ECONOMICS

Working Paper Series

No. 08-2025

Value of Partial Information

Jürgen Eichberger
Alfred Weber Institute for Economics
Heidelberg University
juergen.eichberger@awi.uni-heidelberg.de

Illia Pasichnichenko
University of Sussex
i.pasichnichenko@sussex.ac.uk

Abstract: Blackwell's theorem relates the value of information to the "informativeness" of the information structure. His analysis applies to decision makers who are expected utility maximizers and know the information structure of the decision problem. When decision makers do not know the information structure precisely, the signal generating process and the posterior distributions are often only partially known. This paper studies preferences of decision makers with partial knowledge about signals and posterior probability distributions. The partial information approach allows us to relate the value of information to the decision maker's attitude towards ambiguity. We introduce a new concept of informativeness based on the centroid and prove a theorem in the spirit of Blackwell. Furthermore, we characterize the value of information in terms of the preference relation over information structures. Depending on ambiguity attitude the value of information may be negative.

JEL codes: D81; D83

Key words: Value of information; Ambiguity; Informativeness; Belief functions; Information structure

Value of Partial Information*

Jürgen Eichberger† Illia Pasichnichenko‡

December 5, 2025

Abstract

Blackwell's theorem relates the value of information to the "informativeness" of the information structure. His analysis applies to decision makers who are expected utility maximizers and know the information structure of the decision problem.

When decision makers do not know the information structure precisely, the signal generating process and the posterior distributions are often only partially known. This paper studies preferences of decision makers with partial knowledge about signals and posterior probability distributions. The partial information approach allows us to relate the value of information to the decision maker's attitude towards ambiguity. We introduce a new concept of informativeness based on the centroid and prove a theorem in the spirit of Blackwell. Furthermore, we characterize the value of information in terms of the preference relation over information structures. Depending on ambiguity attitude the value of information may be negative.

JEL classification: D81; D83.

Keywords: Value of information; Ambiguity; Informativeness; Belief functions; Information structure.

1 Introduction

In a world where all information is described by well-specified probabilities, signals about states are naturally evaluated by the value of the improved predictions

*This paper is dedicated to Alain Chateauneuf, a great teacher and good friend of both authors, in honor of his 80th birthday.

†Alfred Weber Institute for Economics, Heidelberg University, Germany. Email: juergen.eichberger@awi.uni-heidelberg.de

‡Department of Economics, University of Sussex, Brighton, United Kingdom. Email: i.pasichnichenko@sussex.ac.uk.

about state-contingent outcomes that will be possible in the light of these signals. Since information that does not improve a decision's outcome can be ignored, the value of information for an expected-utility–maximizing decision maker (DM) can never be negative. Moreover, Blackwell's (1951) seminal theorem links the value of information to the statistical notion of informativeness of the probability distributions over states and signals, that is the "information structure" of the decision problem. Blackwell's analysis applies to DMs who maximize expected utility and have precise knowledge of the information structure. When knowledge about the signal-generating process or about the posterior distributions is incomplete, however, this analysis breaks down.

It is well known (e.g., Grant *et al.*, 1998) that, for agents who do not maximize expected utility, information may have a direct, possibly negative, impact on the DM's well-being. Agents may like or dislike information regardless of its instrumental value. Moreover, even for a probabilistically sophisticated DM with perfect knowledge about the probability distribution over states and signals, Blackwell's theorem needs to be adjusted if the DM is not an expected-utility maximizer (Karni and Safra, 2022). With partial information, a DM's ambiguity regarding updated beliefs over state-contingent outcomes may differ from the ambiguity about signals. Hence, ambiguity attitude with respect to the precision (quality) of the signals can also be different from ambiguity attitude towards updated beliefs (Wang, 2024).[1]

This paper examines preferences over information structures when DMs possess only partial knowledge about signals and posterior probability distributions. We model partial knowledge by belief functions and their associated mass distributions. This approach captures what is objectively known about event probabilities and allows us to relate the value of information to the DM's attitude toward ambiguity. In particular, it combines objective knowledge about likelihoods with subjective beliefs based on the "principle of insufficient reason" when certain probabilities are unknown. Depending on the DM's ambiguity attitude, the value of information may become negative. We illustrate the framework through applications to portfolio choice, data collection, and experimental settings.

Building on this framework, we introduce a new notion of informativeness under partial information, defined via the centroid of the mass distribution, and establish a theorem in the spirit of Blackwell. Specifically, we show that, for a broad class of preferences, one information structure is more informative than another if and only if it yields a higher value in every decision problem. In the

[1] We will discuss the recent literature below in more detail.

special case of complete knowledge about the relationship of states and signals, our result collapses to the standard Blackwell theorem.

Finally, we characterize the value of information through axioms about the DM's preference relation over information structures. We show that our assumptions are both necessary and sufficient for the existence of the value of information proposed in the earlier sections. By relaxing or strengthening these assumptions, we obtain natural generalizations and special cases of the representation.

The paper is organized as follows. Section 2 introduces the formal framework for studying partial information over states, information structures, and the value of information. Section 3 provides examples and applications of the suggested framework. In Section 4, we define the a notion of informativeness for information structures with partial information and establish a theorem in the spirit of Blackwell. Section 5 characterizes the value of information in terms of preferences over information structures. Section 6 concludes.

1.1 Relation to the literature

There is a large literature on decision making under uncertainty. Mostly, however, one finds studies of the extreme cases of (subjective or objective) expected utility maximization either without any uncertainty or under complete uncertainty regarding the probabilities of events. Moreover, if there is ambiguity, DMs are mostly assumed to choose pessimistically the worst expected utility from an exogenously given set of probability distributions. A more balanced approach, that includes these cases as extreme situations, was promoted in Jaffray (1989). It is based on the study of belief functions and their mass distributions that may include information about some but not necessarily all events, a situation we dub partial information. Eichberger and Pasichnichenko (2021) provide a recent axiomatic treatment for this situation.

In a series of papers (Blackwell, 1951, 1953; Blackwell and Girshick, 1954), Blackwell introduced the leading framework for studying information gathering as part of the decision-making-under-uncertainty problem. In the statistical context of expected utility maximization, his famous theorem established that additional information could never be harmful and would be only useless if the additional information was pure noise.

Several papers have extended Blackwell's approach to the non-expected utility framework. Part of this literature (Grant *et al.*, 1998; Karni and Safra, 2022) remained in the probabilistic environment while others (Çelen, 2012; Cheng *et al.*, 2025; Heyen and Wiesenfarth, 2015; Li, 2020) considered beliefs modeled by capacities and sets of multiple priors. Various extensions of Blackwell's theorem relating measures of informativeness to notions of value for information systems have been suggested in this literature. Of crucial importance for these extensions is how the valuation of signals and the valuation of signal-dependent state-contingent actions is modeled. Wang (2024), for example, introduces an auxiliary state space that indexes possible Blackwell experiments and defines two informativeness orders: prior-by-prior dominance, which is robust across all monotone ambiguity preferences, and Wald informativeness, which applies to maxmin DMs.

Of relevance is also the literature on updating belief functions, see Dubois and Prade (1990), Eichberger *et al.* (2010), and Gilboa and Schmeidler (1993). Lin and Payró (2024) proposed a normative framework for belief updating under ambiguous information sources and characterized a broad family of generalized Bayesian updating rules.

Few papers study experimentally the evaluation of information under ambiguity. Kops and Pasichnichenko (2023) and Shishkin and Ortoleva (2023) are experimental studies investigating the possibility of a negative value of information. For example, Kops and Pasichnichenko (2023) find that the majority of ambiguity averse subjects exhibited a negative value of information. Abdellaoui *et al.* (2025) and Gonzalez-Jimenez (2024) study how learning from experience in an environment affects ambiguity attitudes and beliefs.

2 Framework

2.1 Decision making

Consider a finite set of states Ω and a set of outcomes X. Let $\mathcal{A}$ denote the set of all (finitely-valued) functions $a : \Omega \to X$, referred to as actions (acts). In contrast to Savage (1954), we do not presume that probabilities over states are purely subjective. Following Dempster (1967), Shafer (1976) and Jaffray (1989), we assume that partial information from data or prior knowledge provides the DM with a *mass distribution* m over events in the power set of states $\mathcal{P}(\Omega)$, where $m(E) \geq 0$ for all $E \in \mathcal{P}(\Omega)$, $m(\emptyset) = 0$, and $\sum_{E\in\mathcal{P}(\Omega)} m(E) = 1$. A mass distribution can be viewed as a probability distribution over subsets of Ω.[2]

Remark 1. As is well known (see, e.g., Grabisch, 2016), a mass distribution m on

[2] When only singleton sets $\{\omega\}$ have strictly positive weights $m(\{\omega\}) > 0$, the mass distribution m can be identified with a usual probability distribution on Ω.

$\mathcal{P}(\Omega)$ is the Möbius transform of a *belief function* μ^m defined by

$$\mu^m(E) = \sum_{F \subseteq E} m(F) \tag{1}$$

for all $E \in \mathcal{P}(\Omega)$. A belief function resembles a probability distribution, although it is not necessarily additive. In fact, it is a capacity that is monotone of all orders (see Chateauneuf and Jaffray (1989) for details). Given a belief function, the underlying mass distribution can be recovered uniquely. Hence, there is a one-to-one correspondence between belief functions and mass distributions. A belief function is a *convex* capacity and, hence, has a non-empty set of dominating probability distributions, the *core* of the capacity μ^m:

$$\text{core}(\mu^m) = \{p \in \Delta(\Omega) \mid p(E) \geq \mu^m(E) \text{ for all } E \subseteq \Omega\}.$$

Given a von Neumann–Morgenstern utility function u on outcomes in X, the *Choquet expected utility (CEU)* of an action $a \in \mathcal{A}$ with respect to the belief function μ^m can be expressed (Grabisch, 2016, Chapter 7, pp. 377-437) either (i) as the average over the minimal outcome of the events in $\mathcal{P}(\Omega)$ weighted by the mass distribution m or (ii) as the minimal expected utility over the probabilities in the core of μ^m:

$$CEU(a, \mu^m) = \sum_{E \in \mathcal{P}(\Omega) \setminus \emptyset} m(E) \left[\min_{\omega \in E} u(a(\omega)) \right] = \min_{p \in \text{core}(\mu^m)} \sum_{\omega \in \Omega} u(a(\omega)) p(\omega).$$

A disadvantage of the Choquet integral is its bias towards the worst case. In this paper, we will use a more balanced representation of preferences among actions based on a *quasi-average utility* introduced in Eichberger and Pasichnichenko (2021). First, notice that any pair of action $a : \Omega \to X$ and mass distribution m on $\mathcal{P}(\Omega)$ induces a mass distribution $(m * a)$ on $\mathcal{P}(X)$ defined by

$$(m * a)(C) = \sum_{E \subseteq \Omega :\, a(E) = C} m(E) \tag{2}$$

for all $C \subseteq X$. Second, given the mass distribution $n = m * a$ on $\mathcal{P}(X)$, we obtain

$$n = \sum_{C \subseteq X} n(C) e_C,$$

where e_C is the elementary mass distribution that assigns 1 to the set C and

0 to all other sets. Therefore, for any linear evaluation V, we get $V(n) = \sum_{C \subseteq X} n(C)V(e_C)$. The partial information embodied in the mass distribution n allows for an evaluation of the mass distribution as an average, where the values $V(e_C)$ are weighted by the information $n(C)$.

In the spirit of the principle of insufficient reason, one can evaluate an elementary mass distributions e_C by the quasi-average

$$V(e_C) = \phi^{-1}\left(\frac{1}{|C|}\sum_{x \in C}\phi(u(x))\right).$$

Since there is no information about the likelihood of sub-events of C, the quasi-average $V(e_C)$ evaluates the elementary mass distribution e_C by a monotone transformation ϕ of the expected utility of the outcomes $x \in C$ with respect to the uniform distribution $\frac{1}{|C|}$. The uniform distribution reflects the DM's ignorance about the sub-events of C. The monotone transformation ϕ represents the DM's ambiguity attitude. Notice that, for a singleton event $\{x\}$, $V(e_{\{x\}}) = \phi^{-1}\left(\frac{1}{|\{x\}|}\phi(u(x))\right) = u(x)$.

There are two extreme cases:

(i) If the mass distribution n on X is concentrated on singleton events (i.e., $n(C) = 0$ for all C with $|C| > 0$; full information about the probabilities of the outcomes in X), then $V(n) = \sum_{x \in X} n(\{x\})V(e_{\{x\}}) = \sum_{x \in X} n(\{x\})u(x)$ becomes the expected utility with respect to the probability distribution n.

(ii) If the mass distribution n is concentrated on X (i.e., $n(C) = 0$ for all $C \neq X$; no information about the probabilities of the outcomes in X), then $V(n) = V(e_X) = \phi^{-1}\left(\frac{1}{|X|}\sum_{x \in X}\phi(u(x))\right)$. In other words, the expected utility of the outcomes in X is evaluated by the uniform distribution, reflecting the complete ignorance of the DM regarding the likelihoods of outcomes in X.

Most applications, however, concern intermediate cases with partial information, where at least some information about events C with $1 < |C| < |X|$ is available. Eichberger and Pasichnichenko (2021) provide an axiomatization for representing preferences by the quasi-average utility, relate the transformation ϕ to the DM's attitude towards ambiguity, and present applications including the Ellsberg paradox.

2.2 Information structures and the ex-ante value of information

Valuation of the information from an information source proceeds in two stages. In the first stage, a DM must decide whether to enter an information gathering activity (experimentation phase). The information gathering activity, also called experiment, will produce a signal s from a finite set of signals S according to some partially known probability distribution. We will assume that information about the ex-ante signal distribution is partial and ambiguous, hence represented by a mass distribution M over signals in S. In the second stage, once a signal s has been observed, the DM can update the ex-ante mass distribution over states in Ω in the light of the information transmitted by the signal s to a mass distribution m_s. Thus, the signal allows a choice of action given the information of the signal. We dub the possibility to adjust the choice of action in the light of the signal as the *instrumental value of information.*

Following Blackwell (1951), we refer to the environment in which the DM seeks information before choosing an action $a \in \mathcal{A}$ as *information structure* or *experiment.* Formally, an *information structure* consists of (i) a finite set of signals S, (ii) a mass distribution M over signals in S reflecting what is known about the likelihood of the signals, plus (iii) a set of mass distributions m_s over states in Ω conditional on the signal $s \in S$ observed:

$$I = \left(S,\ M,\ \{m_s\}_{s\in S}\right).$$

In the special case where the mass distribution M over signals in S is a probability distribution and all mass distributions m_s are Bayesian updates of a prior probability distribution, the "Bayesian" situation studied by Blackwell (1951) obtains.

Learning that a signal $s \in S$ has occurred will allow the DM to revise the initial mass distribution m to an updated mass distribution m_s. The updated mass distribution m_s on states in Ω together with an action $a : \Omega \to X$ induce a mass distribution $(m_s * a)$ on X according to Equation (2). Denoting by $U_{\phi u}(e_C) := \phi^{-1}\left(\frac{1}{|C|}\sum_{x\in C}\phi(u(x))\right)$ the quasi-average utility of the outcomes $C \subseteq X$, one can write the quasi-average utility of the action a given m_s as[3]

$$V(a, m_s) = \sum_{C\subseteq X} (m_s * a)(C)\ U_{\phi u}(e_C). \tag{3}$$

[3] Note that $(m_s * a)(C) > 0$ only for a finite number of sets $C \subseteq X$.

Consider a finite feasible set of actions $A \subseteq \mathcal{A}$. For a given signal s, denote by $a_s \in \arg\max_{a \in A} V(a, m_s)$ a maximizing action and by

$$\widehat{V}(m_s) = V(a_s, m_s) \tag{4}$$

the maximal quasi-average utility. The value $W(I)$ of an information structure $I = \big(S, M, \{m_s\}_{s \in S}\big)$ is obtained by aggregating the values $\hat{V}(m_s)$ of the signals with the quasi-average of the prior mass distribution over signals $M(s)$:

$$W(I) = \sum_{E \subseteq S} M(E)\, \psi^{-1}\left(\frac{1}{|E|} \sum_{s \in E} \psi\left(\hat{V}(m_s)\right)\right), \tag{5}$$

where ψ is a monotone transformation reflecting the ambiguity attitude of the DM regarding the ambiguity of the signal distribution. It is important to note that ψ need not coincide with ϕ. The former reflects ambiguity attitudes regarding the precision of signals, while the latter reflects ambiguity attitudes toward the resulting updated beliefs.

2.3 Signals and updating mass distributions over states

In this paper, we will not investigate in detail how signals in S and states in Ω are related. Two of the most common ways assume either a prior product space $S \times \Omega$ or a function $\sigma : \Omega \to S$ mapping states to signals.

2.3.1 Signals from a product space $S \times \Omega$

Suppose there is a mass distribution ν over the product space $S \times \Omega$. If ν is concentrated on singleton events in $S \times \Omega$, one can obtain the updated mass distributions m_s as the marginal distributions of ν.

For a general mass distribution ν on the product space $S \times \Omega$ however, there are numerous product measures and a product measure is not well defined (see Hendon *et al.*, 1996). A possible solution to this problem consists in picking a particular product measure, e.g., the Möbius product. The following example illustrates this possibility.

Example 1. Consider an urn with objects of different shapes $\mathfrak{S}$ and colors $\mathfrak{C}$. Past observations provide a mass distribution over subsets of $\mathfrak{C} \times \mathfrak{S}$. Imprecise records may cause ambiguity (see, e.g., Walley, 1996). Objects are randomly drawn from the urn. Agents are informed about the shapes of the objects but not their color. Then agents can bet on the color of the object. Hence, the set of

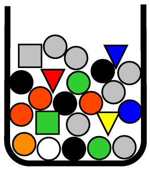

Figure 1: Urn with objects of different shapes and colors

signals is the set of shapes, $S = \mathfrak{S} = \{B, C, T, ...\}$, and the payoff-relevant states are the colors, $\Omega = \mathfrak{C} = \{r, b, g, y, ...\}$.

	state events	signal events									
		$\{B\}$	$\{C\}$	$\{T\}$	$\{B,C\}$	$\{B,T\}$	$\{C,T\}$	$\{B,C,T\}$	m_B	m_C	m_T
	$\{r\}$	$\frac{10}{40}$	$\frac{8}{40}$	0	0	0	0	$\frac{2}{40}$	$\frac{2}{3}$	$\frac{4}{7}$	0
	$\{b\}$	0	0	0	0	0	0	0	0	0	0
	$\{g\}$	0	0	0	0	0	0	0	0	0	0
	$\{r,b\}$	0	0	0	0	0	0	0	0	0	0
	$\{r,g\}$	0	0	0	0	0	0	0	0	0	0
	$\{b,g\}$	0	$\frac{6}{40}$	0	0	0	0	0	0	$\frac{3}{7}$	0
	$\{r,b,g\}$	$\frac{5}{40}$	0	$\frac{9}{40}$	0	0	0	0	$\frac{1}{3}$	0	1
M		$\frac{15}{40}$	$\frac{14}{40}$	$\frac{9}{40}$				$\frac{2}{40}$			

Table 1: Mass distributions conditional on signals

For each signal $s \in S$, beliefs of the agent are assumed to be represented by a mass distribution m_s over the set of states Ω. Information about the likelihood of signals in S is given by a mass distribution M on S. Consider the case of three states $\Omega = \{r, b, g\}$ and three signals $S = \{B, C, T\}$. Suppose there is a mass distribution ν on $\Omega \times S$ that is a Möbius product of two mass distributions over S and Ω. For all $G \in \mathcal{P}(\Omega \times S)$ such that $G = E \times F$ for some $E \subseteq \Omega$ and $F \subseteq S$, $\nu(G)$ is given in Table 1, while $\nu(G) = 0$ for all other $G \in \mathcal{P}(\Omega \times S)$. This mass distribution reflects the partial knowledge the DM may have about the relationship of signals and states. Notice that in this case, one can deduce from the product mass distribution ν both the mass distribution M over S and the updated mass distributions $\{m_s\}_{s \in S}$ over Ω. If the mass distribution were concentrated on the upper left-hand corner of the table, then the updated m_s would correspond to the Bayesian updates of the prior probability distribution over states.

2.3.2 Signals as a function of states

An alternative way to model the relationship between signals and states is by a mapping $\sigma : \Omega \to S$ generating a partition of Ω. If there is no uncertainty regarding the probability distribution governing the likelihood of states, then the mass distributions updated on signals $s \in S$ could be obtained as the Bayesian updates for any event $F \subseteq \Omega$. Notice that in this case the ex-ante distribution over signals is not derived from the model.

Consider an arbitrary ex-ante mass distribution m_0 on the state space Ω and its associated belief function $\mu = \mu^{m_0}$ defined by Equation (1). For an event E from a signal induced partition $\{\sigma^{-1}(s)\}_{s\in S}$ of the state space Ω, one can study numerous notions of updated belief functions and their mass distributions. Specifically, denote by μ_R the belief function updated with respect to the rule R (see Eichberger *et al.*, 2010) assuming the updates are belief functions, e.g.:

- Bayesian: $\mu_B(F|E) = \frac{\mu(F\cap E)}{\mu(E)}$,
- Dempster–Shafer: $\mu_{DS}(F|E) = \frac{\mu(F\cup E^c)-\mu(E^c)}{1-\mu(E^c)}$,
- Generalized Bayesian: $\mu_{GB}(F|E) = \frac{\mu(F\cap E)}{\mu(F\cap E)+1-\mu(F\cup E^c)}$,

for all $F \subseteq \Omega$. The corresponding mass distribution can be derived from the belief function μ_R.

3 Examples and applications

In this section, we present three examples that show the potential of our approach for modeling economic phenomena.

Example 2 (Portfolio Choice with a Consultant)**.** Consider an investor who has to choose between two consultants who offer to provide information about the state-contingent returns of the investor's portfolio. We model the consultants as information structures $I_S = \big(S, M_S, \{m_s\}_{s\in S}\big)$ and $I_T = \big(T, M_T, \{m_t\}_{t\in T}\big)$. A consultant is characterized by a finite set of potential recommendations (signals) S (respectively, T) and signal-dependent mass distributions $\{m_s\}_{s\in S}$ over states (respectively, $\{m_t\}_{t\in T}$) reflecting the partial information regarding the assets' returns given the recommendation $s \in S$ (respectively, $t \in T$). The consultants may be also characterized by prior information regarding the precision of their recommendations. This prior information can be captured by the mass distribution M_S (respectively, M_T).

Assume a set of states $\Omega = \{\omega_1, ..., \omega_n\}$ and a finite set A of portfolios $(a_1, a_2) \in \mathbb{R}^2$ determined by the exogenous supply of stocks and bonds. Given asset prices

	quantity	*price*	*payout in* $\omega \in \Omega$
stock	a_1	q	r_ω
bond	a_2	1	r

Table 2: Asset prices and returns

and returns as in Table 2, the set of portfolios A induces a set of wealth levels

$$X = \{r_\omega a_1 + r a_2 \mid (a_1, a_2) \in A,\ \omega \in \Omega\},$$

which serves as the set of outcomes. Choice of a portfolio $a \in A$ by an investor can be viewed as an action $a : \Omega \to X$ that the investor is assumed to evaluate by its quasi-average utility $V(a, m_s)$ according to Equation (3).

Given a recommendation $s \in S$, the investor will maximize $V(a, m_s)$ by choosing a portfolio a from her budget set $B(q, w) = \{a \in A \mid qa_1 + a_2 \leq w\}$, where w denotes her initial wealth. Denote by a_s a maximizing portfolio and by $\widehat{V}(m_s)$ the maximal quasi-average utility from the recommendation $s \in S$. Hence, the ex-ante value $W(I_S)$ of the information provided by consultant I_S to the investor is given by Equation (5), with an analogous definition for consultant I_T. Notice that ambiguity about the quality (precision, reliability) of the consultants is captured by the mass distributions M_S and M_T over signals and the attitude towards this uncertainty by the function ψ. Given the respective costs of these consultants the investor can choose the one she wants to employ.

The next example illustrates how an information structure may be derived from a data collection process, for instance, through peer recommendations.

Example 3 (Feedback Gathering)**.** Consider a DM who must choose the best option among n available alternatives. To make an informed choice, the DM gathers recommendations from her peers.[4] Let the set of states be $\Omega = \{\omega_1, ..., \omega_n\}$, where ω_i denotes the state in which option i is the best. The set of outcomes is $X = \{x_0, x_1\}$, with x_1 preferred to x_0. The feasible set of actions is $A = \{a_1, ..., a_n\}$, where each action a_i represents a bet on state ω_i: $a_i(\omega_j) = x_1$ if $i = j$, and $a_i(\omega_j) = x_0$ otherwise, for all $i, j \in \{1, ..., n\}$.

Recommendations from k peers can be represented as a sequence $\{o^t\}_{t=1}^k$ of observations over the states. Allowing for imprecise recommendations, each observation is modeled as a non-empty subset $o^t \subseteq \Omega$, where a non-singleton o^t

[4]For example, a hiring committee selecting among n job candidates may solicit feedback from colleagues regarding each candidate's abilities.

indicates that Peer t is uncertain about which option is best. The DM aggregates the recommendations into a mass distribution m_0 over states:

$$m_0(E) = \frac{1}{k}\sum_{t=1}^{k} \mathbf{1}_{(o^t=E)}, \tag{6}$$

for all $E \subseteq \Omega$.[5]

Suppose that collecting an additional observation o^{k+1} entails a cost c^{k+1}. Should the DM obtain o^{k+1} or stop and make a choice based on the current mass distribution m_0? Ex ante, the additional observation defines an information structure $I = \big(S, M, \{m_s\}_{s\in S}\big)$, where the set of signals $S = \{s_E\}_{E\in\mathcal{P}(\Omega)\setminus\emptyset}$ corresponds to the possible recommendations of Peer $k+1$. For each signal s_E, the associated posterior m_{s_E} is formed according to Equation (6), with k replaced by $k+1$. It is natural to assume that $M(\{s_E\}) = \frac{k}{N}m_0(E)$ and $M(S) = \frac{N-k}{N}$, where N denotes the total number of potential observations (e.g., the total number of peers). In particular, when both ϕ and ψ have CARA forms, the DM should obtain the next observation o^{n+1} if and only if $W(I) - c^{n+1} - \hat{V}(m_0) > 0$.

The final example in this section shows how an information structure can be applied to an experimental setup of a laboratory experiment.

Example 4 (Ellsberg Urns)**.** Consider two urns filled with green and blue balls. The exact compositions of the urns are only partially known. The DM's task is to identify which urn is the true urn.[6] The DM can obtain information that reveals the color of a randomly drawn ball from the true urn. What is the value of this information?

Formally, define the states $\Omega = \{\omega_1, \omega_2\}$, outcomes $X = \{x_0, x_1\}$ with $u(x_0) = 0$ and $u(x_1) = 1$, and actions $A = \{a_1, a_2\}$, where each action a_i represents a bet on state ω_i as in the previous example. Define the set of signals as $S = \{g, b\}$. The two states correspond to the urns, Urn I and Urn II, while the two signals represent the possible colors of the randomly drawn ball. For concreteness, suppose Urn I contains 6 green balls, 3 blue balls, and 1 ball of unknown color (blue or green), while Urn II contains 2 green balls, 5 blue balls, and 3 balls of unknown color (blue or green), as illustrated in Figure 2.

The prior mass distribution is given by $m_0(\{\omega_1\}) = m_0(\{\omega_2\}) = \frac{1}{2}$ and

[5] Equation (6) assumes that all peers' recommendations carry equal weight, though this assumption could easily be relaxed.

[6] For a more practical interpretation, urns can be thought of as competing hypotheses.

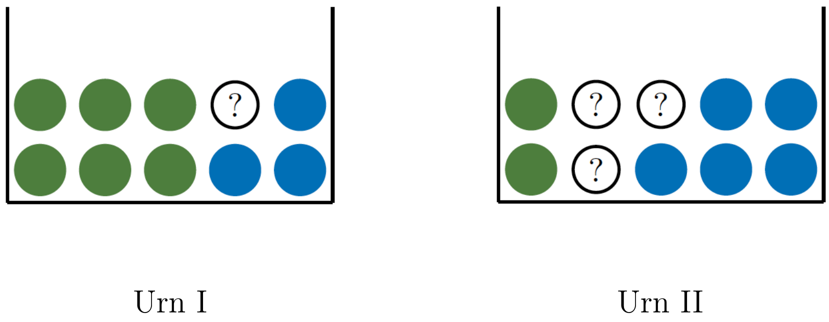

Figure 2: Partially known urn compositions

$m_0(\{\omega_1, \omega_2\}) = 0$.[7] Note that $m_0(\{\omega_1, \omega_2\}) > 0$ would indicate the presence of balls for which the DM does not know in which urn they are located. Similarly, $M(\{g\}) = M(\{b\}) = \frac{8}{20}$ and $M(\{g, b\}) = \frac{4}{20}$, where $M(\{g, b\}) > 0$ represents balls of unknown color. Accordingly, the prior mass distribution on $\mathcal{P}(\Omega \times S)$ is summarized in Table 3.[8]

state events	signal events		
	$\{g\}$	$\{b\}$	$\{g, b\}$
$\{\omega_1\}$	$\frac{6}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
$\{\omega_2\}$	$\frac{2}{20}$	$\frac{5}{20}$	$\frac{3}{20}$
$\{\omega_1, \omega_2\}$	0	0	0

Table 3: Mass distribution on $\mathcal{P}(\Omega \times S)$

To compute m_g and m_b, we must update the mass distribution on $\Omega \times S$ with respect to the events $\{(\omega_1, g), (\omega_2, g)\}$ and $\{(\omega_1, b), (\omega_2, b)\}$. Table 4 shows the results of applying the three updating rules from Section 2.3.2 to our example. Consider the left table in more detail. Under the Bayesian rule, only the balls known with certainty to be green (8 balls) are taken into account, while all others are ignored. The Dempster–Shafer rule, in contrast, treats all balls of unknown color (4 balls) as if they were green. The Generalized Bayesian rule adopts a more cautious stance: it assumes that unknown balls in the current urn are not

[7]For simplicity, we assume that the prior mass distribution is proportional to the number of balls each urn contains.

[8]Note that $\mathcal{P}(\Omega \times S)$ contains 15 non-empty subsets (events), while Table 3 displays only 9 of them. The richness of $\mathcal{P}(\Omega \times S)$ allows for the modeling of more complex information structures. For instance, assigning positive mass to the event $\{(\omega_1, g), (\omega_2, b)\}$ corresponds to balls for which it is only known that they are either in Urn I and green, or in Urn II and blue.

state events	updating rule		
	B	DS	GB
$\{\omega_1\}$	$\frac{6}{8}$	$\frac{7}{12}$	$\frac{6}{11}$
$\{\omega_2\}$	$\frac{2}{8}$	$\frac{5}{12}$	$\frac{2}{9}$
$\{\omega_1, \omega_2\}$	0	0	$\frac{23}{99}$

Signal g (green ball)

state events	updating rule		
	B	DS	GB
$\{\omega_1\}$	$\frac{3}{8}$	$\frac{4}{12}$	$\frac{3}{11}$
$\{\omega_2\}$	$\frac{5}{8}$	$\frac{8}{12}$	$\frac{5}{9}$
$\{\omega_1, \omega_2\}$	0	0	$\frac{17}{99}$

Signal b (blue ball)

Table 4: Posterior mass distributions after observing a signal under the Bayesian, Dempster–Shafer, and Generalized Bayesian rules

green, while those in the opposite urn are green. This procedure leaves a non-zero residual uncertainty mass displayed in the bottom row of the table.

For the analysis that follows, we focus on the Generalized Bayesian rule. Applying Equations (3) and (4), we obtain $\hat{V}(m_0) = \frac{1}{2}$, $\hat{V}(m_g) = \frac{6}{11} + \frac{23}{99}\bar{u}$, and $\hat{V}(m_b) = \frac{5}{9} + \frac{17}{99}\bar{u}$, where $\bar{u} = \phi^{-1}\left(\frac{1}{2}\phi(0) + \frac{1}{2}\phi(1)\right)$. Using Equation (5), we then have

$$W(I) = \frac{2}{5}\hat{V}(m_g) + \frac{2}{5}\hat{V}(m_b) + \frac{1}{5}\bar{V},$$

with $\bar{V} = \psi^{-1}\left(\frac{1}{2}\psi(\hat{V}(m_g)) + \frac{1}{2}\psi(\hat{V}(m_b))\right)$. For an ambiguity averse DM with $\bar{u} = \frac{1}{4}$, both $\hat{V}(m_g)$ and $\hat{V}(m_b)$ are approximately $\frac{3}{5}$. Consequently, $W(I) \approx \frac{3}{5}$ and $W(I) - \hat{V}(m_0) \approx \frac{1}{10}$. The value of information decreases as the DM becomes more averse to outcome uncertainty and as the information becomes more incomplete.[9]

4 Informativeness with partial information

In the framework of expected utility theory, Blackwell (1951) provides a concept for comparing information structures according to their informativeness. In particular, Blackwell shows that informativeness of an experiment corresponds to statistical sufficiency of the experiment, i.e., compared to the superior experiment, the inferior experiment yields only additional noise.

For non-expected utility frameworks, the notion of informativeness of an infor-

[9] To see how additional uncertainty affects the results, suppose that, beyond the 20 original balls in the urns of Figure 2, 10 more balls are introduced, whose colors and urn assignments are completely unknown to the DM. These additional balls are represented by assigning a mass of $\frac{10}{30}$ to the universal event $\Omega \times S$, while the masses of all events in Table 3 are adjusted proportionally. Reapplying Equations (3), (4) and (5), we find that, depending on the DM's attitude toward outcome uncertainty (captured by $\bar{u}$), the information structure I may in fact be rejected: $W(I) - \hat{V}(m_0) < 0$.

mation structure needs to be reconsidered (see Grant *et al.*, 1998; Karni and Safra, 2022; Li and Zhou, 2016; Schlee, 1990; Wakker, 1988). If beliefs are represented by mass distributions, new information no longer concerns only updating properties but requires also a notion of precision for signals. Consider two information structures:

$$I_S = \left(S, M_S, \{m_s\}_{s\in S}\right) \text{ and } I_T = \left(T, M_T, \{m_t\}_{t\in T}\right).$$

Partial information and ambiguity enters the information structures I_S and I_T at two stages:

(i) Ex post, once beliefs over states have been updated in the light of the signals from S, respectively T, leading to modified mass distributions $\{m_s\}_{s\in S}$ and $\{m_t\}_{t\in T}$.

(ii) Ex ante, regarding the relative precision of the mass distributions M_S and M_T when the signal has not yet been realized.

In this section, we suggest a notion of informativeness for information structures under partial information and prove a Blackwell-like equivalence result. Since we allow for partial information at both stages of the process, the suggested notion of informativeness will comprise two conditions. Condition (i) relates to the ex post stage when the signal is known. It expresses informativeness in terms of posterior mass distributions. Condition (ii) concerns the ex ante stage and compares information about the signal. For this comparison of the partial information about the likelihood of the signal, we need a notion of precision for the signal. The concept of the centroid of the core of the capacity μ^M, that corresponds to the mass distribution (Möbius transform) M provides us with an appropriate notion.

4.1 Partial information and the centroid of core(μ^M)

Partial information about the mass distribution M over events is reflected in the core of its associated belief function μ^M. The core of the belief function μ^M that is associated with the mass distribution M contains all probability distributions p yielding, for each event $E \subseteq S$, a probability $p(E) = \sum_{s\in E} p(s) \geq \mu^M(E) = \sum_{F\subseteq E} M(F)$. If a mass distribution is concentrated on the singleton events, the core consists of a single probability distribution. Otherwise, the core is a polyhedron with extreme points reflecting the partial information about the probability distributions.

For a mass distribution M, the *centroid* is itself a probability distribution P_M in the core of the capacity μ^M:

$$P_M(s) = \sum_{E \ni s} \frac{1}{|E|} M(E)$$

for all $s \in S$.[10] The centroid P_M of a mass distribution M associates with each signal s the uniform average of the masses $M(E)$ of all events E that contain s. By construction the centroid is a probability distribution that reflects the relative weights of the information about non-singleton events.

4.2 Comparing information structures

The following definition of relative informativeness is inspired by the respective definitions in Grant *et al.* (1998) and Karni and Safra (2022). These approaches work within a framework of joint probability distributions over the product space of states and signals. Hence, they assume as primitive concept a joint probability distribution on $S \times \Omega$ and derive both the prior distribution over signals and the posterior probability distributions conditional on the realized signal from it.

In context of mass distributions reflecting partial information of the DM, we characterize "informativeness" of an information structure by (i) a condition on the signal-updated mass distributions $\{m_s\}_{s \in S}$ and (ii) a condition on the dispersedness of the mass distribution M over signals. In order to formalize the latter condition, we use the centroid of the belief function μ^M as a measure of diffuseness of the partial information over signals.

Definition 1. Take two information structures $I_S = \left(S, M_S, \{m_s\}_{s \in S}\right)$ and $I_T = \left(T, M_T, \{m_t\}_{t \in T}\right)$. We say that I_S is *more informative* than I_T if there exist numbers $\beta_{st} \geq 0$ such that $\sum_{s \in S} \beta_{st} = 1$ for all $t \in T$,
(i) $\sum_{s \in S} \beta_{st} m_s = m_t$ for all $t \in T$, and
(ii) $\sum_{t \in T} \beta_{st} P_{M_T}(t) = P_{M_S}(s)$ for all $s \in S$.

For a better understanding of Definition 1, consider the special case with no ambiguity, where all mass distributions are concentrated on singletons (i.e., the corresponding belief functions are additive). Then, for any signal $s \in S$, m_s is the (posterior) probability distribution over states $\omega \in \Omega$ given signal $s \in S$, $\sum_{\omega \in \Omega} m_s(\{\omega\}) = 1$. The centroid P_{M_S} is the (prior) probability distribution over signals $s \in S$ since, for additive belief functions, the centroid P_{M_S} effectively

[10] The centroid is also known as the Steiner point of the core. In co-operative game theory, the centroid is known as the Shapley value. For convex games, e.g., belief functions, the Shapley value is an element of the core (Shapley, 1971). Miranda and Montes (2023) provide a careful study of the centroid and compare it to other central distributions in the core of a capacity.

coincides with M_S. The only difference is that Definition 1 is stated in terms of posteriors rather than the likelihoods of signals given states.

Remark 2. Suppose there is a joint probability distribution Q over the product space $S \times \Omega$ of signals and states.[11] Then the Bayesian updates of the signals $s \in S$ on the state space Ω, $Q(\omega|s) = \frac{Q(s,\omega)}{\sum_{\omega \in \Omega} Q(s,\omega)} = \frac{Q(s,\omega)}{Q(s)}$ are the signal-contingent probability distributions over states, where $Q(s)$ denotes the marginal distribution on S. Similarly, $Q(s|\omega) = \frac{Q(s,\omega)}{\sum_{s \in S} Q(s,\omega)} = \frac{Q(s,\omega)}{Q'(\omega)}$ is the likelihood distribution over signals given state $\omega \in \Omega$, where $Q'(\omega)$ denotes the marginal distribution on Ω. Hence, given a joint probability distribution Q over states and signals, we can define an information structure $I = \left(S, M, \{m_s\}_{s \in S}\right)$ by $M(s) = Q(s)$ and $m_s(\omega) = Q(\omega|s)$.

In the context of probability distributions and Bayesian updating, Blackwell (1951) and Cremer (1982) define one information structure as more informative than (or "sufficient for") another if it satisfies a condition analogous to Definition 1, formulated in terms of the likelihood distributions of signals conditional on states. Their analysis assumes that DMs are expected-utility maximizers, which allows them to establish that one information structure is more informative than another if and only if it is more valuable for the DM. More recently, Karni and Safra (2022) demonstrate that being more informative in the sense of Blackwell (1951) is no longer necessary and sufficient when agents deviate from expected utility maximization.[12]

Below we will prove a Blackwell-type equivalence theorem within our framework, which significantly generalizes the standard expected-utility setting. In this context, we allow for ambiguity both at the stage of updated mass distributions and at the stage of mass distributions over signals. We obtain an equivalence theorem for the case in which DMs are ambiguity neutral with respect to the mass distribution on the signal space, i.e., when ψ is linear. Importantly, in this case no restriction is required on the ambiguity attitude ϕ used for the evaluation of the signal-contingent mass distributions $\{m_s\}_{s \in S}$. In other words, the theorem assumes ambiguity-neutrality with respect to signal-related ambiguity, while imposing no restriction on the degree of ambiguity attitude regarding outcome uncertainty. Moreover, we do not confine the evaluation of uncertainty to the

[11]Equivalently to a joint probability distribution, one can also assume as a priori given (i) a set of conditional probability distributions $\{\mu(s|\omega)\}_{\omega \in \Omega}$ plus (ii) a prior distribution $\pi(\omega)$ over the states as in Karni and Safra (2022, p. 2).

[12]Introducing the notion of "hybrid valuable", Karni and Safra (2022, p. 5) show that the equivalence between informativeness and value can be restored for dynamic expected utility preferences that fail to satisfy the compound lotteries axiom.

quasi-average utility $U_{\phi u}(e_C)$. Instead, we allow for an arbitrary evaluation rule $U(e_C)$ under full ignorance with respect to an event $C \subseteq X$. Such additional generality of the DM's utility function will allow us to show that a more valuable signal is also necessarily more informative in the sense of Definition 1.

A *decision problem* is a pair (A, U), where A is a finite feasible set of actions and U is a real function on the set of elementary mass distributions $\{e_C \mid C \subseteq X\}$. Given a decision problem (A, U), the value $W(I)$ of an information structure $I = \left(S, M, \{m_s\}_{s \in S}\right)$ is obtained by aggregating the values

$$\hat{V}(m_s) = \max_{a \in A} \sum_{C \subseteq X} (m_s * a)(C)\, U(e_C)$$

of the signals with the quasi-average of the prior mass distribution M over signals S as in Equation (5).

Theorem 1. *Under a linear ψ, an information structure I_S is more informative than another information structure I_T if and only if $W(I_S) \geq W(I_T)$ for any decision problem.*

The proof of Theorem 1 is contained in Appendix A.

In case that M_S, M_T and all $\{m_s\}_{s \in S}$ and $\{m_t\}_{t \in T}$ are all concentrated on singletons, Theorem 1 reduces to the standard Blackwell's informativeness theorem.

Remark 3. One may consider stronger comparative notions of informativeness than the one suggested in Definition 1. Such stronger notions may imply that the more informative structure is also more valuable for arbitrary ambiguity attitude ψ regarding the signal. One such stronger definition and the associated result is presented in Appendix B.

The following example illustrates the relationship between informativeness and the value of partial information.

Example 5 (Informativeness)**.** Take Ω, X, and A as in Example 4. Let $I_S = \left(S, M_S, \{m_s\}_{s \in S}\right)$ with $S = \{s_1, s_2\}$, $M_S(\{s_1\}) = M_S(\{s_2\}) = \frac{1}{2}$, and $I_T = \left(T, M_T, \{m_t\}_{t \in T}\right)$ with $T = \{t\}$ and $M(T) = 1$. The posterior mass distributions are given in Table 5. Clearly, $\frac{1}{2}m_{s_1} + \frac{1}{2}m_{s_2} = m_t$ and $\frac{1}{2}P_{M_T}(t) = P_{M_S}(s_1) = P_{M_S}(s_2)$, so that I_S is more informative than I_T. Comparing the values of the two structures, we get $W(I_S) = \frac{1}{2} + \frac{1}{3}U(e_{\{x_1,x_2\}})$ and $W(I_T) = \frac{1}{3} + \frac{1}{3}U(e_{\{x_1,x_2\}})$, so that $W(I_S) \geq W(I_T)$ for any $U \in \mathcal{U}$.

Finally, one may also compare information structures by the ambiguity of their associated mass distributions. For instance, Eichberger and Pasichnichenko (2021)

	m_{s_1}	m_{s_2}	m_t
$\{\omega_1\}$	$\frac{2}{3}$	0	$\frac{1}{3}$
$\{\omega_2\}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
$\{\omega_1, \omega_2\}$	0	$\frac{2}{3}$	$\frac{1}{3}$

Table 5: Posterior mass distributions in Example 5

propose a model-free approach to comparing mass distributions in terms of ambiguity, based on the notion of an elementary increase in ambiguity. We can then establish the following. Suppose two information structures $I = \big(S, M, \{m_s\}_{s\in S}\big)$ and $I' = \big(S, M', \{m_s\}_{s\in S}\big)$ share the same posteriors $\{m_s\}_{s\in S}$, but M' is more ambiguous than M. Then $W(I) \geq W(I')$ if and only if ψ is concave. Likewise, consider two information structures $I = \big(S, M, \{m_s\}_{s\in S}\big)$ and $I' = \big(S, M, \{m'_s\}_{s\in S}\big)$ that share the same M and all posteriors except one for $s^* \in S$, where m'_{s^*} is more ambiguous than m_{s^*}. In this case, one obtains $W(I) \geq W(I')$ if and only if ϕ is concave.

5 Preferences over information structures and the value of information

So far, we have derived a value for an information structure by using the two-stage valuation in Equations (3)–(5). These formulas were obtained in analogy to the standard way of evaluating information structures suggested in Blackwell (1951) for the case of probability distributions and expected-utility-maximizing agents. In this section, we will provided an axiomatic derivation of the valuation functional in Equations (3)–(5) based on the DM's preferences over information structures.

Consider a universal set of signals $\mathbb{S}$, such that for any information structure $I = \big(S, M, \{m_s\}_{s\in S}\big)$, we have $S \subseteq \mathbb{S}$. Without loss of generality, assume that each signal induces a unique posterior mass distribution over states, independent of the information structure in which it appears. Formally, for any $s, t \in \mathbb{S}$, $s = t$ if and only if $m_s = m_t$. Consequently, an information structure is fully characterized by its mass distribution M over signals in $\mathbb{S}$. Moreover, we fix a finite feasible set $A \subseteq \mathcal{A}$ of actions.

For an arbitrary set K, let $\mathcal{F}_K$ denote the collection of finite subsets of K. Let $\mathcal{E}_K$ be an algebra of subsets of K that contains all finite subsets. Define $\mathcal{M}(K)$ as the set of mass distributions on $\mathcal{E}_K$ that are concentrated on a finite subset. In other words, for any $m \in \mathcal{M}(K)$, there exist finitely many sets $E_1, \ldots, E_n \in \mathcal{F}_K$ such that $\sum_{i=1}^n m(E_i) = 1$. A special case of such a mass distribution is a finitely supported probability measure (a lottery) on $\mathcal{E}_K$. For notational convenience, we do not distinguish between an element $k \in K$ and the degenerate lottery in $\mathcal{M}(K)$ concentrated on the singleton $\{k\}$. A preference relation on $\mathcal{M}(K)$ is called a *quasi-average preference*, if it satisfies Axioms 1–7 of Eichberger and Pasichnichenko (2021).[13]

In particular, $\mathcal{M}(X)$ denotes the set of mass distributions over outcomes, while $\mathcal{M}(\mathbb{S})$ denotes the set of mass distributions over signals. By assumption, sets X and $\mathbb{S}$ are sufficiently rich to allow for countably many distinct certainty equivalents for any lottery in $\mathcal{M}(X)$ and $\mathcal{M}(\mathbb{S})$, respectively.

We assume that the DM has a preference relation $\succcurlyeq$ on $\mathcal{M}(X)$, representing preferences over mass distributions over outcomes. Similarly, the DM has a preference relation $\succcurlyeq^*$ on $\mathcal{M}(\mathbb{S})$, representing preferences over information structures. Both relations are assumed to possess a quasi-average preference structure:

Assumption 1. The preference relation $\succcurlyeq$ on $\mathcal{M}(X)$ is a quasi-average preference.

Assumption 2. The preference relation $\succcurlyeq^*$ on $\mathcal{M}(\mathbb{S})$ is a quasi-average preference.

Quasi-average preferences combine partially known likelihoods of outcomes and signals with subjective beliefs according to the principle of insufficient reason whenever the likelihood of an event is unknown. In the special case, when likelihoods are known, these preferences coincide with expected utility maximization. At the opposite extreme, under complete ignorance, evaluation is based on a uniform probability distribution and a transformation function that captures the DM's attitude toward ambiguity.[14]

The following assumptions establish a connection between the two preference relations, $\succcurlyeq$ and $\succcurlyeq^*$. To this end, we first extend $\succcurlyeq^*$ to mass distributions over Ω by means of degenerate information structures. For any $m, m' \in \mathcal{M}(\Omega)$, let $m \succcurlyeq^* m'$

[13] Axioms 1–3 have the same interpretation as in the standard von Neumann–Morgenstern framework, applied to a preference relation on $\mathcal{M}(K)$. Axioms 4–7 characterize the DM's ordering of the elementary mass distributions $\{e_E \mid E \in \mathcal{F}_K\}$, which correspond to situations of complete ignorance restricted to E. These axioms imply that each elementary mass distribution e_E is evaluated by the quasi-average $V(e_E) = \phi^{-1}\left(\frac{1}{|E|}\sum_{k\in E}\phi\left(u\left(k\right)\right)\right)$ in the spirit of the principle of insufficient reason. Combined with Axioms 1–3, this leads to the quasi-average utility representation $V(m) = \sum_{E\subseteq K} m(E)V(e_E)$ for all $m \in \mathcal{M}(K)$.

[14] Eichberger and Pasichnichenko (2021) provide a more detailed discussion.

if there exist mass distributions $M, M' \in \mathcal{M}(\mathbb{S})$ such that $M(\{s\}) = M'(\{t\}) = 1$, $m_s = m$, $m_t = m'$, and $M \succcurlyeq^* M'$. Intuitively, a degenerate information structure M with $M(\{s\}) = 1$ corresponds to a situation with prior mass distribution m_s and a completely uninformative signal.

Given a preference relation $\succcurlyeq^*$ extended in this way, we can now link the preference relation $\succcurlyeq^*$ on $\mathcal{M}(\Omega)$ to the preference relation $\succcurlyeq$ on $\mathcal{M}(X)$ through the actions A and the induced mass distributions.

Assumption 3. For any $m, m' \in \mathcal{M}(\Omega)$, we have $m \succcurlyeq^* m'$ if and only if there exists $\widetilde{a} \in A$ such that $m * \widetilde{a} \succcurlyeq m' * a$ for all $a \in A$.

In other words, m is preferred to m' if the DM can induce a more favorable mass distribution $m * \widetilde{a} \in \mathcal{M}(X)$ over outcomes through her choice of action $\widetilde{a}$ under m than under m'.

Finally, we link the DM's risk attitude with respect to signals and her risk attitude with respect to outcomes. To begin with, observe that for any signal s and action a, the DM is indifferent between the mass distribution $m_s * a \in \mathcal{M}(X)$ and its *certainty equivalent* denoted by $c(m_s * a) \in X$. By Assumption 1, such a certainty equivalent always exists.[15] When the DM observes a signal s, she selects an optimal action a_s, leading to the certainty equivalent $c(m_s * a_s)$. Thus, when signals are distributed according to $M \in \mathcal{M}(\mathbb{S})$, facing M is like facing a mass distribution $T_M \in \mathcal{M}(X)$ concentrated on the certainty equivalents $c(m_s * a_s)$ for all s in the support of M:

Definition 2. For $M \in \mathcal{M}(\mathbb{S})$, define $T_M \in \mathcal{M}(X)$ as the mass distribution over certainty equivalents induced by M, given by

$$T_M(C) = M\left(\tau^{-1}(C)\right)$$

for all $C \in \mathcal{E}_X$ and some injective mapping $\tau : \mathbb{S} \to X$ such that $\tau(s) = c(m_s * a_s)$ for all $s \in \mathbb{S}$.

Richness of X guarantees that it is always possible to find distinct certainty equivalents $c(m_s * a_s)$ and $c(m_t * a_t)$.[16] Consequently, such an injective mapping τ exists. Although τ is not unique, any two choices of τ yield mass distributions on X that are equivalent for the DM (Assumption 1).

[15]Since $m_s * a$ is concentrated on a finite set, it follows from the representation of $\succcurlyeq$ that $x^0 \succcurlyeq m_s * a \succcurlyeq x_0$ for some $x^0, x_0 \in X$. Hence, the DM is indifferent between $m_s * a$ and a lottery over x^0 and x_0, which itself possesses a certainty equivalent by our structural assumption.

[16]Even if the DM is indifferent between $m_s * a_s$ and $m_t * a_t$.

When the mass distributions M and M' are concentrated on singletons (so that they are effectively lotteries), it is natural to assume that the DM ranks the corresponding induced outcome distributions in the same way.

Assumption 4. For all $M, M' \in \mathcal{M}(\mathbb{S})$ concentrated on singletons, $M \succsim^* M'$ if and only if $T_M \succsim T_{M'}$.

Hence, the DM exhibits the same risk attitudes for preferences over information structures as for preferences over lotteries on outcomes. However, this assumption does not impose a restriction on her ambiguity attitudes across the two domains.

The following theorem shows that Assumptions 1–4 characterize the preference representation over information structures proposed in Equations (3)–(5).

Theorem 2. *The DM's preferences satisfy Assumptions 1–4 if and only if there exists a representation W of the preference relation $\succsim^*$ on the set of information structures such that*

$$W(I) = \sum_{E \subseteq S} M(E)\, \psi^{-1}\left(\frac{1}{|E|} \sum_{s \in E} \psi\left(\hat{V}\left(m_s\right)\right)\right),$$

where

$$\hat{V}\left(m_s\right) = \max_{a \in A} \sum_{C \subseteq X} \left(m_s * a\right)(C)\, U_{\phi u}(e_C),$$

u is a von Neumann–Morgenstern utility function on X, and ψ and ϕ are continuous and strictly increasing functions on their respective domains. The representation W is unique up to a positive linear transformation. Given W, the functions u and $\hat{V}$ are uniquely determined, while ψ and ϕ are unique up to positive linear transformations.

We prove Theorem 2 in Appendix C.

Notice that the ambiguity attitudes captured by ψ and ϕ in Theorem 2 need not be identical. In Appendix D, we present two alternative representations—one more general than Theorem 2 and one more specific. The more general representation is obtained by dropping Assumption 4. In this case, $\hat{V}\left(m_s\right)$ is not necessarily equal to $\max_{a \in A} \sum_{C \subseteq X} \left(m_s * a\right)(C)\, U_{\phi u}(e_C)$, but is related to it through a strictly increasing transformation. The more specific representation results from strengthening Assumption 4 by allowing mass distributions that are not necessarily concentrated on singletons. Under this stronger assumption, we obtain a representation in which the DM exhibits identical ambiguity attitudes toward uncertainty about signals and uncertainty about outcomes.

6 Concluding remarks

We present a general model of decision making over information structures, in which signals are evaluated based on the beliefs they induce over states of the world. Partial uncertainty (both about signals and about states) is modeled by belief functions and the value of information is assessed through a quasi-average utility of the induced outcomes. When there is uncertainty about the signal generating process, the value of information may be negative.

We propose the centroid of the belief function as a measure of the informativeness of signals and prove a result similar to Blackwell's theorem in this framework. Finally, we show that the representation of the value of information by a two-stage quasi-average utility is derived from the DM's preference order over information structures.

References

Abdellaoui, M., Hill, B., Kemel, E. and Maafi, H. (2025). Learning under ambiguity: An experimental investigation. *Journal of Economic Theory*, **230**, 106093.

Blackwell, D. (1951). Comparison of experiments. In J. Neymann (ed.), *Proceedings of Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, pp. 93–102.

— (1953). Equivalent comparisons of experiments. *Annals of Mathematical Statistics*, **24**, 265–272.

— and Girshick, M. (1954). *Theory of Games and Statistical Decisions*. New York: Dover Publications.

Çelen, B. (2012). Informativeness of experiments for MEU. *Journal of Mathematical Economics*, **8**, 404–406.

Chateauneuf, A. and Jaffray, J.-Y. (1989). Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, **17** (3), 263–283.

Cheng, X., Klibanoff, P., Mukerji, S. and Renou, L. (2025). *Persuasion with Ambiguous Communication.* Working paper, https://arxiv.org/abs/2410.05504.

Cremer, J. (1982). A simple proof of Blackwell's comparison of experiments theorem. *Journal of Economic Theory*, **27**, 439–443.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, **38** (2), 325–339.

Dubois, D. and Prade, H. (1990). Updating with belief functions, ordinal conditional functions and possibility measures. In *UAI'90: Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, Elsevier, pp. 311–330.

Eichberger, J., Grant, S. and Kelsey, D. (2010). Comparing three ways to update choquet beliefs. *Economics Letters*, **107**, 91–94.

— and Pasichnichenko, I. (2021). Decision-making with partial information. *Journal of Economic Theory*, **198**, 105369.

Gilboa, I. and Schmeidler, D. (1993). Updating ambiguous beliefs. *Journal of Economic Theory*, **59**, 33–49.

Gonzalez-Jimenez, D. (2024). Learning gaps? Described versus experienced signals. *Working Paper*.

Grabisch, M. (2016). *Set Functions, Games and Capacities in Decision Making*, *Theory and Decision Library C*, vol. 46. Springer.

Grant, S., Kajii, A. and Polak, B. (1998). Intrinsic preference for information. *Journal of Economic Theory*, **83** (2), 233–259.

Hendon, E., Jacobsen, H. J., Sloth, B. and Tranaes, T. (1996). The product of capacities and belief functions. *Mathematical Social Sciences*, **32**, 95–108.

Heyen, D. and Wiesenfarth, B. R. (2015). Informativeness of experiments for MEU – a recursive definition. *Journal of Mathematical Economics*, **57**, 28–30.

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

Karni, E. and Safra, Z. (2022). Hybrid decision model and the ranking of experiments. *Journal of Mathematical Economics*, **101**, 102700.

Kops, C. and Pasichnichenko, I. (2023). Testing negative value of information and ambiguity aversion. *Journal of Economic Theory*, **213**, 105730.

Li, J. (2020). Preferences for partial information and ambiguity. *Theoretical Economics*, **15**, 1059–1094.

— and Zhou, J. (2016). Blackwell's imformativeness ranking with ambiguity averse preferences. *Games and Economic Behavior*, **96**, 18–29.

Lin, Y.-H. and Payró, F. (2024). Updating under imprecise information. *Barcelona School of Economics Working Papers*, **1424**.

Miranda, E. and Montes, I. (2023). Centroids of the core of exact capacities: a comparative study. *Annals of Operations Research*, **321**, 409–449.

Savage, L. J. (1954). *Foundations of Statistics*. New York: Wiley.

Schlee, E. (1990). The value of information in anticipated utility theory. *Journal of Risk and Uncertainty*, **3**, 83–92.

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

Shapley, L. S. (1971). Cores of convex games. *International Journal of Game Theory*, **1**, 11–26.

Shishkin, D. and Ortoleva, P. (2023). Ambiguous information and dilation: An experiment. *Journal of Economic Theory*, **208**, 105610.

Wakker, P. (1988). Nonexpected utility as aversion to information. *Journal of Behavioral Decision Making*, **1**, 169–175.

Walley, P. (1996). Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **58** (1), 3–34.

Wang, Z. (2024). Informativeness orders over ambiguous experiments. *Journal of Economic Theory*, **222**, 105937.

Appendix

A. Proof of Theorem 1

Note that for a given information structure $I = \left(S, M, \{m_s\}_{s\in S}\right)$ and a given decision problem, we obtain

$$\begin{aligned} W(I) &= \sum_{E\subseteq S} M\,(E)\psi^{-1}\left(\frac{1}{|E|}\sum_{s\in E}\psi\,(V(a_s, m_s))\right) \\ &= \sum_{E\subseteq S} M\,(E)\,\frac{1}{|E|}\sum_{s\in E} V(a_s, m_s) \\ &= \sum_{s\in S} P_M(s)V(a_s, m_s), \qquad (7) \end{aligned}$$

where a_s denotes a maximizer of $V(a, m_s)$ in the decision problem, the second equality follows from the linearity of ψ, and the third from the definition of the centroid. Furthermore, for any action a, one can express $V(a, m_s)$ as

$$\begin{aligned} V(a, m_s) &= \sum_{C\subseteq X} (m_s * a)\,(C)\; U(e_C) \\ &= \sum_{o\in\mathcal{O}} m_s(o)U(e_{a(o)}), \qquad (8) \end{aligned}$$

where $\mathcal{O}$ denotes the finite power set of Ω.

(i) "$\Rightarrow$" If $I_S = \left(S, M_S, \{m_s\}_{s\in S}\right)$ is more informative than $I_T = \left(T, M_T, \{m_t\}_{t\in T}\right)$, then $W(I_S) \geq W(I_T)$ for any decision problem:

For a decision problem (A, U) and for all $a \in A$, we obtain

$$
\begin{aligned}
W(I_S) &= \sum_{s\in S} P_{M_S}(s)V(a_s, m_s) \\
&\geq \sum_{s\in S} P_{M_S}(s)V(a, m_s) \\
&= \sum_{s\in S} P_{M_S}(s)\sum_{o\in\mathcal{O}} m_s(o)U(e_{a(o)}) \\
&= \sum_{s\in S} \underbrace{\left[\sum_{t\in T}\beta_{st}P_{M_T}(t)\right]}_{=P_{M_S}(s) \text{ by Definition 1(ii)}} \sum_{o\in\mathcal{O}} m_s(o)U(e_{a(o)}) \\
&= \sum_{t\in T} P_{M_T}(t)\sum_{o\in\mathcal{O}} \underbrace{\left[\sum_{s\in S}\beta_{st}m_s(o)\right]}_{=m_t(o) \text{ by Definition 1(i)}} U(e_{a(o)}) \\
&= \sum_{t\in T} P_{M_T}(t)\sum_{o\in\mathcal{O}} m_t(o)U(e_{a(o)}) \\
&= \sum_{t\in T} P_{M_T}(t)V(a, m_t),
\end{aligned}
$$

where the first equality follows from Equation (7), the inequality from the fact that a_s is an optimal action given signal s, and both the third and final lines from Equation (8). Moreover, since $W(I_S) \geq \sum_{t\in T} P_{M_T}(t)V(a, m_t)$ for all $a \in A$, it follows in particular that $W(I_S) \geq \sum_{t\in T} P_{M_T}(t)V(a_t, m_t) = W(I_T)$. Thus, the inequality holds for any decision problem.

(ii) "$\Leftarrow$" If $W(I_S) \geq W(I_T)$ for any decision problem, then I_S is more informative than I_T:

We begin with the proof of the following lemma.

Lemma 1. *If an information structure $I_S = \big(S, M_S, \{m_s\}_{s\in S}\big)$ is not more informative than another information structure $I_T = \big(T, M_T, \{m_t\}_{t\in T}\big)$, then there is $d \in \mathbb{R}^{|T|\times|\mathcal{O}|}$ such that*

$$
\sum_{t\in T} P_{M_T}(t)\sum_{o\in\mathcal{O}} m_t(o)d_t(o) > \sum_{s\in S} P_{M_S}(s)\max_{t\in T}\sum_{o\in\mathcal{O}} m_s(o)d_t(o).
$$

Proof. For all $o \in \mathcal{O}$, define $\pi(o) = \sum_{s\in S} P_{M_S}(s)m_s(o)$. For all $s \in S$, $t \in T$, and $o \in \mathcal{O}$, set $\pi^S(s|o) = \frac{P_{M_S}(s)m_s(o)}{\pi(o)}$ and $\pi^T(t|o) = \frac{P_{M_T}(t)m_t(o)}{\pi(o)}$.

Let G denote the set of vectors $g \in \mathbb{R}^{|T|\times|\mathcal{O}|}$ such that $g_t(o) = \sum_{s\in S}\lambda_{st}\pi^S(s|o)$ for some numbers $\lambda_{st} \geq 0$ satisfying $\sum_{t\in T}\lambda_{st} = 1$. We identify π^T with a vector in $\mathbb{R}^{|T|\times|\mathcal{O}|}$. If I_S is not more informative than I_T, we will show that $\pi^T \notin G$.

Assume, to the contrary, that there exist numbers $\{\lambda_{st}\}$ as above. Then for all $t \in T$ and $o \in \mathcal{O}$, we obtain

$$\begin{aligned} m_t(o) &= \pi^T(t|o)\frac{\pi(o)}{P_{M_T}(t)} \\ &= \sum_{s\in S}\lambda_{st}\pi^S(s|o)\frac{\pi(o)}{P_{M_T}(t)} \\ &= \sum_{s\in S}\frac{\lambda_{st}P_{M_S}(s)}{P_{M_T}(t)}m_s(o) \\ &= \sum_{s\in S}\beta_{st}m_s(o), \end{aligned}$$

where $\beta_{st} = \frac{\lambda_{st}P_{M_S}(s)}{P_{M_T}(t)}$, which corresponds to Definition 1(i). Moreover, we have $\sum_{t\in T}\beta_{st}P_{M_T}(t) = P_{M_S}(s)\sum_{t\in T}\lambda_{st} = P_{M_S}(s)$, which corresponds to Definition 1(ii). Finally, the equality $\sum_{s\in S}\beta_{st} = 1$ follows from

$$\begin{aligned} P_{M_T}(t) &= \sum_{o\in\mathcal{O}}\pi^T(t|o)\pi(o) \\ &= \sum_{s\in S}\lambda_{st}\sum_{o\in\mathcal{O}}\pi^S(s|o)\pi(o) \\ &= \sum_{s\in S}\lambda_{st}P_{M_S}(s). \end{aligned}$$

Thus, the constructed numbers $\{\beta_{st}\}$ satisfy Definition 1, leading to a contradiction. Therefore, $\pi^T \notin G$.

Since $\pi^T \notin G$, the separating hyperplane theorem guarantees the existence of some $q \in \mathbb{R}^{|T|\times|\mathcal{O}|}$ such that

$$\sum_{t\in T}\sum_{o\in\mathcal{O}}q_t(o)\pi^T(t|o) > \sum_{t\in T}\sum_{o\in\mathcal{O}}q_t(o)\sum_{s\in S}\lambda_{st}\pi^S(s|o)$$

holds for all families $\{\lambda_{st}\}$ of numbers with $\lambda_{st} \geq 0$ and $\sum_{t\in T}\lambda_{st} = 1$. In particular, changing the order of summation on the right-hand side and considering a special choice of $\{\lambda_{st}\}$ yields

$$\sum_{t\in T}\sum_{o\in\mathcal{O}}q_t(o)\pi^T(t|o) > \sum_{s\in S}\max_{t\in T}\sum_{o\in\mathcal{O}}q_t(o)\pi^S(s|o).$$

By setting $d_t(o) = \frac{q_t(o)}{\pi(o)}$, we obtain

$$\sum_{t\in T}\sum_{o\in\mathcal{O}} d_t(o)P_{M_T}(t)m_t(o) > \sum_{s\in S}\max_{t\in T}\sum_{o\in\mathcal{O}} d_t(o)P_{M_S}(s)m_s(o),$$

which concludes the proof of the lemma. □

Let us now return to the proof of the theorem. We proceed by contradiction. Suppose $W(I_S) \geq W(I_T)$ for any decision problem, while I_S is *not* more informative than I_T. Define $A = \{a^t\}$ as a set of actions indexed by $t \in T$, such that all sets $\{a^t(o) \mid t \in T, o \in O\}$ are distinct. For each $t \in T$ and $o \in \mathcal{O}$, let the utility function be given by $U(e_{a^t(o)}) = d_t(o)$. For the decision problem (A, U), we obtain

$$\begin{aligned}
W(I_S) &= \sum_{s\in S} P_{M_S}(s)\max_{t\in T}\sum_{o\in\mathcal{O}} m_s(o)U(e_{a^t(o)}) \\
&< \sum_{t\in T} P_{M_T}(t)\sum_{o\in\mathcal{O}} m_t(o)U(e_{a^t(o)}) \\
&\leq \sum_{t\in T} P_{M_T}(t)\max_{t\in T}\sum_{o\in\mathcal{O}} m_t(o)U(e_{a^t(o)}) \\
&= W(I_T),
\end{aligned}$$

where the strict inequality follows from the lemma. This contradicts the assumption that $W(I_S) \geq W(I_T)$, thereby completing the proof of the theorem.

B. A stronger notion of informativeness

This section examines a stronger notion of informativeness that guarantees a higher value for any DM, regardless of their ambiguity attitude as represented by ψ. This is obtained by replacing Definition 1(ii) with a strong dominance condition on the cores of M_S and M_T. A drawback of this definition is that higher-valued information structures are not always more informative in the strong sense.

Definition 3. Take two information structures $I_S = \left(S, M_S, \{m_s\}_{s\in S}\right)$ and $I_T = \left(T, M_T, \{m_t\}_{t\in T}\right)$. We say that I_S is *strongly more informative* than I_T if there exist numbers $\beta_{st} \geq 0$ such that $\sum_{s\in S}\beta_{st} = 1$ for all $t \in T$,

(i) $\sum_{s\in S}\beta_{st}m_s = m_t$ for all $t \in T$, and
(ii) $\sum_{t\in T}\beta_{st}q(t) \leq p(s)$ for all $s \in S$, $q \in \text{core}\left(\mu^{M_T}\right)$, $p \in \text{core}\left(\mu^{M_S}\right)$.

We can now show that this notion of informativeness implies a higher value for all decision problems and ambiguity attitudes.

Theorem 3. *If an information structure I_S is strongly more informative than another information structure I_T, then $W(I_S) \geq W(I_T)$ for any decision problem and any continuous and strictly increasing function ψ.*

Proof. Consider a decision problem (A, U). From Definition 3(i), we obtain

$$\begin{aligned}
\hat{V}(m_t) &= \sum_{C \subseteq X} (m_t * a_t)(C) U(e_C) \\
&= \sum_{o \in \mathcal{O}} m_t(o) U(e_{a_t(o)}) \\
&= \sum_{o \in \mathcal{O}} \underbrace{\left[\sum_{s \in S} \beta_{st} m_s(o) \right]}_{= m_t(o) \text{ by Definition 3(i)}} U(e_{a_t(o)}) \\
&= \sum_{s \in S} \beta_{st} \sum_{o \in \mathcal{O}} m_s(o) U(e_{a_t(o)}) \\
&= \sum_{s \in S} \beta_{st} V(a_t, m_s) \\
&\leq \sum_{s \in S} \beta_{st} V(a_s, m_s) \\
&= \sum_{s \in S} \beta_{st} \hat{V}(m_s),
\end{aligned}$$

where the inequality follows from the fact that a_s is an optimal action given signal s. Moreover, since for any $E \subseteq S$ we have

$$\min_{s \in E} \hat{V}(m_s) \leq \psi^{-1}\left(\frac{1}{|E|} \sum_{s \in E} \psi\left(\hat{V}(m_s) \right) \right) \leq \max_{s \in E} \hat{V}(m_s),$$

it follows that

$$\psi^{-1}\left(\frac{1}{|E|} \sum_{s \in E} \psi\left(\hat{V}(m_s) \right) \right) = \sum_{s \in E} \alpha^E(s) \hat{V}(m_s)$$

for some probability distribution α^E on E. Hence, we can represent $W(I)$ as

$$
\begin{aligned}
W(I_S) &= \sum_{E\subseteq S} M_S(E)\psi^{-1}\left(\frac{1}{|E|}\sum_{s\in E}\psi\left(\hat{V}\left(m_s\right)\right)\right)\\
&= \sum_{E\subseteq S} M_S(E)\sum_{s\in E}\alpha^E(s)\hat{V}\left(m_s\right)\\
&= \sum_{s\in S}\underbrace{\sum_{E\ni s} M_S(E)\alpha^E(s)}_{p(s)}\hat{V}\left(m_s\right)\\
&= \sum_{s\in S} p(s)\hat{V}\left(m_s\right),
\end{aligned}
$$

where p is a probability distribution in the core of μ^{M_S}. Similarly,

$$
W(I_T) = \sum_{t\in T} q(t)\hat{V}\left(m_t\right)
$$

for some probability distribution q in the core of μ^{M_T}. Combining these results, we obtain

$$
\begin{aligned}
W(I_T) &= \sum_{t\in T} q(t)\hat{V}\left(m_t\right)\\
&\leq \sum_{t\in T} q(t)\sum_{s\in S}\beta_{st}\hat{V}\left(m_s\right)\\
&= \sum_{s\in S}\sum_{t\in T}\beta_{st}q(t)\hat{V}\left(m_s\right)\\
&\leq \sum_{s\in S} p(s)\hat{V}\left(m_s\right)\\
&= W(I_T),
\end{aligned}
$$

where the second inequality follows from Definition 3(ii), since all $\hat{V}\left(m_s\right)$ can, without loss of generality, be taken to be non-negative. Therefore, the inequality holds for any decision problem and any continuous strictly increasing function ψ. $\square$

C. Proof of Theorem 2

Since the preference relation $\succcurlyeq$ on $\mathcal{M}(X)$ is a quasi-average preference (Assumption 1), it follows from Theorem 9 in Eichberger and Pasichnichenko (2021) that there exist a von Neumann–Morgenstern utility function u on X and a continuous and strictly increasing function ϕ on $u(X)$ such that $\succcurlyeq$ is represented by $\sum_{C \subseteq X} n\,(C)\,U_{\phi u}(e_C)$ for all $n \in \mathcal{M}(X)$. Hence, for any $a, b \in A$ and $m, m' \in \mathcal{M}(\Omega)$, we have $m * a \succcurlyeq m' * b$ if and only if

$$\sum_{C \subseteq X} (m * a)\,(C)\,U_{\phi u}(e_C) \geq \sum_{C \subseteq X} (m' * b)\,(C)\,U_{\phi u}(e_C).$$

Combining this with Assumption 3, it follows that $m \succcurlyeq^* m'$ if and only if there exists $\widetilde{a} \in A$ such that

$$\sum_{C \subseteq X} (m * \widetilde{a})\,(C)\,U_{\phi u}(e_C) \geq \sum_{C \subseteq X} (m' * a)\,(C)\,U_{\phi u}(e_C)$$

for all $a \in A$. Therefore, the functional

$$H(m) = \max_{a \in A} \sum_{C \subseteq X} (m * a)\,(C)\,U_{\phi u}(e_C)$$

represents the preference relation $\succcurlyeq^*$ on $\mathcal{M}(\Omega)$, or, equivalently, on the set of degenerate information structures.

Because the preference relation $\succcurlyeq^*$ on $\mathcal{M}(\mathbb{S})$ is a quasi-average preference (Assumption 2), Theorem 9 in Eichberger and Pasichnichenko (2021) implies that there exist a von Neumann–Morgenstern utility function v on $\mathbb{S}$ and a continuous and strictly increasing function ψ on $v(\mathbb{S})$ such that $\succcurlyeq^*$ is represented by

$$\sum_{E \subseteq S} M\,(E)\,U_{\psi v}(e_E)$$

for all $M \in \mathcal{M}(\mathbb{S})$, where S denotes the finite subset on which M is concentrated.

For all $m \in \mathcal{M}(\Omega)$, define $\hat{V}(m) = v(s)$ for the unique $s \in \mathbb{S}$ such that $m_s = m$. Note that $\hat{V}(m) \geq \hat{V}(m')$ if and only if $m \succcurlyeq^* m'$. Hence, both $\hat{V}$ and H represent the same preference relation $\succcurlyeq^*$ on $\mathcal{M}(\Omega)$ and are therefore related by a strictly increasing transformation. Moreover, Assumption 4 implies that $\hat{V}$ and H capture identical risk attitudes over lotteries on Ω. Consequently, $\hat{V}$ must be a positive linear transformation of H, i.e., $\hat{V} = aH + b$ for some $a, b \in \mathbb{R}$, $a > 0$. Applying the same linear transformation to the functional representing $\succcurlyeq$ on $\mathcal{M}(X)$ and redefining H accordingly gives $\hat{V} = H$, establishing the desired representation.

The proof that the representation implies the assumptions is straightforward. The uniqueness result follows directly from Theorem 9 in Eichberger and Pasichnichenko (2021).

D. Alternative representations

If it is necessary to distinguish between the DM's risk attitudes toward signals and outcomes, a more general representation than that of Theorem 2 is required. Such a representation can be obtained simply by dropping Assumption 4.

Theorem 4. *The DM's preferences satisfy Assumptions 1–3 if and only if there exists a representation W_1 of the preference relation $\succsim^*$ on the set of information structures such that*

$$W_1(I) = \sum_{E \subseteq S} M(E)\, \psi^{-1}\left(\frac{1}{|E|} \sum_{s \in E} \psi\left(\hat{V}_1(m_s) \right) \right),$$

where $\hat{V}_1 : \mathcal{M}(\Omega) \to \mathbb{R}$ is related to

$$\hat{V}(m_s) = \max_{a \in A} \sum_{C \subseteq X} (m_s * a)(C)\, U_{\phi u}(e_C),$$

by a strictly increasing transformation, u is a von Neumann–Morgenstern utility function on X, and ψ and ϕ are continuous and strictly increasing functions on their respective domains. The representation W_1 and the function u are unique up to positive linear transformations. Given W_1 and u, the function $\hat{V}_1$ is uniquely determined, while ψ and ϕ are unique up to positive linear transformations.

Proof. The first part of the proof follows the same reasoning as the proof of Theorem 2. By analogous arguments, we obtain that the function

$$H(m) = \max_{a \in A} \sum_{C \subseteq X} (m * a)(C)\, U_{\phi u}(e_C)$$

represents the preference relation $\succsim^*$ on $\mathcal{M}(\Omega)$. Equivalently, it represents preferences over the set of degenerate information structures.

Because the preference relation $\succsim^*$ on $\mathcal{M}(\mathbb{S})$ is a quasi-average preference, Theorem 9 in Eichberger and Pasichnichenko (2021) implies that there exist a von Neumann–Morgenstern utility function v on $\mathbb{S}$ and a continuous and strictly increasing function ψ on $v(\mathbb{S})$ such that $\succsim^*$ is represented by

$$\sum_{E \subseteq S} M(E)\, U_{\psi v}(e_E) \tag{9}$$

for all $M \in \mathcal{M}(\mathbb{S})$, where S denotes the subset on which M is concentrated. For each $m \in \mathcal{M}(\Omega)$, define $\hat{V}_1(m) = v(s)$ for the unique $s \in \mathbb{S}$ such that $m_s = m$.

By construction, $\hat{V}_1(m) \geq \hat{V}_1(m')$ if and only if $m \succcurlyeq^* m'$. Thus, both $\hat{V}_1$ and H represent the preference relation $\succcurlyeq^*$ on $\mathcal{M}(\Omega)$, and therefore they are related by a strictly increasing transformation. Finally, substituting $v(s)$ with $\hat{V}_1(m_s)$ in Equation (9) yields the desired representation $W_1(I)$.

The proof that the representation implies the assumptions is straightforward. The uniqueness properties follow from Theorem 9 in Eichberger and Pasichnichenko (2021). □

The representation W_1 is more general than W, as $\hat{V}_1(m_s)$ is merely a strictly increasing transformation of the maximal value under the posterior m_s. Consequently, the equality $\hat{V}_1(m_s) = \hat{V}(m_s)$ does not hold in general.

Alternatively, Assumption 4 can be strengthened by extending it to all mass distributions in $\mathcal{M}(\mathbb{S})$, rather than restricting it to those concentrated on singletons:

Assumption 4'. For all $M, M' \in \mathcal{M}(\mathbb{S})$, $M \succcurlyeq^* M'$ if and only if $T_M \succcurlyeq T_{M'}$.

This allows us to specialize Theorem 2 to the case $\psi = \phi$. In other words, the resulting representation W_2 does not distinguish between the DM's attitude toward uncertainty in signals and uncertainty in outcomes. Assumptions 2 and 3 are no longer required, since Assumption 4' links $\succcurlyeq$ and $\succcurlyeq^*$ and makes $\succcurlyeq^*$ inherit a quasi-average preference structure from $\succcurlyeq$.

Theorem 5. *The DM's preferences satisfy Assumptions 1 and 4' if and only if there exists a representation W_2 of the preference relation $\succcurlyeq^*$ on the set of information structures such that*

$$W_2(I) = \sum_{E \subseteq S} M(E)\, \phi^{-1}\left(\frac{1}{|E|} \sum_{s \in E} \phi\left(\hat{V}\left(m_s\right)\right)\right),$$

where

$$\hat{V}\left(m_s\right) = \max_{a \in A} \sum_{C \subseteq X} \left(m_s * a\right)(C)\, U_{\phi u}(e_C),$$

u is a von Neumann–Morgenstern utility function on X, and ϕ is a continuous and strictly increasing function on its domain. The representation W_2 is unique up to a positive linear transformation. Given W_2, the functions u and $\hat{V}$ are unique, while ϕ is unique up to a positive linear transformation.

Proof. By Assumption 4', $M \succcurlyeq^* M'$ if and only if $T_M \succcurlyeq T_{M'}$. From Assumption 1 and Theorem 9 in Eichberger and Pasichnichenko (2021), it follows that $T_M \succcurlyeq T_{M'}$

if and only if

$$\sum_{C\subseteq X} T_M(C)\, U_{\phi u}(e_C) \geq \sum_{C\subseteq X} T_{M'}(C)\, U_{\phi u}(e_C).$$

Since $T_M(C) = M(E)$ for all $M \in \mathcal{M}(\mathbb{S})$, $C \in \mathcal{E}_X$, and $E = \tau^{-1}(C)$, we obtain $M \succsim^* M'$ if and only if

$$\sum_{E\subseteq S} M(E)\ \phi^{-1}\left(\frac{1}{|E|}\sum_{s\in E}\phi\left(u(\tau(s))\right)\right) \geq \sum_{E'\subseteq S'} M'(E')\ \phi^{-1}\left(\frac{1}{|E'|}\sum_{s'\in E'}\phi\left(u(\tau(s'))\right)\right).$$

Define $\hat{V}$ by $\hat{V}(m_s) = u(\tau(s))$ for all $s \in \mathbb{S}$. Note that

$$u(\tau(s)) = u(c(m_s * a_s)) = \sum_{C\subseteq X} (m_s * a_s)(C)\, U_{\phi u}(e_C) = \max_{a\in A} \sum_{C\subseteq X} (m_s * a)(C)\, U_{\phi u}(e_C).$$

Hence, we obtain the desired representation.

The proof that the representation implies the assumptions is straightforward. The uniqueness result follows from Theorem 9 in Eichberger and Pasichnichenko (2021). □