Blackwell Monotonicity and Motivated Reasoning[*]

Inga Deimen[†] Ernesto Rivera Mora[‡]

March 30, 2026

Abstract

When does more information increase an agent's expected utility across all decision problems?—when does *Blackwell monotonicity* hold? In the standard setup all agents under Bayesian updating weakly benefit from more information. This paper goes beyond the standard setup by studying settings with *motivated reasoning* and *belief-dependent preferences*. That is, (1) we allow agents to choose their subjective belief and (2) we allow for preferences that directly depend on the subjective belief. We show that for unconstrained reasoning Blackwell monotonicity holds. Under constrained reasoning, however, it often fails. We characterize the class of reasoning constraints that satisfy Blackwell monotonicity.

JEL: D81, D83, D91

Keywords: *Blackwell experiments, belief choice, psychological decision problems*

---

[*]We are grateful to Nageeb Ali, Davide Bordoli, Philipp Strack, Ignacio Esponda, Amanda Friedenberg, Jeanne Hagenbach, Navin Kartik, Botond Kőeszegi, Alessandro Pavan, and the audiences at Arizona, Edinburgh, Loyola Marymount, Paris Game Theory Festival 2025, SAET (Ischia 2025), Stony Brook GT 2025.

[†]University of Arizona, ideimen@arizona.edu

[‡]University of Colorado Boulder, ernesto.riveramora@colorado.edu

1 Introduction

Individuals often face situations in which learning the truth is psychologically painful. When information carries emotional costs, individuals may prefer to remain ignorant rather than confront uncomfortable realities. As an example, consider an individual concerned about their medical condition. A visit to the doctor could reveal bad news—not only emotionally painful but potentially harmful to physical health.[1] Yet patients often find ways to rationalize or distort the implications of medical information. For example, the patient might convince herself that the doctor is biased by financial incentives for surgery, or that the diagnosis does not apply to her family's unique history. If the patient cannot shield herself from the psychological pain of incoming information, she might prefer to avoid learning about her health condition altogether. These distortions, while emotionally protective, can lead to delayed or forgone treatment. How does the ability to distort her beliefs affect the patient's value for information? When does the benefit of a better-informed decision always outweigh the cost of painful news?

To address these questions, we introduce a novel and flexible framework of motivated reasoning in which (1) the agent's decision problem can feature belief-dependent preferences and (2) the agent has some flexibility in choosing her subjective belief. More precisely, the agent's decision problem is divided into two periods in which "Self 1" and "Self 2" of the agent operate. Although the two selves share identical preferences, they differ in their level of introspection and their ability to act.[2] The first period consists of two phases. In the *learning phase*, Self 1 observes a signal from a (Blackwell) experiment and updates her prior belief according to Bayes rule to an *objective* interim belief. In the *belief-selection phase* Self 1 chooses a *subjective* belief that her second-period self (Self 2) will hold when making the final decision. Thus Self 1 operates at a high level of introspection but cannot directly control the action. Instead, she influences the action indirectly by selecting the optimal subjective belief for Self 2. In the second period, Self 2 takes the subjective belief chosen by Self 1 as her belief, thus operating at a lower level of introspection, and selects the action.[3]

In accordance with the literature on motivated reasoning, we impose two key behavioral

[1] The psychological stress from a difficult diagnosis can have direct physiological consequences, a phenomenon related to the *nocebo effect* in which negative expectations induce adverse physical effects (Faasse, 2019; Todaro, Shen, Niaura, Spiro III, and Ward, 2003).

[2] Equivalently, one can think of Self 1 and Self 2 as two distinct agents with identical preferences. For instance, Self 1 might be a parent who shapes a child's beliefs, while Self 2 is a child who naively acts on the belief the parent induced.

[3] In psychology, dual-process theories distinguish System 1 processing (fast and automatic) from System 2 processing (slow and reflective); see Kahneman (2011); Evans and Stanovich (2013). Our binary-self framework aligns with this split: Self 1 engages in the reflective choice of beliefs, while Self 2 acts on the chosen belief with limited access to how it was formed.

assumptions (often implicitly assumed) on how the agent reasons and acts:

(i) Endogenous subjective beliefs: The agent selects a subjective belief that may not align perfectly with the objective reality but serves some internal purpose or need.

(ii) Internal consistency: Coherent behavior prohibits actions that are inconsistent with the agent's subjective beliefs.

Taken together, these conditions state that, while the agent can choose a subjective belief, her action must remain consistent with her subjective point of view. The binary-self model that we introduce provides a simple micro-foundation for internal consistency and explicitly describes the process of how the individual forms her subjective belief and makes choices.[4]

To illustrate, consider the optimal expectations model of Brunnermeier and Parker (2005), in which an agent optimizes her average felicity by trading off gains from optimistic beliefs against losses from poor choices. In their application, the agent holds subjective optimistic beliefs about the economy while simultaneously maintaining a low level of savings—an action consistent with her subjective optimistic belief. Moreover, as they write: *"While the interaction between optimistic and rational forces can be viewed as a model of a divided self, agents are unaware of this division and of the fact that their beliefs may be biased." (p.1097)*

Within our framework of motivated reasoning, we ask under which conditions the individual benefits from receiving more information. To understand the value of information, we take the classical approach provided by Blackwell's theorem (Blackwell, 1951, 1953). His seminal work establishes equivalences across different rankings of experiments in terms of their level of informativeness. One key implication of Blackwell's results is the principle that being better informed improves welfare. This principle, recently called *Blackwell monotonicity*, states that experiments that are more informative—in terms of the garbling ranking—always yield higher expected utility, regardless of the particular decision problem being considered.[5] To the best of our knowledge, the impact of information on welfare in settings with *motivated reasoning*—that is, for non-Bayesian updating and belief-dependent preferences—remains largely unexplored.[6]

[4] Our results hold beyond our binary-self micro-foundation. In particular, they hold in any framework in which endogenous subjective beliefs satisfy conditions (i) and (ii). Since in our model, the agent's action must be consistent with the subjective belief, the agent faces an endogenous cost of belief distortion. To have a well-defined notion of welfare across all decision problems, we do not impose an additional exogenous cost for belief manipulation as in Caplin and Leahy (2019).

[5] Whitmeyer (2024) introduces this concept for standard decision problems with non-Bayesian updating functions. Our work adapts this terminology to settings with updating correspondences and belief-dependent preferences.

[6] Spiegler (2008) is an important exception. He explores the effect of information acquisition in a two-period model of anticipatory utility (Brunnermeier and Parker, 2005). Our framework allows for a general class of belief-choice restrictions, experiments, and belief-dependent decision problems (including those beyond anticipatory utility).

We first show that when the agent faces no constraints in the belief-selection phase, she benefits from more information in the learning phase for all psychological (and standard) decision problems. Hence, Blackwell monotonicity holds: more informative experiments never reduce the agent's welfare. Intuitively, the agent can always choose a subjective belief that mitigates both the psychological costs of new information and the costs of taking an uninformative action.

Second, we show that this is no longer true in general if the agent's motivated reasoning is constrained. We explicitly model the agent's flexibility and limitations in changing her subjective belief by considering different *reasoning constraints*.[7] Formally, we define each reasoning constraint by an updating correspondence that describes for each objective belief the set of subjective beliefs that Self 1 can choose from in the belief-selection phase. This allows our framework to capture different models of motivated reasoning used in the literature. For example, under "Bayesian" reasoning, the agent can only choose the subjective belief equal to the objective belief that follows from Bayesian updating. In Brunnermeier and Parker (2005), the agent cannot assign positive probability to states proven impossible but retains flexibility within the remaining set of beliefs. We call this restriction "deductive" reasoning, which acknowledges clear evidence while leaving some freedom in interpreting imperfect signals. Our framework also captures agents with systematic updating errors ("exogenously distorted" reasoning), as studied in Jakobsen (2025), de Oliveira (2018), and Whitmeyer (2024). Other examples of reasoning constraints are provided in Section 4.1.

Our main result characterizes the reasoning constraints under which Blackwell monotonicity holds for all decision problems, including those with belief-dependent preferences. The characterization takes a simple form. We show that Blackwell monotonicity holds if and only if the reasoning constraint is *shrinking in support*. That is, if and only if "the set of available subjective beliefs weakly increases as the support of the objective belief decreases." Informally, this condition requires that, as information becomes more precise (i.e., ruling out more states), the agent weakly gains flexibility in constructing an internal subjective narrative.[8]

Our characterization shows that the increased flexibility makes additional information valuable, as the agent can better align her subjective belief with her belief-dependent preferences while simultaneously selecting a more informed action. Formally, we show that

[7] Following the separation principle proposed by Battigalli and Generoso (2023), our model distinguishes between the formal rules of the decision problem (including intrinsic belief preferences) and the decision-maker's personal characteristics (i.e., whether the agent follows Bayesian or other principles). We capture these personal characteristics by introducing different reasoning constraints that describe the limits under which an agent can choose her subjective belief.

[8] We take no normative stand on the shrinking-in-support condition, but keep a positive perspective throughout the paper. (See Section 7.)

when the agent's reasoning is shrinking in support and an experiment E' is a garbling of E, then E weakly improves the agent's welfare across all decision problems. Conversely, when the agent's reasoning constraint fails to satisfy the shrinking in support condition, we construct belief-dependent decision problems in which a garbling E' of E leads to strictly higher welfare.

Our main result extends to weaker versions of Blackwell monotonicity—specifically, when Blackwell monotonicity needs to hold only for certain subclasses of experiments or certain subclasses of decision problems. First, Section 5.1 considers settings with *full-support* experiments, in which all signals occur with positive probability regardless of the state. For these settings, we show that the shrinking-in-support condition can be replaced with the weaker requirement that the agent's reasoning constraint is *constant in the interior*, i.e., within the set of full-support beliefs the choice set does not vary with the prior (see Proposition 2). For example, consider deductive reasoning (which satisfies being constant in the interior). The result implies that more informative experiments benefit deductive agents when restricting attention to experiments that have full support. Thus within the context of Brunnermeier and Parker (2005), an agent that solely observes imperfect signals about the economy will always prefer more informative sources. Nevertheless, a signal that perfectly reveals that the economy is deteriorating, would decrease the agent's welfare—as deductive reasoning fails Blackwell monotonicity.

Second, Section 5.2 studies settings with *non-binding* decision problems, in which the agent's reasoning constraint does not limit the agent's welfare compared to unconstrained reasoning. We prove that for any such non-binding decision problem, more information always improves the agent's welfare (see Proposition 3). For instance, when we focus on decision problems in which preferences are *increasing* in beliefs about the high state, an agent with increasing reasoning always weakly benefits from more informative sources. For example an optimist who cares about good health is not worried about going to the doctor.

To illustrate our results, we examine Blackwell monotonicity in decision problems with binary states and binary actions (see Section 6). While our main interpretation captures medical decisions, our application covers other environments with motivated reasoning. For instance, a student uncertain about her ability might choose to believe she has high ability, even if this belief lowers her motivation to study for an exam. Likewise, a voter may form beliefs about a candidate's competence that match her ideological preferences rather than using objective evidence. We analyze how additional information influences behavior and payoffs in binary decision problems assuming the agent has some flexibility in selecting her subjective belief.

Literature review

Our paper contributes to several strands of literature. First, we build on the literature on belief-dependent preferences, pioneered by Geanakoplos, Pearce, and Stacchetti (1989) and further developed by Battigalli and Dufwenberg (2009). These papers establish the theoretical foundations for analyzing environments in which agents' utilities depend directly on their beliefs. Lipnowski and Mathevet (2018) point out that information may hurt Bayesian agents with psychological preferences. Dana, Weber, and Kuang (2007) provide empirical evidence that individuals actively avoid information in certain contexts. In our paper, we characterize conditions such that more information is welfare enhancing across all psychological (and standard) decision problems.

Second, we contribute to the literature on motivated reasoning.[9] The psychological foundations of motivated reasoning were established by Kunda (1990). In economics, Akerlof and Dickens (1982) pioneer the study of motivated beliefs by analyzing workers' belief manipulation about workplace accidents. Several papers study anticipatory utility, in which agents' current preferences depend on their beliefs about future outcomes and the emotions these beliefs generate. This often leads to strategic manipulation of information and biased updating of new information (Caplin and Leahy, 2001; Brunnermeier and Parker, 2005; Kőszegi, 2006; Spiegler, 2008; Ortoleva, 2024).

More broadly, Bénabou (2015) and Bénabou and Tirole (2016) argue that beliefs can fulfill psychological and functional needs beyond anticipatory utility. Caplin and Leahy (2019) show that "wishful thinking" can result in procrastination, confirmation bias, polarization, and the endowment effect. In an experiment, Zimmermann (2020) studies the impact of feedback on motivated beliefs and how this impact depends on the timeline. Niehaus (2020) studies the biases of an altruist who wants to believe in his good deeds and avoids information that reveals the impact of his actions. Little, Schnakenberg, and Turner (2022) study the effect of motivated reasoning in the context of democratic accountability.

Our framework provides a way to model self-deception. Like in Von Hippel and Trivers (2011), our model captures *"... the non-unitary of the mind and how different types of psychological dualism enable the same person to be both deceiver and deceived."* However, unlike in Sobel (2020, 2024)—where deceiving can generate "damage" through suboptimal actions—in our framework, subjective beliefs directly affect utility and can improve welfare even when they differ from reality. Hagenbach and Koessler (2022) analyze equilibria of a disclosure game in which an informed Self 1 can choose to "memorize" and disclose information to Self 2, or to "forget" so that Self 2 is uninformed when taking the action.

[9] The literature also uses the term *motivated beliefs*.

Our analysis contributes to the literature on overconfidence that results from non-Bayesian updating rules. Compte and Postlewaite (2004) show that selective memory—in which agents recall successes more easily than failures—can generate optimistically biased beliefs that improve welfare. Similarly, Bénabou and Tirole (2002) and Köszegi (2006) show that individuals strategically suppress negative information to preserve their self-image. Moreover, overconfidence may trigger misguided learning about other fundamentals (Heidhues, Kőszegi, and Strack, 2018). More generally, Bohren and Hauser (2023) link non-Bayesian updating and model misspecification.

We contribute to the literature examining whether information improves the welfare of non-Bayesian agents. For standard decision problems, Morris and Shin (1997), Jakobsen (2025), and Whitmeyer (2024) investigate whether classical welfare results on information extend to agents with non-Bayesian updating rules. Our approach differs from these papers in two ways. First, our model incorporates belief-dependent preferences. Second, we give agents some flexibility to choose their subjective beliefs.

The structure of the paper is as follows. The next section introduces the model. Section 3 studies the value of information for unconstrained reasoning. In Section 4, we introduce reasoning constraints, provide examples, and state preliminary results. In Section 5, we state our main result that provides necessary and sufficient conditions such that Blackwell monotonicity holds for all decision problems with belief-dependent preferences and motivated reasoning. We apply our equivalence result to binary decision problems for different reasoning constraints, in Section 6. Section 7 discusses extensions. All omitted proofs can be found in the appendix.

2 Model

We study decision making by a single agent. There is a finite set of payoff relevant states, Θ, and an arbitrary set of actions A. Write $\Delta(\Theta)$ for the set of probability distributions over Θ. The agent's preferences are described by a utility function $u : A \times \Theta \times \Delta(\Theta) \to \mathbb{R}$, in which $\Delta(\Theta)$ captures the agent's interim subjective beliefs, i.e., the beliefs the agent has when taking the action. Write (A, u) for the agent's **decision problem**. Note that this includes **standard** decision problems—in which u is constant in the agent's beliefs—as well as **psychological** decision problems—in which the agent not only cares about the action and the state, but also has preferences about what beliefs she has at the time of making the decision.[10]

The agent's belief develops over multiple periods. In period 0, the agent has an exogenous and objective prior $\mu_0 \in \Delta(\Theta)$ about the state of the world θ. The agent then observes information in the form of a signal realization. The signal realization $s \in S$ comes from **experiment** (S, π) where S is a finite set of **signals** and $\pi : \Theta \to \Delta(S)$ is the **experiment's protocol** that prescribes a signal distribution for each state. In period 1, the signal induces an **objective belief** μ_1 for the agent, that is, the agent updates her belief from μ_0 to μ_1 using Bayes rule. The agent then has the opportunity to distort her belief, i.e., the agent can choose a **subjective belief** $\mu_2 \in \Delta(\Theta)$. In period 2, the agent has subjective belief μ_2 and takes an action $a \in A$. In period 3, payoffs are realized.

For convenience, we call the agent in period 1 "Self 1" and in period 2 "Self 2." It is assumed that Self 2 takes belief μ_2 at face value. Self 1, however, understands that the objective belief is μ_1. While the model defines Self 1 and Self 2 as different individuals, both selves get a realized payoff $u(a, \theta, \mu_2)$ and hence are interpreted as different selves of the same agent.

To summarize, the timing is as follows:

Period 0: Nature selects a state $\theta \in \Theta$ according to the prior $\mu_0 \in \Delta(\Theta)$ and selects a signal $s \in S$ according to $\pi : \Theta \to \Delta(S)$.

Period 1: Self 1 observes the signal s and updates the belief to μ_1 using Bayes rule. Self 1 chooses a subjective belief $\mu_2 \in \Delta(\Theta)$ for Self 2.

Period 2: Self 2 takes μ_2 as her belief and selects an action $a \in A$.

Period 3: The agent gets payoff $u(a, \theta, \mu_2)$.

We will start analyzing the problem by focusing on the agent's expected payoff in period 2. Assume that each Self i has belief μ_i. For all actions $a \in A$ and states $\theta \in \Theta$ write

$$\mathbb{E}_{\mu_i}\big[u(a, \theta, \mu_2)\big] := \sum_{\theta \in \Theta} u(a, \theta, \mu_2) \cdot \mu_i(\theta) \quad \text{ for } \mu_i = \mu_1, \mu_2.$$

So, given that Self 2 has belief μ_2, Self 2 chooses

$$a \in \arg\max_{a \in A} \mathbb{E}_{\mu_2}\big[u(a, \theta, \mu_2)\big].$$

[10]We could have similarly assumed that the agent's payoff depends also on the ex-post beliefs. Section 7.2 shows that this restriction is without loss of generality: our results extend to environments in which the agent has preferences about both interim and posterior beliefs.

Note that an optimal action for Self 2 always exists if u has a finite range or if u is continuous and A is compact. (See Lemma A.1)

Notice, Self 1's expected utility from Self 2's action is $\mathbb{E}_{\mu_1}[u(a, \theta, \mu_2)]$, which has an expectation based on the objective belief μ_1 instead of the subjective belief μ_2. In this sense, Self 1 acknowledges the true prior μ_1 but evaluates the utility function u at the subjective belief μ_2. So, while Self 1 and Self 2 differ in their anticipatory utility, they receive the same realized utility $u(a, \theta, \mu_2)$. Moreover, since Self 1 anticipates the behavior of Self 2's, Self 1 chooses a belief μ_2 that maximizes expected utility, considering both the value that u assigns to different interim beliefs and the costs of misleading Self 2's action.

Remark 1. The welfare analysis is based on Self 1's perspective, who chooses the best belief among all beliefs that are optimal for Self 2. Accordingly, we say that "the agent is weakly better off" whenever Self 1's expected payoff is weakly higher.

3 The Value of Information

This section analyzes the value of information in our setting with "unconstrained" motivated reasoning, that is, when Self 1 can freely choose Self 2's belief. Following Blackwell (1951, 1953) we use the garbling ranking as our measure of informativeness.

Definition 1. Fix experiments (S, π) and (S', π'). Say (S', π') is a **garbling** of (S, π) if there is a mapping $\gamma : S \to \Delta(S')$ such that

$$\pi'(s' \mid \theta) = \sum_{s \in S} \gamma(s' \mid s) \cdot \pi(s \mid \theta).$$

If (S', π') is a garbling of (S, π), say (S, π) is **more informative** than (S', π').

Informally, (S', π') is a garbling of (S, π) if (S', π') adds noise to the signal revealed by (S, π).[11] This definition induces a partial order over the space of experiments. Blackwell connects this ranking of information to a welfare ranking. In particular, his results imply that a Bayesian agent prefers more information over less, regardless of the standard decision problem that the agent faces.[12] A key implication is that—in standard settings—observing any experiment weakly improves the agent's welfare compared to observing no experiment at all.

[11] In the literature, (S, π) is said to be "sufficient" for (S', π').

[12] In the original formulation, Blackwell's theorem links informativeness with loss minimization. For a formulation of Blackwell's theorem in terms of expected utilities, see Kihlstrom (1984), Crémer (1982), or de Oliveira (2018).

By contrast, it is well known that information can be detrimental to Bayesian agents who face psychological decision problems (Bénabou and Tirole (2016), Lipnowski and Mathevet (2018)). Thus for Bayesian agents, the implications of Blackwell's theorem do not extend to psychological decision problems. We now broaden the picture by studying psychological decision problems in our context of motivated reasoning in which the agent can choose her subjective belief.

Proposition 1. *Assume (S, π) is more informative than (S', π'). For all decision problems (A, u) and all priors $\mu_0 \in \Delta(\Theta)$, an agent without reasoning constraints is weakly better off under (S, π) than under (S', π').*

We delay the formal proof to Section 5 and provide some intuition here. Proposition 1 follows from the fact that—absent any restrictions in the choice of the subjective belief—the psychological decision problem can be transformed into a standard decision problem. To see this, write

$$\mathcal{F} := \Big\{(\mu_2, a) \in \Delta(\Theta) \times A \ : \ a \in \arg\max_{a \in A} \mathbb{E}_{\mu_2}[u(a, \theta, \mu_2)]\Big\},$$

for the feasible set of belief-action pairs that satisfy internal consistency, i.e., the agent's choice of action must be optimal given the subjective belief μ_2. Since Self 1 is unconstrained in choosing the subjective belief μ_2, the feasible set $\mathcal{F}$ is independent of the objective belief μ_1. In the auxiliary problem the "action set" is given by $\hat{A} := \mathcal{F}$. Each action $\hat{a} = (\mu_2, a) \in \hat{A}$ yields utility $\hat{u}(\hat{a}, \theta) := u(a, \theta, \mu_2)$ for each state $\theta \in \Theta$. So, the auxiliary problem $(\hat{A}, \hat{u})$ is standard: Self 1 faces a fixed action set $\mathcal{F}$ and standard expected utility maximization of $\mathbb{E}_{\mu_1}[\hat{u}(\hat{a}, \theta)]$ under the objective belief μ_1. Blackwell's theorem therefore applies directly, ensuring that the value function

$$V(\mu_1) := \sup \big\{\mathbb{E}_{\mu_1} u(a, \theta, \mu_2) : (\mu_2, a) \in \mathcal{F}\big\}$$

is convex in the objective belief μ_1. Hence, more informative experiments weakly improve the agent's welfare.

4 Constrained Reasoning

The previous analysis focused on unconstrained reasoning, in which Self 1 can choose any subjective belief μ_2 from the entire set of possible beliefs $\Delta(\Theta)$. However, some agents might face constraints on their ability to select beliefs. This section extends the basic model to incorporate these constraints through an **updating correspondence** $B : \Delta(\Theta) \rightrightarrows \Delta(\Theta)$ that maps each objective belief μ_1 to a subset $B(\mu_1) \subseteq \Delta(\Theta)$ of feasible subjective beliefs.

The updating correspondence B identifies the agent's **reasoning constraints** by specifying which beliefs Self 1 can choose for Self 2 given the objective belief μ_1.[13] When $B(\mu_1) = \Delta(\Theta)$ for all μ_1, the agent has unconstrained reasoning and can choose any belief. When $B(\mu_1) \subsetneq \Delta(\Theta)$ for some μ_1, the agent faces reasoning constraints and must select a subjective belief from the restricted set $B(\mu_1)$. The timing is now as follows:

Period 0: Nature selects a state $\theta \in \Theta$ according to the prior $\mu_0 \in \Delta(\Theta)$ and selects a signal $s \in S$ according to $\pi : \Theta \to \Delta(S)$.

Period 1: Self 1 observes the signal s and updates the belief to μ_1 using Bayes rule. Self 1 then chooses a subjective belief $\mu_2 \in B(\mu_1)$ for Self 2, where the constraint $B(\mu_1)$ limits the available choices.

Period 2: Self 2 takes μ_2 as her belief and selects an action $a \in A$.

Period 3: The agent gets payoff $u(a, \theta, \mu_2)$.

4.1 Examples of reasoning constraints

Agents may face very different constraints in their belief choice, or, in other words, agents can vary substantially in their level of flexibility in selecting their beliefs. In the following, we provide a list of reasoning constraints that feature natural properties, some of which have been considered in the literature. Figure 4.1 illustrates (in blue) examples of reasoning constraints for a binary state space.

Example 1. Bayesian. $B(\mu) = \{\mu\}$ for all $\mu \in \Delta(\Theta)$.

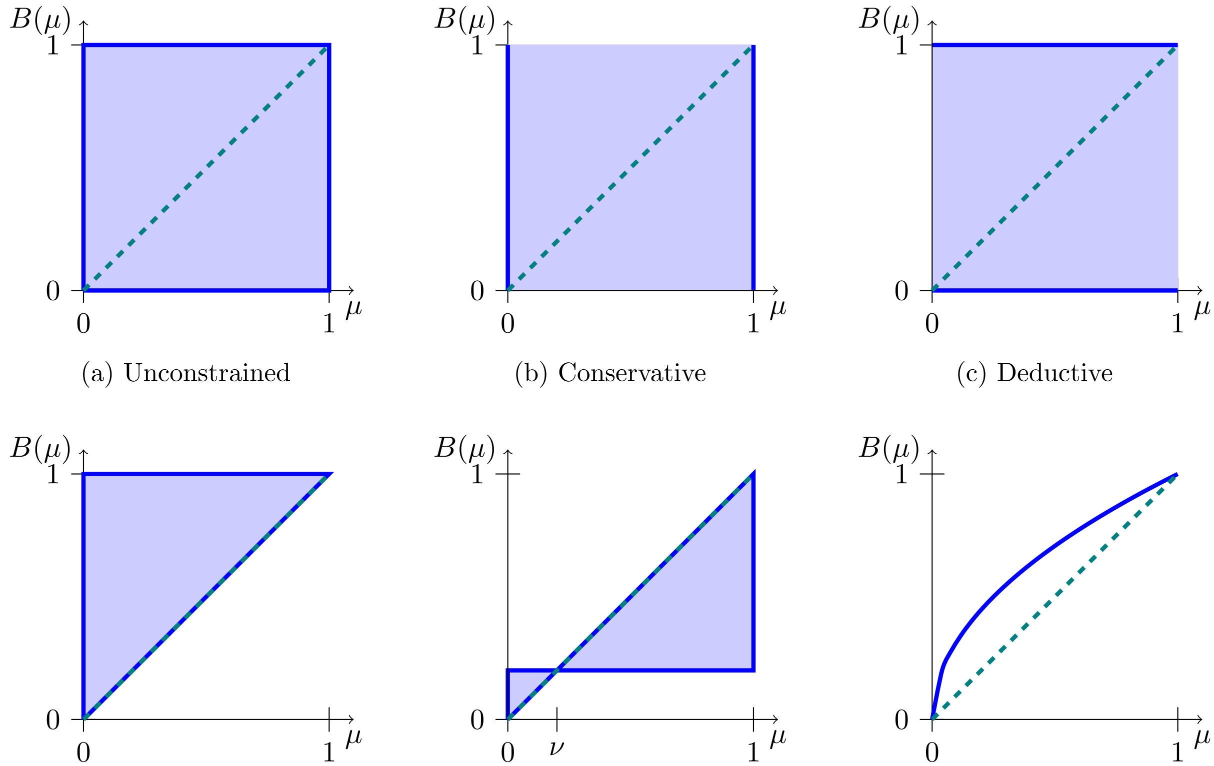
If the reasoning constraint is "Bayesian," the agent updates her beliefs using Bayes rule. (See the teal-dashed line in Figure 4.1.) Since the agent cannot choose to have a different belief, her beliefs are never inaccurate. In psychological decision problems, Bayesian beliefs have been used, for example, in Lipnowski and Mathevet (2018) and Köszegi (2006).

Example 2. Conservative. $B(\mu) = \left\{\mu' \in \Delta(\Theta) : \text{Supp}\,(\mu) \subseteq \text{Supp}\,(\mu')\right\}$ for all $\mu \in \Delta(\Theta)$.

If the reasoning constraint is "conservative," the agent can change beliefs but cannot discard states that are still possible. For instance, consider an individual with *illness anxiety disorder* who exhibits excessive concern about her health.[14] Doctor visits or negative test results give her only little reassurance as these cannot rule out illness with certainty.

[13] The concept of an updating correspondence is very flexible and allows us to capture a large range of possible applications. In this sense our paper offers a unifying approach linking different branches of the literature of motivating reasoning.

[14] See Scarella, Boland, and Barsky (2019).



(a) Unconstrained (b) Conservative (c) Deductive

(d) Increasing (e) Reference-biased ($\nu = 0.2$) (f) Exogenously distorted

Figure 4.1. Examples of reasoning constraints in a setting with binary state space $\Theta = \{\underline{\theta}, \overline{\theta}\}$. Each belief $\mu \in \Delta(\Theta)$ is described by the probability $\mu(\overline{\theta}) \in [0, 1]$. The updating correspondence is depicted in blue. Bold blue lines indicate where the boundary is included. The teal-dashed line depicts Bayesian reasoning.

Example 3. Deductive. $B(\mu) = \left\{\mu' \in \Delta(\Theta) : \text{Supp}\,(\mu') \subseteq \text{Supp}\,(\mu)\right\}$ for all $\mu \in \Delta(\Theta)$.

If the reasoning constraint is "deductive," the agent cannot assign positive probability to the states that cannot happen. This reasoning constraint is used, for example, in Brunnermeier and Parker (2005), Spiegler (2008), and Niehaus (2020).[15]

Example 4. Increasing. Assume $\Theta \subset \mathbb{R}$, finite.
$B(\mu) = \left\{\mu' : \mu' \text{ first-order stochastically dominates } \mu\right\}$ for all $\mu \in \Delta(\Theta)$.

Here, we assume Θ is a finite subset of $\mathbb{R}$. If the reasoning constraint is "increasing," the agent can only shift the prior belief towards higher states. For instance, consider an individual belonging to a family in which everyone strongly believes to be very healthy. The individual can easily adopt beliefs that align with her family's view, even without supporting evidence. However, the individual requires compelling evidence to justify holding beliefs about having

[15] In Niehaus (2020) the agent has a deductive reasoning constraint and chooses both interim and ex-post beliefs. (See Section 7.2.)

some disease.

Example 5. Reference-biased. Fix a reference belief $\nu \in \Delta(\Theta)$.
$B(\mu) = \big\{\lambda\mu + (1-\lambda)\nu : \lambda \in [0,1]\big\}$ for all $\mu \in \Delta(\Theta)$.

Consider an agent who holds a reference belief ν based on some reference point or prior belief. When receiving new information, she may partially incorporate this information but always maintains some weight on her reference belief. Even when evidence strongly suggests a different conclusion, she might choose to believe something closer to her reference point rather than fully accepting the new information.

Example 6. Exogenously distorted. $B(\mu) = \{f(\mu)\}$ for all $\mu \in \Delta(\Theta)$ and some mapping $f : \Delta(\Theta) \to \Delta(\Theta)$.

The agent's belief can be "exogenously distorted" via a mapping $f : \Delta(\Theta) \to \Delta(\Theta)$. The mapping f aims to capture "mistakes" in the way that the agent updates her beliefs. A systematic distortion of beliefs is, for example, used in Jakobsen (2025), de Clippel and Zhang (2022), and Whitmeyer (2024). In the particular case in which the mapping f is constant, the agent is stubborn and will not respond to information at all.

4.2 Preliminaries

Before analyzing the agent's value of information, we first ensure that the decision problem is well defined. To do so, we will assume that at each belief at which Self 2 is indifferent between multiple actions in period 2, Self 2 takes the preferred action of Self 1. Put differently, we assume that Self 1 can select the action for period 2, provided that action is among the optimal actions for Self 2.

Fix a decision problem (A, u) and a reasoning constraint B. For all $\mu_1 \in \Delta(\Theta)$, write

$$\mathcal{F}_B(\mu_1) := \big\{(\mu_2, a) \in \Delta(\Theta) \times A : \mu_2 \in B(\mu_1) \text{ and } a \in \arg\max_{a \in A} \mathbb{E}_{\mu_2}[u(a, \theta, \mu_2)]\big\}$$

for the pairs of beliefs and actions that can be induced by Self 1. Observe that—in contrast to the unconstrained case—this set depends on μ_1. Write

$$V_B(\mu_1) := \sup\big\{\mathbb{E}_{\mu_1} u(a, \theta, \mu_2) : (\mu_2, a) \in \mathcal{F}_B(\mu_1)\big\}$$

for the supremum of the values that are achievable by Self 1 (and hence, also for Self 2) given an objective belief μ_1. Say (A, u) is **well defined for *B*** if for each objective belief $\mu_1 \in \Delta(\Theta)$ there exists $(\mu_2, a) \in \mathcal{F}_B(\mu_1)$ such that $\mathbb{E}_{\mu_1} u(a, \theta, \mu_2) = V_B(\mu_1)$.

Lemma A.1 in the appendix provides two sufficient conditions for a well-defined decision problem. The first, ensures that every continuous utility achieves its maximum on a compact set of alternatives. The second guarantees that when u has a finite range, the supremum in $V_B(\mu_1)$ is attained by some value in a finite set.[16]

We now introduce a key property that will be central to our characterization result.

Definition 2. The reasoning constraint $B : \Delta(\Theta) \rightrightarrows \Delta(\Theta)$ is **shrinking in support** if for all $\mu, \mu' \in \Delta(\Theta)$, $\text{Supp}\,(\mu') \subseteq \text{Supp}\,(\mu)$ implies $B(\mu) \subseteq B(\mu')$.

A reasoning constraint B is shrinking in support if ruling out states (in the objective sense) can only enlarge the set of beliefs that Self 1 can choose from. In other words, if the agent has more information, she is better able to construct a coherent narrative that justifies subjective beliefs.

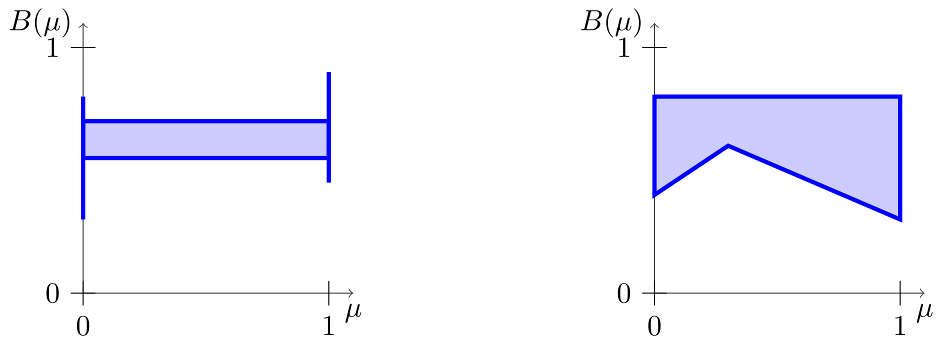


Figure 4.2. Examples of reasoning constraints for binary states: left panel, $B(\mu)$ is shrinking in support, right panel $B(\mu)$ is not shrinking in support.

Observe, if B is shrinking in support, then $B(\mu) = B(\mu')$ for each pair μ, μ' that satisfies $\text{Supp}\,(\mu) = \text{Supp}\,(\mu')$. In other words, the belief choice set determined by B is "constant in the interior." On the boundary beliefs, by contrast, the belief set can be arbitrarily enlarged. Figure 4.2 (left panel) illustrates an example of a reasoning constraint that is shrinking in support in a setting with binary states.

Going back to our examples, observe that the Bayesian reasoning constraint is *not* shrinking in support. By contrast, the unconstrained and the conservative reasoning constraints are shrinking in support. If the reasoning constraint is deductive, reference-biased, or increasing, then it is *not* shrinking in support. Finally, exogenously distorted reasoning constraints are shrinking in support if and only if the distortion mapping f is constant.

[16] Some examples in the literature use utility functions that are discontinuous in the belief (Geanakoplos, Pearce, and Stacchetti, 1989; Rivera Mora, 2024). From our examples, only the conservative reasoning constraint (Example 2) fails to be compact-valued. A finite range of u is a sufficient (but not necessary) condition for each problem to be well-defined for any reasoning constraint.

The next lemma provides an equivalent characterization of shrinking in support that emphasizes the nested structure of sets of available beliefs within linear combinations. This alternative formulation will be useful for the proof of our main result.

Lemma 1. *A reasoning constraint B is shrinking in support if and only if for all $\mu, \mu' \in \Delta(\Theta)$ and all $\lambda \in (0,1)$, $B(\lambda\mu + (1-\lambda)\mu') \subseteq B(\mu) \cap B(\mu')$.*

Lemma 1 provides a characterization of reasoning constraints that are shrinking in support. It states that B is shrinking in support if and only if each subjective belief that can be reached from objective beliefs in the interior of the simplex can also be reached from objective beliefs on the boundary of the simplex.

5 Blackwell Monotonicity

We are now equipped to study the value of information for constrained motivated reasoning.

Definition 3. The reasoning constraint B satisfies **Blackwell monotonicity** if whenever (S, π) is more informative than (S', π') the following holds: for all decision problems (A, u) and all priors $\mu_0 \in \Delta(\Theta)$, the agent is weakly better off under (S, π) than under (S', π').

Notice, in contrast to Whitmeyer (2024)—who introduces this terminology—we analyze a framework of endogenous belief choice and consider both standard and psychological decision problems. Our next result characterizes under which reasoning constraints Blackwell monotonicity holds:

Theorem 1. *The agent's reasoning constraint B satisfies Blackwell monotonicity if and only if B is shrinking in support.*

Following the implications of Blackwell (1951, 1953), our result connects garblings to a robust welfare criterion. To cover both standard and psychological decision problems, the reasoning constraint B has to be shrinking in support. Informally, this condition states that the agent is more flexible in choosing her subjective belief when she receives more information. Intuitively, this flexibility allows the agent to protect herself from any painful information. Conversely, when shrinking in support fails, more information can actually harm the agent, as more information may constrain her ability to choose welfare-improving beliefs. Sections 5.1 and 5.2 provide milder conditions for Blackwell monotonicity to hold in settings in which either the set of experiments or the set of decision problems are restricted.[17]

[17] Without reasoning constraints (Section 3) the agent's reasoning is shrinking in support. Hence Proposition 1 follows from Theorem 1.

Proof. Sufficiency: Assume that the reasoning constraint B is shrinking in support. Fix a decision problem (A, u) that is well defined for B. Write $\mathcal{F}_B(\mu_1) \subseteq B(\mu_1) \times A$ for the set of feasible subjective beliefs and optimal actions consistent with B and write $V_B(\mu_1)$ for the maximum value that is achievable by Self 1 given belief μ_1. The proof is divided into two steps. The first shows that V_B is weakly convex. The second uses the convexity to show that B satisfies Blackwell monotonicity.

Step 1. Fix $\lambda \in (0, 1)$ and priors $\mu_1, \mu_1' \in \Delta(\Theta)$. Let $\overline{\mu}_1 = \lambda\mu_1 + (1 - \lambda)\mu_1'$. Notice, since the reasoning constraint is shrinking in support, $B(\overline{\mu}_1) \subseteq B(\mu_1) \cap B(\mu_1')$. (See Lemma 1.) Therefore,

$$\begin{aligned}\mathcal{F}_B(\overline{\mu}_1) &= \big\{(\mu_2, a) \in \Delta(\Theta) \times A : \mu_2 \in B(\overline{\mu}_1) \text{ and } a \in \arg\max_{a\in A} \mathbb{E}_{\mu_2}[u(a, \theta, \mu_2)]\big\} \\ &\subseteq \big\{(\mu_2, a) \in \Delta(\Theta) \times A : \mu_2 \in B(\mu_1) \text{ and } a \in \arg\max_{a\in A} \mathbb{E}_{\mu_2}[u(a, \theta, \mu_2)]\big\} \\ &= \mathcal{F}_B(\mu_1).\end{aligned}$$

An analogous argument shows $\mathcal{F}_B(\overline{\mu}_1) \subseteq \mathcal{F}_B(\mu_1')$. Hence,

$$\begin{aligned}V_B(\overline{\mu}_1) &= \max\big\{\mathbb{E}_{\overline{\mu}_1}[u(a, \theta, \mu_2)] : (\mu_2, a) \in \mathcal{F}_B(\overline{\mu}_1)\big\} \\ &= \max\big\{\lambda\mathbb{E}_{\mu_1}[u(a, \theta, \mu_2)] + (1 - \lambda)\mathbb{E}_{\mu_1'}[u(a, \theta, \mu_2)] : (\mu_2, a) \in \mathcal{F}_B(\overline{\mu}_1)\big\} \\ &\leq \max\big\{\lambda\mathbb{E}_{\mu_1}[u(a, \theta, \mu_2)] : (\mu_2, a) \in \mathcal{F}_B(\overline{\mu}_1)\big\} + \max\big\{(1 - \lambda)\mathbb{E}_{\mu_1'}[u(a, \theta, \mu_2)] : (\mu_2, a) \in \mathcal{F}_B(\overline{\mu}_1)\big\} \\ &\leq \lambda \max\big\{\mathbb{E}_{\mu_1}[u(a, \theta, \mu_2)] : (\mu_2, a) \in \mathcal{F}_B(\mu_1)\big\} + (1 - \lambda) \max\big\{\mathbb{E}_{\mu_1'}[u(a, \theta, \mu_2)] : (\mu_2, a) \in \mathcal{F}_B(\mu_1')\big\} \\ &= \lambda V_B(\mu_1) + (1 - \lambda)V_B(\mu_1').\end{aligned}$$

So, V_B is weakly convex.

Step 2. Write $\tau, \tau' \in \Delta(\Delta(\Theta))$ for the posterior distributions associated to the experiments (S, π) and (S', π'), respectively. Assume that (S', π') is a garbling of (S, π). This implies that τ is a mean preserving spread of τ'. (See Blackwell (1951, 1953); Leshno and Spector (1992).) Observe, if V_B is convex, then, by Jensen's inequality,

$$\int_{\Delta(\Theta)} V_B(\mu_1)\; d\tau \geq \int_{\Delta(\Theta)} V_B(\mu_1)\; d\tau',$$

implying that the agent is weakly better off under (S, π) than under (S', π'). Since (A, u) is arbitrary, it follows that B is Blackwell monotone.

Necessity: This part of the proof is by contrapositive. Assume B is not shrinking in support. Then, there exist $\mu, \mu' \in \Delta(\Theta)$ and $\lambda \in (0, 1)$ such that for the convex combination $\overline{\mu} =$

$\lambda\mu + (1-\lambda)\mu'$ it holds that $B(\overline{\mu}) \nsubseteq B(\mu)$. (See Lemma 1.) Let $\widetilde{\mu} \in B(\overline{\mu}) \backslash B(\mu)$. We will use $\widetilde{\mu}, \overline{\mu}, \mu$, and μ' to construct a decision problem (A, u), a prior μ_0, and the experiments (S, π) and (S', π') such that Blackwell monotonicity is violated.

Let $A = \{a\}$ and $u : \Theta \times A \times \Delta(\Theta) \to \mathbb{R}$ be defined by $u(\theta, a, \mu_2) := \mathbb{1}[\mu_2 = \widetilde{\mu}]$. Set the prior $\mu_0 = \overline{\mu}$. We now introduce an experiment that spreads the prior belief $\overline{\mu}$ into posteriors μ and μ'. Let (S, π) be such that $S = \{s, s'\}$ and $\pi : \Theta \to \Delta(S)$ satisfies

$$\pi(s \mid \theta) \cdot \overline{\mu}(\theta) = \mu(\theta) \cdot \lambda \quad \text{and} \quad \pi(s' \mid \theta) \cdot \overline{\mu}(\theta) = \mu'(\theta) \cdot (1-\lambda) \qquad \forall \theta \in \Theta.$$

Notice, the ex-ante probability of signal s is $\sum_\theta \overline{\mu}(\theta) \cdot \pi(s \mid \theta) = \lambda$ and the ex-ante probability of signal s' is $\sum_\theta \overline{\mu}(\theta) \cdot \pi(s' \mid \theta) = 1 - \lambda$. Moreover,

$$\mu(\theta) = \frac{\overline{\mu}(\theta) \cdot \pi(s \mid \theta)}{\sum_{\theta'} \overline{\mu}(\theta') \cdot \pi(s \mid \theta')} \quad \text{and} \quad \mu'(\theta) = \frac{\overline{\mu}(\theta) \cdot \pi(s' \mid \theta)}{\sum_{\theta'} \overline{\mu}(\theta') \cdot \pi(s' \mid \theta')} \qquad \forall \theta \in \Theta.$$

So, after observing the signal s, Self 1 updates her belief from $\mu_0 = \overline{\mu}$ to $\mu_1 = \mu$. Likewise, after observing the signal s', Self 1 updates her belief from $\mu_0 = \overline{\mu}$ to $\mu_1 = \mu'$.

Note that observing (S, π) induces an expected utility strictly less than 1: Signal s induces belief $\mu_1 = \mu$. Since $\widetilde{\mu} \notin B(\mu)$, Self 1 can only achieve a utility of 0 after signal s. Moreover, signal s has ex-ante probability $\lambda > 0$.

Let (S', π') be an experiment such that S' is a singleton. So, trivially, (S', π') is a garbling of (S, π). Moreover, observing (S', π') induces belief μ_1 equal to the prior $\overline{\mu}$. Since $\widetilde{\mu} \in B(\overline{\mu})$, Self 1 can choose $\mu_2 = \widetilde{\mu} \in B(\overline{\mu})$ and achieve a utility of 1 with probability one. Thus, the agent has strictly higher expected utility under (S', π') than under (S, π). □

The first part of the proof shows that, provided that B is shrinking in support, the agent is unambiguously (weakly) better off with more informative signals. In particular, observing some experiment is always (weakly) better than observing no experiment. The proof shows that the value function $V_B(\mu_1)$ is convex in the objective belief μ_1. While it is standard to show convexity for the value of information, the emphasis on the role of the support is, to the best of our knowledge, new. The key insight rests on an "irreversibility property" of signals: a signal can rule out states by moving an interior objective belief to the boundary of the simplex, but the converse is impossible: once a state is ruled out, no signal can bring the objective belief back to the interior. As a result, the shrinking-in-support condition ensures that Self 1's choice set $\mathcal{F}_B(\mu_1)$ weakly expands as beliefs move towards the boundaries, making $V_B(\cdot)$ convex. Jensen's inequality then implies that the agent benefits from mean-preserving spreads of beliefs, precisely what more informative experiments generate.

The second part of the proof shows that Blackwell monotonicity fails for each B that is not shrinking in support. It proceeds by constructing a counterexample. If B is not shrinking in support, we can find a subjective belief $\widetilde{\mu}$ that can be chosen given the prior belief $\overline{\mu}$ but cannot be attained given an updated belief μ that results from a more informative experiment. The proof constructs a decision problem in which (1) the action and the state are not payoff relevant, and (2) the agent achieves her maximal payoff at the subjective belief $\widetilde{\mu}$.[18] Figure 5.1 provides an example in which the blue region describes the set of objective beliefs that attain a maximal payoff (i.e., the objective beliefs such that B allows for $\widetilde{\mu}$). Observe that when the experiment is uninformative (i.e., when the signal set is a singleton), the agent retains her prior belief, yielding her maximal payoff. By contrast, any informative experiment that makes the desirable subjective belief $\widetilde{\mu}$ infeasible after some signal strictly decreases the agent's payoff. Summarizing, for any reasoning constraint that is not shrinking in support, we can construct a decision problem in which more information harms the agent.

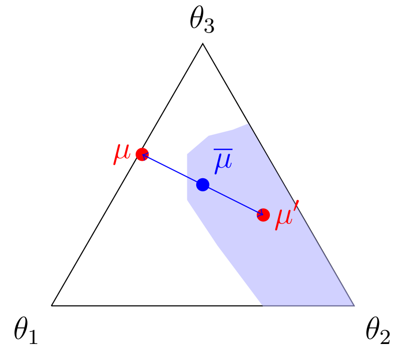


Figure 5.1. Illustration of the posterior spread used in the proof of necessity in a setting with three states. The blue shaded region represents the set of objective beliefs such that B allows for the desired subjective belief $\widetilde{\mu}$. Notice, this correspondence B is not shrinking in support.

5.1 Restricting the set of experiments

Theorem 1 provides a welfare result that holds across all decision problems and all experiments. However, parallel results can be provided when the analyst restricts either the set of experiments or the set of decision problems. We now present a result for a specific subclass of experiments that we call *full support* experiments. This class of experiments has been used, for example, in Caplin and Leahy (2019), Hagenbach and Koessler (2022), and Jakobsen (2025).

Definition 4. An experiment (S, π) has **full support** if, for each state $\theta \in \Theta$, the conditional distribution $\pi(\cdot \mid \theta) \in \Delta(S)$ has full support.

[18] Notice that Part (1) is not essential for the counterexample in the proof, one can add a small term that depends on the action and the state without impacting the final result.

An experiment has full support if all signals are possible for all states. So, observing the outcome of a full-support experiment does not preclude any state. Therefore, if the prior belief μ_0 has full support, the objective interim belief μ_1 also has full support.

Definition 5. The reasoning constraint B is **constant in the interior** if for each pair of beliefs $\mu, \mu' \in \Delta(\Theta)$ with full support, it holds that $B(\mu) = B(\mu')$.

The condition encodes a reversibility property among interior beliefs: for any two full-support beliefs $\mu, \mu' \in \Delta(\Theta)$, there exists a full-support experiment mapping μ to μ' via Bayesian updating, and vice versa. A reasoning constraint that assigns different sets of admissible choices to two interior beliefs would therefore introduce an asymmetry. Being constant in the interior rules out this asymmetry.[19]

Proposition 2. *Assume B is constant in the interior and assume (S, π) and (S', π') are full-support experiments. If (S, π) is more informative than (S', π'), then for all decision problems (A, u) and all full-support priors $\mu_0 \in \Delta\Theta$, the agent is weakly better off under (S, π) than under (S', π').*

Proof. Because the experiments have full support, they always induce beliefs that are in the interior of the belief space $\Delta(\Theta)$. Thus, it is sufficient to consider a reasoning constraint B that is constant in the interior (as opposed to shrinking in support) and apply an argument analogous to the proof of Theorem 1. □

As an example, consider the deductive reasoning constraint. While this correspondence is not shrinking in support, it is constant in the interior. Consequently, Theorem 1 does not apply, but Proposition 2 does. This means that, when the analysis is limited to experiments with full support, additional information is beneficial for agents with a deductive reasoning. Our results are thus helpful for understanding the value of information in Brunnermeier and Parker (2005) and Spiegler (2008) who consider deductive reasoning in their models of anticipatory utility.

5.2 Restricting the set of decision problems

Theorem 1 provides a characterization of reasoning constraints for which information improves welfare across *all* (including psychological) decision problems. This subsection provides a result for a class of decision problems in which the restriction of B is "non-binding."

[19] Being constant in the interior is a weaker condition than being shrinking in support. If B is shrinking in support, then $B(\mu) = B(\mu')$ for any pair of beliefs μ, μ' with $\text{Supp}(\mu) = \text{Supp}(\mu')$, which implies that B is constant in the interior.

That is, decision problems in which the agent—with reasoning constraint B—can achieve the same value as with unconstrained reasoning ($B(\mu_1) = \Delta(\Theta)$). To describe this formally, fix a psychological decision problem (A, u) and recall that $V(\mu_1)$ is the maximal value that an agent with unconstrained reasoning can achieve for an objective belief μ_1 given (A, u).

Definition 6. Fix a psychological problem (A, u) and reasoning constraint B. Say (A, u) **is non-binding for $\boldsymbol{B}$** if for all $\mu_1 \in \Delta(\Theta)$, $V_B(\mu_1) = V(\mu_1)$.

So, (A, u) is non-binding for reasoning constraint B if the restrictions on B do not constrain the utility that the agent can achieve. The following Lemma provides a sufficient condition for B to be non-binding in the case of separable decision problems. Separable problems are, for example, used when modeling ego utility (Köszegi, 2006), stubbornness (Lipnowski and Mathevet, 2018) or selecting memory (Hagenbach and Koessler, 2022).

Lemma 2. *Let (A, u) be a decision problem such that $u(a, \theta, \mu_2) = v(a, \theta) + \phi(\mu_2)$, for each $(a, \theta, \mu_2) \in A \times \Theta \times \Delta\Theta$. If a reasoning constraint B and the sub-utility $\phi : \Delta(\Theta) \to \mathbb{R}$ satisfy*

$$\left\{\mu_2 \in \Delta(\Theta) : \phi(\mu_2) \geq \phi(\mu_1)\right\} \subseteq B(\mu_1) \quad \text{for each } \mu_1 \in \Delta(\Theta),$$

then (A, u) is non-binding for B.

For each psychological sub-utility ϕ, we can define the dual belief correspondence $B_\phi(\mu_1) := \{\mu_2 \in \Delta(\Theta) : \phi(\mu_2) \geq \phi(\mu_1)\}$ that satisfies the condition in Lemma 2. Therefore, for each decision problem (A, u) that is separable with respect to ϕ, the problem is non-binding for B_ϕ. In this sense, each sub-utility ϕ naturally induces a belief correspondence that allows the agent to maximize her utility without restriction. For example, the result applies to binary-state settings in which (1) the reasoning constraint B is increasing, and (2) the decision problem is separable with an increasing belief-dependent sub-utility ϕ. (See Section 6.)

The following result shows that if a given decision problem (A, u) is non-binding for B, then the welfare result is restored and more information is weakly better.

Proposition 3. *Fix a decision problem (A, u) and a reasoning constraint B. Assume (A, u) is non-binding for B. If (S, π) is more informative than (S', π'), then for all $\mu_0 \in \Delta(\Theta)$, the agent is weakly better off under (S, π) than under (S', π').*

Proof. By assumption, for the given decision problem (A, u), the utility V_B achieved by the agent coincides with the utility of an agent with unconstrained reasoning, $V_B = V$. Since Proposition 1 implies that V is convex, V_B is also convex. □

6 Application: binary decision problems

In this section, we apply our characterization result to binary decision problems. We will consider the reasoning constraints introduced in Section 4.1 in the context of a binary state space, binary actions, and additive belief-dependent preferences.

Consider an uncertain patient facing a medical decision. The patient may have decent health ($\theta = H$) or low health ($\theta = L$) but does not know her condition. A minor treatment ($a = h$) would be optimal if the patient is in decent health, while surgery ($a = l$) would be optimal if her health is low. The patient receives a psychological reward ϕ that depends on her belief about her condition, say, she benefits from believing she has decent health (ϕ is increasing). Before deciding on the treatment, suppose the patient can choose between more or less informative tests about her health status. How should the patient rank different tests? Depending on which beliefs the patient can choose to have—according to her reasoning constraint B—the characterization in Theorem 1 identifies whether being better informed is always beneficial for the patient.

Formally, take a simple binary decision problem with state space $\Theta = \{L, H\}$ and action space $A = \{l, h\}$ in which the agent wishes to match the state with the right action (surgery $a = l$ for low health ($\theta = L$) and a minor treatment $a = h$ for decent health ($\theta = H$)). Action $a = l$ provides a payoff of $x \in (0, 1)$ regardless of the state. Action $a = h$ provides a payoff of 0 if the state is L and a payoff 1 if the state is H.

The agent's belief-dependent utility function is given by

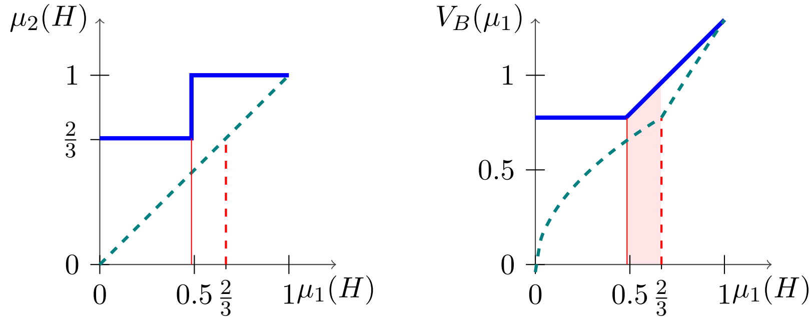
$$u(a, \theta, \mu_2) = x \cdot \mathbb{1}[a = l] + \mathbb{1}[a = h \text{ and } \theta = H] + \phi\big(\mu_2(H)\big), \tag{1}$$

where $\mu_2(H)$ is the agent's subjective belief that the state is H in period 2 when taking the action and $\phi : [0, 1] \to \mathbb{R}$ is a increasing mapping that captures the agent's belief-dependent preferences.[20] The application assumes additive preferences, thereby separating the effects of the action from the preferences over beliefs.[21]

To provide more structure to the problem, we assume $x = \frac{2}{3}$ and a uniform prior belief that assigns equal probability to each state. We assume that psychological preferences are captured by the utility mapping $\phi(\mu_2(H)) = \sqrt{\mu_2(H)} - \sqrt{0.5}$. So, ϕ exhibits diminishing returns, provides a psychological reward for optimistic beliefs (higher than 0.5), and provides a psychological cost for pessimistic beliefs (lower than 0.5). Given an objective belief μ_1,

[20]Absent belief-dependent preferences, the expected benefit from $a = h$ is μ_1 and the benefit from $a = \ell$ is x. Hence, the agent chooses $a = h$ whenever $\mu_1 \geq x$.

[21]Note, our main theorem does not require this separability assumption.



(a) Subjective beliefs (b) Value functions

Figure 6.1. Optimal subjective belief (left) and expected utility (right) of an unconstrained/conservative/increasing agent (blue solid) and a Bayesian agent (teal dashed).

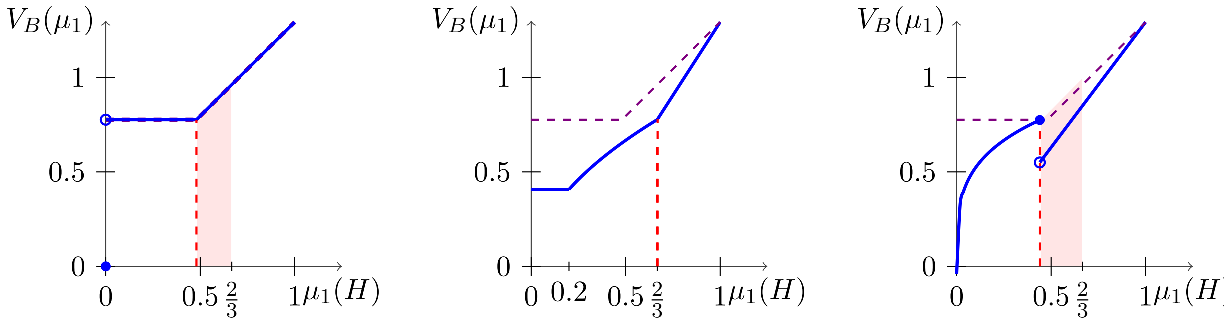
Self 1's optimization problem is:

$$\max_{\mu_2 \in B(\mu_1)} \quad \mathbb{E}_{\mu_1}\left[\tfrac{2}{3} \cdot \mathbb{1}[a^*(\mu_2) = l] + \mathbb{1}[a^*(\mu_2) = h \text{ and } \theta = H] + \sqrt{\mu_2(H)} - \sqrt{0.5}\right]$$
$$\text{s.t.} \quad a^*(\mu_2) \in \arg\max_{a \in \{l,h\}} \mathbb{E}_{\mu_2}\left[\tfrac{2}{3} \cdot \mathbb{1}[a = l] + \mathbb{1}[a = h \text{ and } \theta = H]\right].$$

We will study the agent's beliefs and value functions for unconstrained reasoning and the reasoning constraints discussed in Section 4.1.

Unconstrained and Bayesian We first compare unconstrained and Bayesian reasoning. Under Bayesian reasoning, the subjective belief μ_2 equals the objective belief μ_1. The left panel in Figure 6.1 illustrates in teal-dashed the beliefs of an agent with Bayesian reasoning. The agent changes her action at $\mu_1(H) = 2/3$ from $a = \ell$ to $a = h$. The teal-dashed line in the right panel describes the agent's expected utility function $V_B(\mu_1)$. Since the Bayesian reasoning constraint is not shrinking in support, more information can make the agent worse off. Indeed, the expected utility function $V_B(\mu_1)$ is not convex in μ_1.

Now consider a patient with unconstrained reasoning who can freely choose any belief about her health, regardless of past evidence or conditions, $\mu_2 \in \Delta(\Theta)$. The left panel in Figure 6.1 illustrates the optimally chosen subjective belief $\mu_2(H)$ (blue-solid) as a function of the objective belief $\mu_1(H)$. Since in our example the agent's payoff increases in the belief, all subjectively chosen beliefs are weakly larger than the objective Bayesian belief (teal dashed). Observe, however, that Self 1 does not always choose the maximal belief $\mu_2(H) = 1$, which would induce Self 2 to always take action $a = h$. Instead, for low levels of the objective belief $\mu_1(H)$, Self 1 chooses the maximal belief $\mu_2(H) = \frac{2}{3}$ that is consistent with the low action $a = \ell$. The right panel illustrates the value function (blue-solid) of an agent with unconstrained reasoning. Note that indeed the value function is convex. By Proposition 1

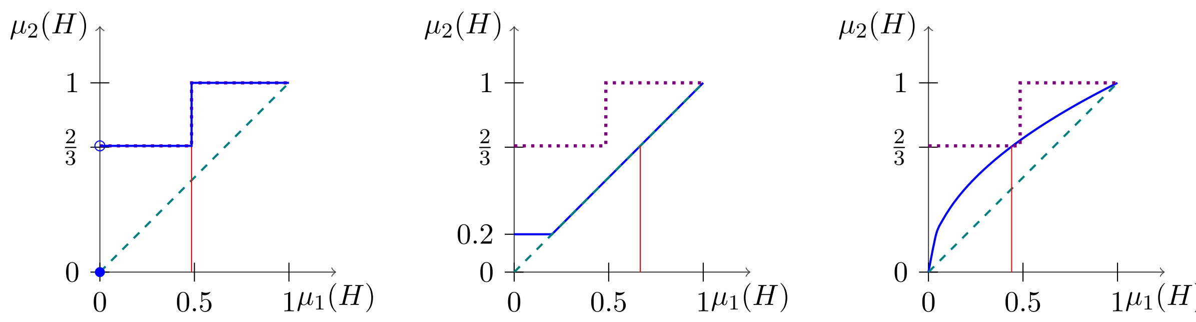


(a) Deductive (b) Reference-biased ($\nu = 0.2$) (c) Exog. distorted ($\sqrt{\cdot}$)

Figure 6.3. Value functions (blue-solid) for (a) deductive, (b) reference-biased, and (c) exogenously distorted reasoning constraints. As benchmarks in each figure is the value function for unconstrained/conservative/increasing (purple dotted) reasoning.

information always increases welfare for patients with unconstrained reasoning. The intuition is straightforward: since the patient can freely choose her belief, she can simply disregard and forget any unfavorable information revealing low health.

Note that in our example, an unconstrained agent changes her action at the cutoff-belief $\mu_1(H) = 0.48$ from $a = \ell$ to $a = h$. The red-shaded area in the right panel illustrates the objective beliefs at which Self 2 chooses a suboptimal action (from Self 1's perspective). The optimal action is $a = l$, as the belief satisfies $\mu_1(H) \leq \frac{2}{3}$, but instead, the agent chooses action $a = h$. This illustrates the trade-off for Self 1 between choosing "pleasant" beliefs and the potential wrong action that Self 2 might take based on the chosen subjective belief. Intuitively, in this region, the gain from an increased belief is larger than the loss from an increased action.



(a) Deductive (b) Reference-biased ($\nu = 0.2$) (c) Exogenously distorted ($\sqrt{\cdot}$)

Figure 6.2. The agent's optimally chosen subjective beliefs $\mu_2(H)$ (blue solid) as a function of the objective beliefs $\mu_1(H)$, for (a) deductive, (b) reference-biased, and (c) exogenously distorted reasoning constraints. As benchmarks in each figure are the optimal beliefs of the Bayesian (teal dashed) and the unconstrained/conservative/increasing (purple dotted) reasoning.

Conservative Consider a patient with conservative reasoning constraint who can change her beliefs about her health; however, as long as there is no certainty about her condition she cannot discard either state. At first glance, it may not be obvious whether more information would help or hurt such a patient. However, observe that the conservative reasoning constraint $B(\mu) = \{\mu' \in \Delta(\Theta) : \text{Supp}(\mu) \subseteq \text{Supp}(\mu')\}$ is shrinking in support. Consequently, for any well-defined decision problem, by Theorem 1 more information is welfare-improving for a patient with conservative reasoning.[22] The agent's optimal subjective belief coincides with the unconstrained belief, illustrated in blue-solid in the left panel in Figure 6.1. Likewise, considering the supremum value (as opposed to the maximum), the conservative value function V_B coincides with the unconstrained value function depicted by the blue-solid line in the right panel in Figure 6.1.

Deductive Consider a patient with deductive reasoning who can choose any possible belief, i.e., any belief that has not been ruled out with certainty. The deductive reasoning constraint $B(\mu) = \{\mu' \in \Delta(\Theta) : \text{Supp}(\mu') \subseteq \text{Supp}(\mu)\}$ is *not* shrinking in support.[23] Therefore, more information can be detrimental for patients with deductive reasoning. Note, however, the only case in which information can indeed be harmful is when the medical test fully reveals that the agent has low health. Thus, if it is the case that only full-support experiments are available, Proposition 2 implies that a patient with deductive reasoning is unambiguously better off with more information. In Figure 6.3, the value function of a patient with deductive reasoning (blue-solid) coincides with the unconstrained value function (purple-dashed), except for the case that $\mu(H) = 0$ in which the patient knows that she has low health with certainty. Compared to the optimal action of a Bayesian patient, the deductive patient switches from surgery to minor treatment at a lower belief. The left panel in Figure 6.2 illustrates the agent's beliefs in blue-solid.

Reference-biased Consider a patient with reference-biased reasoning who always incorporates an exogenous reference belief ν in her belief choice (assigning, say, probability 0.2 to state H). For example, the patient might stick to the belief that she has good health with probability ν, as she convinced of some health pattern in her family history. The agent's subjective belief under a reference-biased reasoning constraint is illustrated in the central panel in Figure 6.2 in blue-solid. Note that the reference-biased reasoning constraint

[22] In this application the decision problem is may not be well defined as the conservative correspondence is not compact-valued. However, the decision problem is well defined whenever the psychological reward ϕ has finite range. (See Lemma A.1.)

[23] For instance, consider μ with $\text{Supp}(\mu) = \{1\}$ and ν with $\text{Supp}(\nu) = \{0, 1\}$. Then $\text{Supp}(\mu) \subseteq \text{Supp}(\nu)$ but $B(\nu) \nsubseteq B(\mu)$.

$B(\mu) = \{\lambda\mu + (1-\lambda)\nu : \lambda \in [0,1]\}$ is not shrinking in support.[24] Thus, more information is not always beneficial for patients with reference-biased reasoning. In the central panel in Figure 6.3, the reference-biased value function (blue-solid) is not convex and lies below the unconstrained value function in violet-dashed. Intuitively, the patient lacks the flexibility to avoid harmful beliefs.

Increasing Consider an optimistic patient who can ignore negative test results about her health condition. As a result, the patient can bias her belief towards good health but cannot bias it towards low health. The agent's subjective belief under an increasing reasoning constraint is illustrated in the left panel in Figure 6.1 in blue solid. It coincides with the optimal unconstrained belief. Notice, the increasing reasoning constraint $B(\mu) = \{\mu' :$ μ' first-order stochastically dominates $\mu\}$ is *not* shrinking in support.[25] Therefore, more information can be harmful for *some* psychological decision problems. However, in this application the belief-dependent part ϕ of the utility is increasing. So, the patient would never choose to decrease her subjective belief. In other words, (A, u) is non-binding for B and, as a consequence, information weakly increases the patient's welfare in this particular problem. (See Lemma 2 and Proposition 3.) In the left panel in Figure 6.1, the increasing value function is identical to the (convex) unconstrained value function. However, if ϕ were decreasing (i.e., the patient would prefer to think that her health is low), then some information sources would hurt the patient.

Exogenously distorted Consider a patient who has a non-Bayesian distortion in the way she updates her beliefs. An exogenously distorted reasoning constraint $B(\mu) = \{f(\mu)\}$ is shrinking in support if and only if f is constant.[26] So, Blackwell monotonicity holds if and only if the patient is unresponsive to information. The distorted subjective belief with $f(\cdot) = \sqrt{\cdot}$ is illustrated in the right panel in Figure 6.2. The associated value function is non-convex and below the unconstrained value function (depicted in blue-solid in the right panel in Figure 6.3). Moreover, the patient switches from surgery to minor treatment at a belief that is lower than the cutoff of a patient with unconstrained reasoning.

[24] Observe that each belief with full support $\mu \neq \nu$ satisfies $B(\mu) \neq B(\nu) = \{\nu\}$. Since shrinking in support requires the same set of feasible beliefs on all beliefs with full support, B is not shrinking in support.

[25] To see this, consider beliefs μ assigning probability 0.3 to state H and μ' assigning probability 0.7 to state H. Let $\overline{\mu} = 0.5\mu + 0.5\mu'$, which assigns probability 0.5 to state H. Then $B(\overline{\mu})$ contains beliefs that assign probability 0.6 to state H, which are not in $B(\mu')$ because $\overline{\mu}$ does not stochastically dominate μ'. Thus, $B(\overline{\mu}) \not\subseteq B(\mu) \cap B(\mu')$, violating the condition for shrinking in support. (See Lemma 1.)

[26] To see this, note that if f is constant, then $B(\mu)$ is the same for all μ, trivially satisfying the shrinking-in-support condition. Conversely, if f is not constant, there exist μ and μ' such that $f(\mu) \neq f(\mu')$, which implies that $B(\mu) \cap B(\mu') = \emptyset$, violating the shrinking-in-support condition.

7 Discussion

Our paper presents a precise characterization of when individuals engaging in motivated reasoning will seek more informative sources. We show that agents always acquire the most informative sources if and only if their reasoning constraint is shrinking in support—a condition that allows them to maintain desirable beliefs even when confronted with challenging evidence. In particular, when an agent faces no restrictions, the agent always welcomes more information and thus Blackwell monotonicity is satisfied.

7.1 Information avoidance

Do individuals always seek more information? As Caplin and Leahy (2001) (p.9) write, "It is now widely accepted that not all individuals want or benefit from information [...]." For agents that avoid information, Blackwell monotonicity does not apply, and our characterization result implies that (i) those agents have restrictions in their belief choice, and (ii) the reasoning of such agents violates the shrinking-in-support condition. Shrinking-in-support allows the agent to protect herself by reframing the possibilities: as information rules out states, she gains weakly more freedom in interpreting, narrating, or distorting the remaining uncertainty. Under such reasoning, information is never intrinsically harmful, as it never forces the agent into a narrower set of beliefs.

We take no normative stand on the shrinking-in-support condition, and we do not claim whether or not it is a reasonable description of belief formation. Some agents may have limited ability to reframe narratives that shape beliefs, so the condition fails to hold. In contexts in which agents use incoming information as raw material to construct richer narratives, the condition may be more plausible.[27]

Our paper provides a positive analysis of information acquisition, providing a unifying lens through which a wide class of models of motivated-reasoning can be classified according to whether agents seek or avoid information. Our main result identifies the shrinking-in-support condition as necessary and sufficient for Blackwell monotonicity. Moreover, it provides a testable prediction about endogenous information acquisition in the context of motivated reasoning: Agents who actively avoid sources of information must have non-trivial boundaries on what beliefs they can achieve, as the unrestricted belief correspondence satisfies Blackwell monotonicity.

[27] In particular children can often be very creative in generating internal protective beliefs in response to devastating facts. For instance, Wente et al. (2020) find evidence that children bias their beliefs toward desired outcomes, letting wishful thinking shape how they interpret the world they live in. More generally, as part of narrative therapy, individuals learn to reinterpret life events and create life-enhancing narratives (APA, 2026).

7.2 Preferences for ex-post beliefs

Our main model considers preferences for *interim* beliefs, i.e., preferences over the beliefs the agent holds at the time of taking the action. Preferences for ex-post beliefs capture important psychological motivations such as regret or self-esteem. (See, for instance, Battigalli and Dufwenberg (2009); Mannahan (2023).) This subsection extends our framework to decision problems in which the agent has preferences over both *interim* and *ex-post* beliefs.

In this extended setting, the agent first chooses an action and then observes an outcome $y \in Y$. The outcome y is drawn from a distribution that depends on both the chosen action $a \in A$ and the true state $\theta \in \Theta$. Write $\gamma : A \times \Theta \to \Delta(Y)$ for the (exogenous) distribution of outcomes. Observe that the outcome y provides partial (or complete) information about the state, depending on the outcome mapping γ. The agent's preferences are captured by a measurable utility function

$$v : A \times Y \times \Theta \times \Delta(\Theta) \times \Delta(\Theta) \to \mathbb{R}.$$

So, $v(a, y, \theta, \mu_2, \mu_3)$ describes the agent's utility from a path in which she chooses action a, receives outcome y, the state is θ, has interim belief μ_2, and has ex-post belief μ_3. Note, since the utility function v depends on μ_3 and θ, it can be interpreted as the reduced-form of a multi-stage setting, allowing v to capture the expected utility over all future uncertainty.

The agent's distribution of ex-post beliefs is given by a measurable mapping

$$b : A \times Y \times \Theta \times \Delta(\Theta) \to \Delta(\Theta).$$

So, $\mu_3 = b(a, y, \theta, \mu_2)$ is the ex-post belief conditional on the chosen action a, the outcome y, the state θ, and the interim belief μ_2. We allow this ex-post belief to emerge through Bayes rule, an exogenous distortion, or any other belief-formation mechanism. Note, this extended framework includes models such as Niehaus (2020) (who considers a deductive correspondence) in which the agent can choose both interim and ex-post beliefs.

Within this extended framework, define the mapping $u : A \times \Theta \times \Delta(\Theta) \to \mathbb{R}$ by

$$u(a, \theta, \mu_2) := \int_{y \in Y} v\big(a, y, \theta, \mu_2, b(a, y, \theta, \mu_2)\big)\ d\gamma(a, \theta).$$

Thus, conditional on action a, state θ, and interim belief μ_2, $u(a, \theta, \mu_2)$ is the expectation over the outcomes Y of the agent's utility v. By averaging the utility function v over the outcomes and the ex-post beliefs that they induce, the resulting function u expresses ex-post belief preferences in terms of interim beliefs. The function u thus describes the agent's preferences in terms of the action a, the state θ, and the interim belief μ_2. Consequently, all results in this paper apply to this extended setting as well.

A Appendix

Lemma A.1. *Fix a decision problem (A, u) and a reasoning constraint B. Suppose either (i) that u is continuous, A is compact, and B is compact-valued, or (ii) that the range of u is finite, then (A, u) is well defined for B.*

Proof. We divide the proof in two cases.

Case 1: Assume that u is continuous, A is compact, and B is compact-valued. First, we establish that $\mathcal{F}_B(\mu_1)$ is non-empty. This follows from two observations: (1) $B(\mu_1)$ is non-empty, and (2) for any belief $\mu_2 \in B(\mu_1)$, the set of optimal actions $A^*(\mu_2) := \arg\max_{a \in A} \mathbb{E}_{\mu_2} u(a, \theta, \mu_2)$ is non-empty due to the compactness of A and the continuity of u.

Next, we show that $\mathcal{F}_B(\mu_1)$ is compact. Since $B(\mu_1)$ and A are compact sets, their cartesian product $B(\mu_1) \times A$ is also compact. Since $\mathcal{F}_B(\mu_1)$ is a subset of $B(\mu_1) \times A$, to show $\mathcal{F}_B(\mu_1)$ compact it suffices to show that it is closed. Consider a convergent sequence $((\mu_2^n, a^n) : n \in \mathbb{N}) \subseteq \mathcal{F}_B(\mu_1)$ with limit point (μ_2^*, a^*). By definition, for each n, we have $\mu_2^n \in B(\mu_1)$ and $a^n \in \arg\max \mathbb{E}_{\mu_2^n} u(\cdot, \theta, \mu_2^n)$. Since $B(\mu_1)$ is closed (as it is compact), we have $\mu_2^* \in B(\mu_1)$. We will show that $a^* \in A^*(\mu_2^*)$. To show this, assume $a^* \notin A^*(\mu_2^*)$. Then, there is some $a' \in A$ such that

$$\mathbb{E}_{\mu_2^*} u(a', \theta, \mu_2^*) > \mathbb{E}_{\mu_2^*} u(a^*, \theta, \mu_2^*) = \lim_{n \to \infty} \mathbb{E}_{\mu_2^n} u(a^n, \theta, \mu_2^n),$$

where the equality follows from continuity of u. Moreover, by continuity of u, there is some $n \in \mathbb{N}$ such that

$$\mathbb{E}_{\mu_2^n} u(a', \theta, \mu_2^n) > \mathbb{E}_{\mu_2^n} u(a^n, \theta, \mu_2^n),$$

contradicting the fact that $a^n \in A^*(\mu_2^n)$. Thus, $a^* \in A^*(\mu_2^*)$ and as a consequence $(\mu_2^*, a^*) \in \mathcal{F}_B(\mu_1)$. This proves that $\mathcal{F}_B(\mu_1)$ is closed. Since $\mathcal{F}_B(\mu_1)$ is a closed subset of the compact set $B(\mu_1) \times A$, it is itself compact.

Finally, the function $h(\mu_2, a) = \mathbb{E}_{\mu_1} u(a, \theta, \mu_2)$ is continuous on the compact set $\mathcal{F}_B(\mu_1)$. By the Extreme Value Theorem, h attains its maximum, so there exists $(\mu_2^*, a^*) \in \mathcal{F}_B(\mu_1)$ such that:

$$\mathbb{E}_{\mu_1} u(a^*, \theta, \mu_2^*) = \max_{(\mu_2, a) \in \mathcal{F}_B(\mu_1)} \mathbb{E}_{\mu_1} u(a, \theta, \mu_2) = V_B(\mu_1)$$

Thus, the supremum in $V_B(\mu_1)$ is attained, proving that (A, u) is well defined for B.

Case 2: Denote $R \subseteq \mathbb{R}$ as the range of u and assume that the range R is finite. Fix an objective belief μ_1 and write $\overline{u}(a, \mu_2) = \mathbb{E}_{\mu_1} u(a, \theta, \mu_2)$ for the (objective) expected utility of action a given subjective belief μ_2. Notice, since $\overline{u}$ is a finite sum and the range of u is finite, the mapping $\overline{u}$ also has finite range. Write $\overline{R}$ for the range of $\overline{u}$.

Note that for each $\mu_1 \in \Delta(\Theta)$, the set of feasible expected utilities satisfies

$$\left\{\mathbb{E}_{\mu_1} u(a, \theta, \mu_2) : (\mu_2, a) \in \mathcal{F}_B(\mu_1)\right\} \subseteq \overline{R}.$$

Since $\overline{R}$ is a finite set, the supremum $V_B(\mu_1)$ is attained at some $(\mu_2^*, a^*) \in \mathcal{F}_B(\mu_1)$. Therefore, (A, u) is well defined for B. □

Proof of Lemma 1*. Only if.* Assume B is shrinking in support. Notice, if $\overline{\mu} = \lambda\mu + (1-\lambda)\mu'$, then $\text{Supp}\,(\mu) \subseteq \text{Supp}\,(\overline{\mu})$ and $\text{Supp}\,(\mu') \subseteq \text{Supp}\,(\overline{\mu})$. Hence $B(\overline{\mu}) \subseteq B(\mu)$ and $B(\overline{\mu}) \subseteq B(\mu')$. Therefore, $B(\overline{\mu}) \subseteq B(\mu) \cap B(\mu')$, as desired.

If. Fix $\mu, \overline{\mu} \in \Delta(\Theta)$ such that $\text{Supp}\,(\mu) \subseteq \text{Supp}\,(\overline{\mu})$. To show that B is shrinking in support, it suffices to show that $B(\overline{\mu}) \subseteq B(\mu)$. Observe that for all $\theta \in \text{Supp}\,(\overline{\mu})$,

$$\lim_{\lambda \to 0} \tfrac{1}{1-\lambda}\overline{\mu}(\theta) - \tfrac{\lambda}{1-\lambda}\mu(\theta) = \overline{\mu}(\theta) \in [0, 1].$$

Hence, there is some sufficiently small $\lambda' \in (0, 1)$, such that

$$\mu'(\theta) := \tfrac{1}{1-\lambda'}\overline{\mu}(\theta) - \tfrac{\lambda'}{1-\lambda'}\mu(\theta) \in [0, 1] \quad \forall \theta \in \Theta.$$

Hence, $\mu' \in \Delta(\Theta)$. Notice that $\overline{\mu} = \lambda'\mu + (1-\lambda')\mu'$. By hypothesis, $B(\overline{\mu}) \subseteq B(\mu) \cap B(\mu')$. Therefore, $B(\overline{\mu}) \subseteq B(\mu)$, as desired. □

Proof of Lemma 2*.* Assume that $\left\{\mu_2 \in \Delta(\Theta) : \phi(\mu_2) \geq \phi(\mu_1)\right\} \subseteq B(\mu_1)$.

Fix an objective belief $\mu_1 \in \Delta(\Theta)$. Observe, the expected utility of an agent with unconstrained reasoning is maximized at some

$$(\mu_2^*, a^*) \in \arg\max_{\Delta(\Theta)\times A} \mathbb{E}_{\mu_1} u(a, \theta, \mu_2) = \arg\max_{\Delta(\Theta)\times A} \mathbb{E}_{\mu_1}\big[v(a, \theta) + \phi(\mu_2)\big].$$

So, in particular, $\phi(\mu_2^*) \geq \phi(\mu_1)$ and thus $\mu_2^* \in B(\mu_1)$. Therefore, $V_B(\mu_1) = V(\mu_1)$ and the decision problem (A, u) is non-binding for B. □

References

George A Akerlof and William T Dickens. The economic consequences of cognitive dissonance. *American Economic Review*, 72(3):307–319, 1982.

American Psychological Association APA. Dictionary of psychology, 2026. URL https://dictionary.apa.org/narrative-therapy.

Pierpaolo Battigalli and Martin Dufwenberg. Dynamic psychological games. *Journal of Economic Theory*, 144(1):1–35, 2009.

Pierpaolo Battigalli and Nicolò Generoso. Information flows and memory in games. *Available at SSRN 4435785*, 2023.

Roland Bénabou. The economics of motivated beliefs. *Revue d'économie politique*, 125(5):665–685, 2015.

Roland Bénabou and Jean Tirole. Self-confidence and personal motivation. *Quarterly Journal of Economics*, 117(3):871–915, 2002.

Roland Bénabou and Jean Tirole. Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–164, 2016.

David Blackwell. Comparison of experiments. *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, 2:93–103, 1951.

David Blackwell. Equivalent comparisons of experiments. *The annals of mathematical statistics*, pages 265–272, 1953.

J Aislinn Bohren and Daniel N Hauser. *The behavioral foundations of model misspecification: A decomposition*. Penn Institute for Economic Research, 2023.

Markus K Brunnermeier and Jonathan A Parker. Optimal expectations. *American Economic Review*, 95(4):1092–1118, 2005.

Andrew Caplin and John Leahy. Psychological expected utility theory and anticipatory feelings. *The Quarterly Journal of Economics*, 116(1):55–79, 2001.

Andrew Caplin and John V Leahy. Wishful thinking. Technical report, National Bureau of Economic Research, 2019.

Olivier Compte and Andrew Postlewaite. Confidence-enhanced performance. *American Economic Review*, 94(5):1536–1557, 2004.

Jacques Crémer. A simple proof of Blackwell's "comparison of experiments" theorem. *Journal of Economic Theory*, 27(2):439–443, 1982.

Jason Dana, Roberto A Weber, and Jason Xi Kuang. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33:67–80, 2007.

Geoffroy de Clippel and Xu Zhang. Non-Bayesian persuasion. *Journal of Political Economy*, 130(10):2594–2642, 2022.

Henrique de Oliveira. Blackwell's informativeness theorem using diagrams. *Games and Economic Behavior*, 109:126–131, 2018.

Jonathan St BT Evans and Keith E Stanovich. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3):223–241, 2013.

Kate Faasse. Nocebo effects in health psychology. *Australian Psychologist*, 54(6):453–465, 2019.

John Geanakoplos, David Pearce, and Ennio Stacchetti. Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60–79, 1989.

Jeanne Hagenbach and Frédéric Koessler. Selective memory of a psychological agent. *European Economic Review*, 142:104012, 2022.

Paul Heidhues, Botond Kőszegi, and Philipp Strack. Unrealistic expectations and misguided learning. *Econometrica*, 86(4):1159–1214, 2018.

Alexander M Jakobsen. Coarse Bayesian updating. *Working paper*, 2025.

Daniel Kahneman. *Thinking, fast and slow.* macmillan, 2011.

Richard E Kihlstrom. A 'Bayesian'exposition of Blackwell's theorem on the comparison of experiments. *Bayesian models in economic theory*, pages 13–31, 1984.

Botond Köszegi. Ego utility, overconfidence, and task choice. *Journal of the European Economic Association*, 4(4):673–707, 2006.

Botond Kőszegi. Emotional agency. *The Quarterly Journal of Economics*, 121(1):121–155, 2006.

Ziva Kunda. The case for motivated reasoning. *Psychological Bulletin*, 108(3):480, 1990.

Moshe Leshno and Yishay Spector. An elementary proof of Blackwell's theorem. *Mathematical Social Sciences*, 25(1):95–98, 1992.

Elliot Lipnowski and Laurent Mathevet. Disclosure to a psychological audience. *American Economic Journal: Microeconomics*, 10(4):67–93, 2018.

Andrew T Little, Keith E Schnakenberg, and Ian R Turner. Motivated reasoning and democratic accountability. *American Political Science Review*, 116(2):751–767, 2022.

Rachel Mannahan. Self-esteem and rational self-handicapping. *Available at SSRN 4422597*, 2023.

Stephen Morris and Hyun Song Shin. The rationality and efficacy of decisions under uncertainty and the value of an experiment. *Economic Theory*, 9:309–324, 1997.

Paul Niehaus. A theory of good intentions. *working paper*, 2020.

Pietro Ortoleva. Alternatives to Bayesian updating. *Annual Review of Economics*, 16, 2024.

Ernesto Rivera Mora. Mechanism design with belief-dependent preferences. *Journal of Economic Theory*, 216:105782, 2024.

Timothy M Scarella, Robert J Boland, and Arthur J Barsky. Illness anxiety disorder: psychopathology, epidemiology, clinical characteristics, and treatment. *Psychosomatic medicine*, 81(5):398–407, 2019.

Joel Sobel. Lying and deception in games. *Journal of Political Economy*, 128(3):907–947, 2020.

Joel Sobel. On the relationship between damage and deception. *working paper*, 2024.

Ran Spiegler. On two points of view regarding revealed preference and behavioral economics. *The foundations of positive and normative economics: A handbook*, pages 95–115, 2008.

John F Todaro, Biing-Jiun Shen, Raymond Niaura, Avron Spiro III, and Kenneth D Ward. Effect of negative emotions on frequency of coronary heart disease (the normative aging study). *The American journal of cardiology*, 92(8):901–906, 2003.

William Von Hippel and Robert Trivers. The evolution and psychology of self-deception. *Behavioral and brain sciences*, 34(1):1–16, 2011.

Adrienne O Wente, Mariel K Goddu, Teresa Garcia, Elyanah Posner, María Fernández Flecha, and Alison Gopnik. Young children are wishful thinkers: The development of wishful thinking in 3-to 10-year-old children. *Child development*, 91(4):1166–1182, 2020.

Mark Whitmeyer. Blackwell-monotone updating rules. *Journal of Political Economy (Forthcoming)*, 2024.

Florian Zimmermann. The dynamics of motivated beliefs. *American Economic Review*, 110 (2):337–363, 2020.