, $\frac{Z}{2} + \frac{Z}{2} \succeq \frac{1}{2}0 + \frac{Z}{2}$, so $Z \succeq \frac{Z}{2}$. Second, $\frac{Z}{2} + \frac{-Z}{2} \succeq \frac{1}{2}0 + \frac{-Z}{2}$, so $0 \succeq \Leftrightarrow\frac{Z}{2}$. $\square$

Lemma 2. *If $X \succeq Y$ then $X \Leftrightarrow Y \succeq 0$.*

Proof. If $X \succeq Y$ then $\frac{X}{2} \Leftrightarrow \frac{Y}{2} \succeq \frac{Y}{2} \Leftrightarrow \frac{Y}{2} = 0$ (from the convex combination axiom), so $\frac{X}{2} \Leftrightarrow \frac{Y}{2} \succeq 0$. Since $X \Leftrightarrow Y \succeq \frac{X}{2} \Leftrightarrow \frac{Y}{2}$ (from lemma (1)), by transitivity, $X \Leftrightarrow Y \succeq 0$. $\square$

Lemma 3. *If $X \succeq 0$ and $Y \succeq 0$, then $X + Y \succeq 0$.*

Proof. The following sequence of conclusions is valid:

$$\begin{array}{rrl}
0 \succeq \Leftrightarrow\dfrac{X}{2}, & 0 \succeq \Leftrightarrow\dfrac{Y}{2} & \text{by lemma (1)},\\
Y \succeq \Leftrightarrow\dfrac{X}{2}, & X \succeq \Leftrightarrow\dfrac{Y}{2} & \text{by transitivity},\\
Y + \dfrac{X}{2} \succeq 0, & X + \dfrac{Y}{2} \succeq 0 & \text{by lemma (2)},\\
 & 0 \succeq \Leftrightarrow\dfrac{Y}{2} \Leftrightarrow \dfrac{X}{4} & \text{by lemma (1)},\\
 & X + \dfrac{Y}{2} \succeq \Leftrightarrow\dfrac{Y}{2} \Leftrightarrow \dfrac{X}{4} & \text{by transitivity},\\
 & X + \dfrac{X}{4} + \dfrac{Y}{2} + \dfrac{Y}{2} \succeq 0 & \text{by lemma (2)},\\
 & \dfrac{5X}{4} + Y \succeq 0. &
\end{array}$$

Now use the convex combination axiom with $\lambda = \frac{4}{5}$ and $Z = 0$:

$$\frac{4}{5}\left(\frac{5X}{4} + Y\right) + \frac{1}{5}Y \succeq \frac{4}{5}0 + \frac{1}{5}Y,$$

which yields $X + Y \succeq \frac{Y}{5}$.

By the convex combination axiom with $\lambda = \frac{1}{5}$ and $Z = 0$:

$$\frac{1}{5}Y + \frac{4}{5}0 \succeq \frac{1}{5}0 + \frac{4}{5}0,$$

which yields $\frac{Y}{5} \succeq 0$. By transitivity, $X + Y \succeq 0$. $\square$

Lemma 4. *$X \succeq Y$ if and only if $X \Leftrightarrow Y \succeq 0$. [cancellation]*

Proof. One direction is given by lemma (2). If $X - Y \succeq 0$, then $2X - 2Y \succeq 0$ (lemma (3)). Then $\frac{2X-2Y}{2} + \frac{1}{2}2Y \succeq \frac{1}{2}0 + \frac{1}{2}2Y$, so $X \succeq Y$. □

Lemma 5. *If $X_i \succeq Y_i$ then $\sum X_i \succeq \sum Y_i$.*

Proof. By cancellation, $X_i - Y_i \succeq 0$, and by induction on lemma (3), $\sum_i (X_i - Y_i) \succeq 0$, so $\sum_i X_i - \sum Y_i \succeq 0$. By cancellation, $\sum X_i \succeq \sum Y_i$. □

Lemma 6. *If $X \succeq Y$, then $\lambda X \succeq \lambda Y$ for $\lambda > 0$ [positive homogeneity].*

Proof. For any $\lambda' \in (0, 1]$, $\lambda' X \succeq \lambda' Y$. Now by decomposing every number $\lambda > 0$ as $\lambda = \sum 1 + \lambda'$ where $0 \leq \lambda' \leq 1$, obtain the $\lambda X = \lambda' X + \sum X$ and $\lambda Y = \lambda' Y + \sum Y$. So $\lambda X \succeq \lambda Y$ by lemma 5. □

Lemma 7. *If $X + \lambda \succeq Y$ for all $\lambda > 0$ then $X \succeq Y$ [continuity].*

Proof. If $X + \lambda \succeq Y$ for all $\lambda > 0$, construct the sequence $X_i = X + \frac{\lambda_0}{2^i}$, for $\lambda_0 > 0$, $i = 1, 2, \ldots$. Then $X_i \succeq Y$ by hypothesis and $X_i \to X$, so $X \succeq Y$ (by convergence). □

Lemma 8. *If $X \geq Y$ then $X \succeq Y$ [monotonicity].*

Proof. If $X \geq Y$, then $X + \lambda > Y$ for all $\lambda > 0$; by the dominance axiom, $X + \lambda \succ Y$, so $X + \lambda \succeq Y$ for all $\lambda > 0$. By the continuity theorem, $X \succeq Y$. □

The last theorem implies that preference is reflexive: $X \succeq X$.

Theorem 1. *Giron and Rios' axioms are equivalent to the following axioms:*

1. *If $X = -1$ and $Y = 0$, it is not the case that $X \succeq Y$. [sure gain]*
2. *If $X \succeq Y$ and $Y \succeq Z$ then $X \succeq Z$. [transitivity]*
3. *If $X \geq Y$ then $X \succeq Y$. [monotonicity]*

4. *If* $X \succeq Y$ *and* $\lambda > 0$ *then* $\lambda X \succeq \lambda Y$. *[positive homogeneity]*

5. *If* $X + \lambda \succeq Y$ *for all* $\lambda > 0$ *then* $X \succeq Y$. *[continuity]*

6. $X \succeq Y$ *if and only if* $X \Leftrightarrow Y \succeq 0$. *[cancellation]*

Proof. The first axiom comes directly from dominance; the second is transitivity. All others were derived from Giron and Rios' axioms in the previous lemmas.

To obtain Giron and Rios' axioms from the axioms in the thorem, start with the convex combination axiom. The convex combination axiom follows from the properties of positive homogeneity. If $X \succeq Y$, then $\lambda X \succeq \lambda Y$ and then $\lambda X + (1 \Leftrightarrow \lambda) Z \Leftrightarrow (1 \Leftrightarrow \lambda) Z \succeq \lambda Y$, so by cancellation $\lambda X + (1 \Leftrightarrow \lambda) Z \succeq \lambda Y + (1 \Leftrightarrow \lambda) Z$ (the converse of all implications is true).

Dominance comes primarily from monotonicity. If $X > Y$ then $X \geq Y$, then $X \succeq Y$ by monotonicity. If $X > Y$ then $X \geq Y + \gamma$ for $\gamma = \inf |X \Leftrightarrow Y|$, so $X \succeq Y + \gamma$ by monotonicity. Suppose $Y \succeq X$; then $Y \succeq X$ and $X \succeq Y + \gamma$; this would imply that $Y \succeq Y + \gamma$ (by transitivity) and then $0 \succeq \gamma$ (by cancellation). Using positive homogeneity, obtain $0 \succeq 1$, which conflicts with sure gain. Since $X \succeq Y$ and not $Y \succeq X$, $X \succ Y$.

Finally, convergence requires several axioms. It is always possible to find a large enough n such that $|X_n \Leftrightarrow X| < \lambda$ for all $i > n$, all $\lambda > 0$. This implies that $\lambda \Leftrightarrow |X \Leftrightarrow X_i| > 0$ and then $\lambda \Leftrightarrow |X \Leftrightarrow X_i| \succeq 0$ by monotonicity. Since $X_i \Leftrightarrow Z \succeq 0$, $X \Leftrightarrow (X \Leftrightarrow X_i) \Leftrightarrow Z + \lambda \Leftrightarrow |X \Leftrightarrow X_i| \succeq 0$. This implies that $X \Leftrightarrow Z + \lambda \succeq (X \Leftrightarrow X_i) + |X \Leftrightarrow X_i| \succeq 0$. Since $X \Leftrightarrow Z + \lambda \succeq 0$ for all $\lambda > 0$, $X \Leftrightarrow Z \succeq 0$ by continuity and $X \succeq Z$ by cancellation. Since $Y \Leftrightarrow X_i \succeq 0$, so $Y \Leftrightarrow X \Leftrightarrow (X_i \Leftrightarrow X) + \lambda \Leftrightarrow |X_i \Leftrightarrow X| \succeq 0$. This implies that $Y \Leftrightarrow X + \lambda \succeq (X_i \Leftrightarrow X) + |X_i \Leftrightarrow X| \succeq 0$. Since $Y \Leftrightarrow X + \lambda \succeq 0$ for all $\lambda > 0$, $Y \Leftrightarrow X \succeq 0$ by continuity and $Y \succeq X$ by cancellation. $\square$

B Derivation of Quasi-Bayesian theory

This section collects a number of well-known facts about lower expectations and convex sets of distributions. Some results have been mentioned in the literature without proofs based on preferences (theorems 2, 3, 4 and 7); others have been published with different terminology and notation, are are included here for completeness (theorems 5 and 6).

Theorems 8 and 9 and lemma 9 are fundamental results taken directly from Walley's book [58]; their proofs are omitted.

Theorem 2. *Lower expectations have the following properties:*

1. $\underline{E}[f] \geq \inf f$.
2. $\underline{E}[\lambda f] = \lambda \underline{E}[f]$ *for* $\lambda > 0$. *[positive homogeneity]*
3. $\underline{E}[f+g] \geq \underline{E}[f] + \underline{E}[g]$. *[super-additivity]*

Proof. Since $f \geq \inf f$, $\max_\mu [f \succeq \mu] \geq \inf f$ by monotonicity. Positive homogeneity yields: $\underline{E}[\lambda f] = \max_{\mu_1} [\lambda f \succeq \mu_1] = \max_{\lambda \mu_2} [\lambda f \succeq \lambda \mu_2] = \lambda \max_{\mu_2} [f \succeq \mu_2] = \lambda \underline{E}[f]$. Finally, consider $\max_\mu [f+g \succeq \mu]$. Since $f \succeq \underline{E}[f]$ and $g \succeq \underline{E}[g]$, $f+g \succeq \underline{E}[f] + \underline{E}[g]$ (lemma 5). Then $\max_\mu [f+g \succeq \mu] \geq \underline{E}[f] + \underline{E}[g]$, so $\underline{E}[f+g] \geq \underline{E}[f] + \underline{E}[g]$. □

Theorem 3. $\underline{E}[f \Leftrightarrow g] \geq 0$ *if and only if* $f \succeq g$.

Proof. If $f \succeq g$, then $f \Leftrightarrow g \succeq 0$ (by cancellation) and $\underline{E}[f \Leftrightarrow g] \geq 0$ by expression (2). If $\underline{E}[f \Leftrightarrow g] \geq 0$, then $f \Leftrightarrow g \succeq 0$ by transitivity; then $f \succeq g$ by cancellation. □

Theorem 4. *Any functional satisfying the properties in theorem 2 generates a preference pattern (through expression (3)) satisfying Giron and Rios' axioms.*

Proof. First notice two facts:

- Suppose X is equal to a constant μ. Since $\underline{E}[0] = 0$ by positive homogeneity, $0 = \underline{E}[X + (\Leftrightarrow X)] \geq \underline{E}[X] + \underline{E}[\Leftrightarrow X]$, so $\underline{E}[X] \leq \Leftrightarrow \underline{E}[\Leftrightarrow X] \leq \Leftrightarrow \inf(\Leftrightarrow X) = \sup(X) = \mu$. Since $\underline{E}[X] \geq \inf(X) = \mu$, $\underline{E}[X] = \mu$.

- Suppose X is equal to $Y+\mu$ for a constant μ. By superadditivity, $\underline{E}[X] \geq \underline{E}[Y] + \mu$ and $\underline{E}[Y] \geq \underline{E}[X] \Leftrightarrow \mu$ (since $\underline{E}[\mu] = \mu$). Since the last expression yields $\underline{E}[X] \leq \underline{E}[Y] + \mu$, $\underline{E}[X] = \underline{E}[Y] + \mu$.

Given these facts, the proof of the theorem involves showing that preferences produced through expression (3) satisfy axioms in theorem 1, provided that properties in theorem 2 are satisfied by the functional $\underline{E}$.

Sure gain is directly verified as it states that if $X = \Leftrightarrow 1$ and $Y = 0$, $\underline{E}[\Leftrightarrow 1 \Leftrightarrow 0] = \Leftrightarrow 1 \leq 0$.

Transitivity is directly verified as it states that if $\underline{E}[X \Leftrightarrow Y] \geq 0$ and $\underline{E}[Y \Leftrightarrow Z] \geq 0$, then $\underline{E}[X \Leftrightarrow Z] = \underline{E}[X \Leftrightarrow Y + Y \Leftrightarrow Z] \geq \underline{E}[X \Leftrightarrow Y] + \underline{E}[Y \Leftrightarrow Z] \geq 0$.

Positive homogeneity and cancellation are directly satisfied.

Continuity states that if $\underline{E}[X + \lambda \Leftrightarrow Y] \geq 0$ for all $\lambda > 0$, then $\underline{E}[X \Leftrightarrow Y] \geq 0$. Suppose $\underline{E}[X \Leftrightarrow Y] < 0$. Pick $\lambda' = \Leftrightarrow \underline{E}[X \Leftrightarrow Y]/2$; then $\underline{E}[X \Leftrightarrow Y] \leq \Leftrightarrow \lambda' \leq 0$ and $\underline{E}[X + \lambda' \Leftrightarrow Y] \leq 0$, contradicting the hypothesis. □

A general result about lower expectations is [58, section 2.5.5]:

Lemma 9 (Walley's coherence condition). *A functional $\underline{E}[X]$ is a lower expectation if and only if for all non-negative integers m, n, and all functions X_i, Y:*

$$\sup\left(\sum_{j=1}^{n}(X_j \Leftrightarrow \underline{E}[X_j]) \Leftrightarrow m(Y \Leftrightarrow \underline{E}[Y])\right) \geq 0.$$

Theorem 5. *A functional $\underline{E}[X]$ is a lower expectation if and only if it is represented by its set K of dominating distributions.*

Proof. Suppose a functional $\underline{E}[X]$ is represented by K; for every function, $\underline{E}[X] = \min_{p \in K} E_p[X]$. Since every $E_p[X] \geq \inf(X)$, $\underline{E}[X] \geq \inf(X)$. Since $E_p[\lambda X] = \lambda E_p[X]$, $\underline{E}[\lambda X] \geq \lambda \underline{E}[X]$. And since the minimum operator is superadditive ($\min(X+Y) \geq \min X + \min Y$), $\underline{E}[X+Y] \geq \underline{E}[X] + \underline{E}[Y]$. So the functional is a lower expectation by theorem 4.

Suppose a functional is a lower expectation. To prove that it is represented by its set K of dominating distributions, it is enough to prove that for every function Y, there is a probability distribution in K such that: (1) $E_p[Y] = \underline{E}[Y]$, and (2) $E_p[X] \geq \underline{E}[X]$ for all other functions X.

To prove this, consider the set F_X with all functions $X \Leftrightarrow \underline{E}[X]$. For an arbitrary function Y, construct the set $F_Y = \{F_X \cup (\underline{E}[Y] \Leftrightarrow Y)\}$.

For any $n \geq 1$ and functions $\{Z_i\}_1^n$, $Z_i \in F_Y$, lemma 9 indicates that $\sup\left(\sum_{j=1}^n Z_j\right) \geq 0$. This implies that there is an expectation $E_p[Z]$ such that $E_p[Z] \geq 0$ for all functions in F_Y (this is a consequence of the separating hyperplane theorem, as shown by Walley's separation lemma [58, section 3.3.2]).

Using this result, take a distribution $p_Y(\theta)$ such that:

- $E_{p_Y}[X \Leftrightarrow \underline{E}[X]] \geq 0$ for all X; this implies that $p_Y(\theta) \in K$;
- $E_{p_Y}[\underline{E}[Y] \Leftrightarrow Y] \leq 0$; this implies that $E_{p_Y}[Y] = \underline{E}[Y]$.

This proves that for every Y, $\underline{E}[Y] = \min_{p \in K} E_p[Y]$. □

Theorem 6. *For a preference pattern satisfying Giron and Rios' axioms, there is a closed and convex set of probability distributions K such that:*

$$f \succeq g \Leftrightarrow E_p[f] \geq E_p[g] \text{ for every } p \in K.$$

Proof. For a preference pattern satisfying Giron and Rios' axioms, a lower expectation can be defined through expression (2). By theorem 5, the set of dominating expectations

K represents the lower expectation; this set is closed and convex. If $f \succeq g$, then $\underline{E}[f \Leftrightarrow g] \geq 0$, so $E_p[f \Leftrightarrow g] \geq 0$ for all $p(\theta) \in K$; consequently, $E_p[f] \geq E_p[g]$ for all $p(\theta) \in K$. Conversely, if $E_p[f] \geq E_p[g]$ for all $p(\theta) \in K$, then $E_p[f \Leftrightarrow g] \geq 0$ for all $p(\theta) \in K$; consequently, $f \succeq g$. $\square$

Theorem 7. *For a non-null event A, $\underline{E}[f|A]$ is the only solution of the equation:*

$$\underline{E}[(f \Leftrightarrow \lambda)\, \delta_A] = 0. \tag{9}$$

Proof. By theorem 3, $\max_\mu [(f \Leftrightarrow \mu)\delta_A \succeq 0]$ and $\max_\mu [\underline{E}[(f \Leftrightarrow \mu)\delta_A] \geq 0]$ are equivalent. For $\lambda > \mu$, $\underline{E}[(f \Leftrightarrow \mu)\delta_A] \geq \underline{E}[(f \Leftrightarrow \lambda)\delta_A] + (\lambda \Leftrightarrow \mu)\underline{E}[\delta_A]$. Since $\underline{E}[\delta_A] > 0$, the function $\underline{E}[(f \Leftrightarrow \mu)\delta_A]$ is strictly decreasing with μ; since this function is negative for $\mu > \sup(f)$ and negative for $\mu < \inf(f)$, the value $\arg\max_\mu [\underline{E}[(f \Leftrightarrow \mu)\delta_A] \geq 0]$ is the only solution of equation (6). $\square$

References

[1] J. Berger and E. Moreno. Bayesian robustness in bidimensional model : Prior independence. *Journal of Statistical Planning and Inference*, 40:161–176, 1994.

[2] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, 1985.

[3] C. Bicchieri and G. A. Antonelli. Game-theoretic axioms for local rationality and bounded knowledge. Technical Report CMU-PHIL-37, Carnegie Mellon University, May 1993.

[4] J. S. Breese and K. W. Fertig. Decision making with interval influence diagrams. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 467–478. Elsevier Science, North-Holland, 1991.

[5] P. Cheeseman. In defense of probability. In *Proceeding of International Joint Conference on Artificial Intelligence*, pages 1003–1009, 1985.

[6] C. Cherniak. *Minimal Rationality.* The MIT Press, Cambridge, MA, 1986.

[7] L. Chrisman. Incremental conditioning of lower and upper probabilities. *International Journal of Approximate Reasoning*, 13(1):1–25, 1995.

[8] L. Chrisman. Independence with lower and upper probabilities. *XII Uncertainty in Artificial Intelligence Conference*, pages 169–177, 1996.

[9] L. Chrisman. Propagation of 2-monotone lower probabilities on an undirected graph. *XII Uncertainty in Artificial Intelligence Conference*, pages 178–186, 1996.

[10] F. Cozman and E. Krotkov. Quasi-Bayesian strategies for efficient plan generation: Application to the planning to observe problem. *Proc. Twelfth Conference Uncertainty in Artificial Intelligence*, pages 186–193, 1996.

[11] L. de Campos and S. Moral. Independence concepts for convex sets of probabilities. *XI Uncertainty in Artificial Intelligence*, 1995.

[12] A. P. Dempster. A generalization of Bayesian inference. *Journal Royal Statistical Society B*, 30:205–247, 1968.

[13] J. Earman. *Bayes or Bust?* The MIT Press, Cambridge, MA, 1992.

[14] T. L. Fine. Lower probability models for uncertainty and nondeterministic processes. *Journal of Statistical Planning and Inference*, 20:389–411, 1988.

[15] P. C. Fishburn. *Utility Theory for Decision Making.* John Wiley and Sons, Inc., New York, 1970.

[16] L. Fortnow and D. Whang. Optimality and domination in repeated games with bounded players. In *Proceedings of the 26th ACM Symposium on the Theory of Computing*, pages 741–749, New York, 1994.

[17] P. Gardenfors and N. E. Sahlin. Unreliable probabilities, risk taking and decision making. *Synthese*, 53:361–386, 1982.

[18] F. J. Giron and S. Rios. Quasi-Bayesian behaviour: A more realistic approach to decision making? In J. M. Bernardo, J. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics*, pages 17–38. University Press, Valencia, Spain, 1980.

[19] I. J. Good. *Good Thinking: The Foundations of Probability and its Applications.* University of Minnesota Press, Minneapolis, 1983.

[20] H. E. Kyburg Jr. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, 31:271–293, 1987.

[21] V. Ha and P. Haddawy. Theoretical foundations for abstraction-based probabilistic planning. *XII Uncertainty in Artificial Intelligence Conference*, pages 291–298, 1996.

[22] I. Hacking. Slightly more realistic personal probability. *Philosophy of Science*, pages 311–325, December 1967.

[23] J. Y. Halpern and R. Fagin. Two views of belief: Belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54:275–317, 1992.

[24] D. Heckerman and H. Jimison. A Bayesian perspective on confidence. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*, pages 149–160. Elsevier Science Publishers, North-Holland, 1989.

[25] M. E. Hellman and T. M. Cover. Learning with finite memory. *The Annals of Mathematical Statistics*, 41(3):765–782, 1970.

[26] M. E. Hellman and T. M. Cover. A review of recent results on learning with finite memory. In B. N. Petrov and C. Frigues, editors, *2nd International Symposium on Information Theory*, pages 289–294. Akademiai Kiado, Budapest, 1973.

[27] E. J. Horvitz. Reasoning about beliefs and actions under computational resource contraints. In L. N. Kanal, T. S. Levitt, and J. F. Lemmer, editors, *Uncertainty in*

Artificial Intelligence 3, pages 301–324. Elsevier Science Publishers, North-Holland, 1989.

[28] P. J. Huber. *Robust Statistics*. Wiley, New York, 1980.

[29] A. Kukla. Criteria of rationality and the problem of logical sloth. *Philosophy of Science*, 58:486–490, 1991.

[30] H. E. Kyburg and M. Pittarelli. Set-based Bayesianism. *IEEE Transactions on Systems, Man and Cybernetics A*, 26(3):324–339, May 1996.

[31] H. E. Kyburg Jr. *The Logical Foundations of Statistical Inference*. D. Reidel Publishing Company, New York, 1974.

[32] I. Levi. *The Enterprise of Knowledge*. The MIT Press, Cambridge, Massachusetts, 1980.

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

[34] P. Maher. What is wrong with strict Bayesianism? In *PSA*, volume 1, pages 450–457. Philosophy of Science Association, 1986.

[35] P. Maher. *Betting on Theories*. Cambridge studies in Probability, Induction and Decision Theory. Cambridge University Press, New York, 1993.

[36] H. T. Nguyen and E. A. Walker. On decision making using belief functions. In R. Yager, J. Kacprzyk, and M. Fedrizzi, editors, *Advances in the Dempster-Shafer Theory of Evidence*, chapter 15, pages 311–330. Wiley, 1994.

[37] N. J. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28:71–87, 1986.

[38] J. Pearl. Theoretical bounds on the complexity of inexact computations. *IEEE Trans. on Information Theory*, IT-22(5), September 1976.

[39] J. Pearl. On probability intervals. *International Journal of Approximate Reasoning*, 2:211–216, 1988.

[40] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kauffman, San Mateo, CA, 1988.

[41] J. Pearl. Locality-bounded rationality. Technical Report R-124, Computer Science Department, University of California Los Angeles, February 1989.

[42] M. E. Pollack, D. J. Israel, and M. E. Bratman. Toward an architecture for resource-bounded agents. Technical Report CSLI-87-104, Center for Study of Language and Information, 1987.

[43] Y. Ritov. Robust Bayes decision procedures: Gross error in the data distribution. *The Annals of Statistics*, 13(2):626–637, 1985.

[44] E. H. Ruspini. The logical foundations of evidential reasoning. Technical Report SRIN408, SRI International, 1987.

[45] S. Russell and E. Wefald. *Do the Right Thing.* The MIT Press, Cambridge, MA, 1991.

[46] I. R. Savage. On not being rational. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics: Proceedings of the First International Meeting Held in Valencia*, chapter 7, pages 321–328. University Press Valencia, Valencia, Spain, 1980.

[47] T. Seidenfeld. Outline of a theory of partially ordered preferences. *Philosophical Topics*, 21(1):173–188, Spring 1993.

[48] T. Seidenfeld, M. J. Schervish, and J. B. Kadane. Decisions without ordering. In W. Sieg, editor, *Acting and Reflecting*, pages 143–170. Kluwer Academic Publishers, 1990.

[49] G. Shafer. Probability judgment in artificial intelligence and expert systems. *Statistical Science*, 2(1):3–44, 1987.

[50] G. Shafer. Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning*, 4:323–362, 1990.

[51] E. H. Shortliffe and B. G. Buchanan. *A model of inexact reasoning in medicine*, chapter 11, pages 233–232. Addison-Wesley, 1985.

[52] H. A. Simon. Rational choice and the structure of the environment. *Psychol. Review*, 63:129–138, March 1956.

[53] H. A. Simon. Rational decision making in business organizations. *The American Economic Review*, 69(4):493–513, September 1979.

[54] H. A. Simon. *Models of Bounded Rationality 2*. MIT Press, Cambridge MA, 1982.

[55] H. A. Simon and J. B. Kadane. Optimal problem-solving search: All-or-none solutions. *Artificial Intelligence*, 6:235–247, 1975.

[56] C. A. B. Smith. Consistency in statistical inference and decision. *Journal Royal Statistical Society B*, 23:1–25, 1961.

[57] P. Suppes. The measurement of belief. *Journal Royal Statistical Society B*, 2:160–191, 1974.

[58] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York, 1991.

[59] T. Whalen. Interval probabilities induced by decision problems. In R. Yager, J. Kacprzyk, and M. Fedrizzi, editors, *Advances in the Dempster-Shafer Theory of Evidence*, chapter 16, pages 353–374. Wiley, 1994.