



(1955). Under this premise, an agent faces, apart from the standard action choice, a problem of error management. In economic decision making, some mistakes in perception of payoff-relevant parameters are costlier than others, and this asymmetry has an impact on the frequency of different types of perception errors. In this paper, we relate error management to known observed behavioural biases. In order to understand the direction of biases arising under second-best perception strategies, we need first to understand the associated costs of under- or over- estimating payoff-relevant parameters.

Our agent first receives some statistical information on a state of nature which can be either high or low, and this information translates into an objective belief  $p$  that the state is high. In a choice stage, she recalls this information imperfectly and ends up with a subjective belief  $q$  that the state is high. She then chooses an action that is optimal under her subjective belief  $q$  within a fixed choice set. Due to belief distortion, her choice may differ from the optimal one. We ask how large is the payoff loss incurred from the misperception of  $p$  for  $q$ .

We derive a simple and intuitive characterization of the loss from misperceiving  $p$  as  $q$ . This loss can be expressed as an integral formula that depends only on the second derivative  $v''$  of the value function. The value function  $v(p)$ , for each probability  $p$  of the high state, specifies the payoff of the optimizing agent who has correct perception. This characterization makes the loss from misperception simple to compute in several applications in which the direct computation is cumbersome.

Equipped with this loss characterization, we study a class of error-management problems in which the agent chooses a distribution of perception errors that performs well across all decision problems that she encounters in her environment. We follow principles from the ecological rationality literature, in that our agent is unable to reoptimize the perception strategy in each encountered decision problem separately but, instead, must choose a perception heuristic that fits her environment.<sup>1</sup>

The model is as follows. Based on all available information, the agent forms an objective probability  $p$  of a high payoff state. She then memorizes a probability  $m$  in the perception stage, and recalls a perturbed value  $q = m + \varepsilon$  in the choice stage, where  $\varepsilon$  is a noise term with symmetric density. Then, she encounters a random decision problem drawn from her environment and chooses an action that she perceives as optimal under her belief  $q$  and the encountered utility function.

---

<sup>1</sup>See Simon (1956) and Gigerenzer and Todd (1999) on ecological rationality.

One feasible perception strategy is to memorize the true probability value. Under such a strategy, the recalled probability is in expectation equal to the true one. It turns out, however, that this unbiased perception strategy is generically suboptimal, and the agent benefits from memorizing  $m$  distinct from the true probability  $p$ .

We characterize the direction of the optimal perception bias based on monotonicity properties of the second derivative  $v''$  of the value function. When  $v''$  is decreasing in a neighborhood of the true observed probability, we show that the agent memorizes a probability higher than the true one and thus exhibits an upward perception error bias. When  $v''$  is increasing, her memory exhibits a downward bias. In applications with Gaussian distributions of payoffs, the value function  $v$  is easy to compute, and the direction of monotonicity of  $v''$  is easy to characterize. Our results then allow us to make clear-cut predictions on the direction of the bias in these environments.

We detail here one of our applications to illustrate why second-best perception strategies may be biased. Consider an agent who chooses between two binary lotteries. Each available lottery is defined by the rewards it pays in the high and low states of the world. The true probability of the high state is  $p$ , but the agent perceives this probability to be  $q = m + \varepsilon$ , and chooses the lottery that maximizes the perceived expected reward. The agent's environment contains a whole range of such decision problems, where each decision problem is represented by all four lottery rewards. We model this by assuming that all four rewards are drawn iid. from a standard normal distribution. The perception strategy  $m(p)$  is chosen before the agent observes the realization of these draws, hence independently of the lottery rewards.

A misperception of  $q$  instead of  $p$  causes a loss only when the optimal choices at  $p$  and  $q$  differ. This happens precisely when there exists a probability  $s$  between  $p$  and  $q$  such that the agent is indifferent at  $s$  between the two lotteries. The likelihood that such a tie arises at  $s$  depends on  $s$  in an intuitive way. Since the expected value of each lottery is a convex combination of the two standard normal draws, its variance is lower the closer  $s$  is to  $1/2$ . Thus, the likelihood of a tie at probability  $s$  is a single-peaked function of  $s$  attaining its maximum at  $1/2$ . This implies that misperceptions of  $p$  for a value  $q$  towards the direction of  $1/2$  are more likely to distort choices than misperceptions in the opposite direction: undervaluation of one's own ability to predict the binary state leads to suboptimal choice more often than the symmetric opposite error. Since the agent can shift the distribution of her perception error by controlling the memorized probability value, it is optimal for her to memorize a value further away from  $1/2$  than the truth, hence to exhibit an overprecision

bias.

The above intuition indicates that the optimal perception strategy depends crucially on at which probabilities the agent is likely to experience a tie between the available actions. Given a draw of the lottery rewards, the value function is piecewise linear and has a kink (or non-differentiability point) at a probability leading to the tie. Ex ante, across all the decision problems,  $v''(q)$  thus reflects the likelihood that the agent encounters a decision problem in which such a tie occurs at  $q$ . This explains why  $v''(q)$ , the main object of our analysis, captures the likelihood of a tie occurring at probability  $q$ .

Apart from the overprecision application, we illustrate our methodology in three additional examples: illusion of control, order effect, and probability weighting. In each application, we make natural assumptions about the distribution of utility functions encountered by the agent in her environment and derive the optimal bias that arises. We relate these biases to stylized facts from psychology and behavioural economics. In the illusion-of-control example, the agent overvalues her impact on her own well-being. In the order-effect application, she exaggerates the quality difference among two available objects observed in the first of the two periods. In the probability-weighting example we revisit the model of Steiner and Stewart (2016), who derive the overweighting of small probabilities as a second-best perception heuristic.

## 1.1 Related literature

Rate-distortion theory studies optimal communication via a noisy channel, Shannon (1948, 1959). One of the primitives of this theory is an exogenous map that specifies a loss to each input and output of the communication. A popular loss function is the square error. In our paper, the loss function is derived from the agent's environment and is equal to the average welfare loss caused by the distortion.<sup>2</sup> Once we establish the relevant loss function, we let the agent engage in optimal error management: she avoids the costlier types of the errors. Beyond information theory, error management has been studied in several scientific disciplines such as biology and psychology. See Johnson et al. (2013) for an interdisciplinary literature review and Alaoui and Penta (2016) for a recent axiomatization of the cost-benefit approach to error management within economics. We contribute to this literature a formal characterization of loss from misperception based on a statistical description of the agent's

---

<sup>2</sup>An appropriate definition of the loss function is a nontrivial and important part of the data compression problem. See the discussion of the loss function for compression of images in Wang et al. (2004).

environment.

Sims (1998, 2003) has introduced an exogenous information-theoretic capacity constraint to economics. The subsequent literature on rational inattention studies the information acquisition and processing of an agent who, as in our model, is unsure about a payoff-relevant parameter but who, unlike our agent, knows her payoff function. In that setting (assuming signal cost is nondecreasing in Blackwell informativeness), the agent acquires only an action recommendation and no additional information beyond that needed for choice. In our model, the agent does not know her payoff function when she processes information and thus forms beliefs beyond those needed for the mere action choice in any given problem. The resulting optimal information structure can then be naturally interpreted as a perception of the payoff parameter.<sup>3</sup>

Our assumption that the perception strategy is optimized across many decision problems has appeared in the literature on the evolution of the utility functions. This literature studies the performance of decision processes across a distribution of the fitness rewards to the available actions. Depending on the constraints assumed, the optimal decision criterion is either the expected utility maximization, as in Robson (2001), or its behavioural variants, as in Rayo and Becker (2007) and Netzer (2009). (See Robson and Samuelson (2010) for a literature review.) Most of this literature focuses on choice under certainty. Two exceptions who, like us, focus on the probability perception are Herold and Netzer (2010), who propose that biases in probability perception serve as a correction of another behavioural distortion; and Compte and Postlewaite (2012), who conclude that an agent with limited memory benefits from ignoring weakly informative signals.

We contribute to the branch of the bounded rationality literature that emphasizes limited memory. Mullainathan (2002) studies the behavioural implications of exogenous imperfect memory usage. Like us, Dow (1991), Hirshleifer and Welch (2002) and Wilson (2014) study the behavioural implications of an agent who optimizes her usage of limited memory, but they examine effects that are different from ours. In his survey, Lipman (1995) focuses on memory-related frictions.

This paper generalizes the model of Steiner and Stewart (2016), who study optimal probability perception in a choice between a binary lottery with random rewards and a fixed outside option. The contribution of this paper beyond Steiner and Stewart is threefold: (1) We significantly enlarge the class of settings by allowing arbitrary payoff distributions

---

<sup>3</sup>Within the rational inattention literature, Woodford (2012a,b) studies the behavioural biases resulting from perception frictions; in these studies he assumes the mean square error loss criterion.

and arbitrary action sets; (2) we characterize the perception loss function from the more fundamental value of information function, and (3) we deliver new behavioural insights.

Two of our derived biases, the illusion of control and overprecision, are classified among so-called positive illusions (Taylor and Brown, 1988). Such positive belief biases are often rationalized by assuming that the agent derives felicity from holding favourable beliefs about herself; see for example Brunnermeier and Parker (2005), Caplin and Leahy (2001), and Köszegi (2006). In contrast, belief distortions in our model are purely instrumental—they guide choice—and arise in the absence of any felicity benefits.

The paper is organized as follows. Section 2 introduces the model, and our loss characterization is presented in section 3. We develop general results on the direction of biases in section 4, and applications in section 5. We show an example of a sophisticated Bayesian agent in section 6. Section 7 concludes.

## 2 A Model of Misperceptions

An agent faces a randomly drawn decision problem. She chooses an action  $a \in A = \{1, \dots, n\}$ ,  $n \geq 2$ , and receives payoff  $u(a, \theta)$  where  $\theta \in \{0, 1\}$  is a payoff state. The objective probability of the state  $\theta = 1$  given all agent's information is  $p$ . The agent's perception of the probability distribution over  $\theta$  is imperfect. Correspondingly, we distinguish between an objective probability  $p$  and a subjective probability  $q$  of the state 1. The objective and subjective probabilities of the state 0 are  $1 - p$  and  $1 - q$  respectively. The payoff functions  $u(a, \cdot)$  are extended from  $\theta \in \{0, 1\}$  to all beliefs in  $[0, 1]$  through the expected utility formula  $u(a, p) = pu(a, 1) + (1 - p)u(a, 0)$ .

We study a two-stage decision process in which a *perception stage* is followed by a *choice stage*. In the *choice stage*, the agent knows the realized payoff function  $u$  that defines the decision problem she faces, mistakenly perceives the true probability  $p$  as  $q$ , and chooses an action

$$a_{q,u}^* = \arg \max_{a \in A} u(a, q)$$

that maximizes her perceived payoff. Finally, her realized payoff is  $u(a_{q,u}^*, p)$ .<sup>4</sup>

The payoff function is uncertain in the preceding *perception stage*. Accordingly, the payoff function  $u = (u(a, \theta))_{a,\theta}$  and the probability  $p$  of state 1 are random variables

---

<sup>4</sup>In our framework, the maximizer is unique with probability one. We therefore ignore cases with multiple optima.

with supports  $\mathbb{R}^{2n}$  and  $[0, 1]$ , respectively, both admitting atomless density. We assume  $u$  and  $p$  are independent to ensure that the agent does not draw inferences about  $\theta$  from the observation of the payoffs  $u$ , and vice versa. The payoff of the optimizing agent who misperceives the objective probability  $p$  as  $q$ , averaged out across the draws of the payoff function  $u$ , is

$$V(p, q) = \mathbb{E}_u u(a_{q,u}^*, p).$$

The value function  $v(p) = V(p, p)$  represents the expected payoff of the optimizing agent who perceives the objective probability  $p$  correctly. The expected loss due to misperceiving the true probability  $p$  as  $q$  is then

$$L(p, q) = v(p) - V(p, q).$$

Our aim is to characterize how the distribution of payoff functions  $u$  translates into the loss function  $L$ .

### 3 Loss Characterization

Our main technical contribution is a characterization of the loss function  $L$  in terms of the value function  $v$ .<sup>5</sup>

**Lemma 1.** *Assume  $v$  is twice differentiable on  $[0, 1]$ . Then, the expected payoff loss due to misperceiving the probability  $p$  as  $q$  is*

$$L(p, q) = \int_p^q v''(s)(s - p)ds. \tag{1}$$

As defined, the loss function  $L$  depends on the multidimensional distribution of underlying payoff functions  $u$ , and it depends on two parameters,  $p$  and  $q$ . Lemma 1 shows that  $L$  is entirely characterized through the second derivative  $v''$  of the value function. It is remarkable that the underlying distribution of  $u$  acts on  $L$  only through the one-parameter function  $v''$ . The lemma is particularly useful in settings where the loss function  $L$  cannot

---

<sup>5</sup>We have stated the loss characterization result in Lemma 1 under the assumption that  $v''$  exists. When the distribution of the utility functions exhibits atoms, then  $v''$  has kinks and thus it is not twice differentiable. Yet, with the use of the generalized functions of Schwartz (1957), Lemma 1 extends to those cases. Assume for illustration that with probability one the agent's payoff is  $1/2$  if she correctly announces the realized state and 0 otherwise. Then  $v(p) = \max\{p - 1/2, 1/2 - p\}/2$ ,  $v''(p)$  is the Dirac delta function  $\delta(p - 1/2)$ , and  $L(p, q) = \int_p^q v''(s)(s - p)ds = |1/2 - p|$  if  $1/2$  is between  $p$  and  $q$ , and the loss is 0 otherwise.

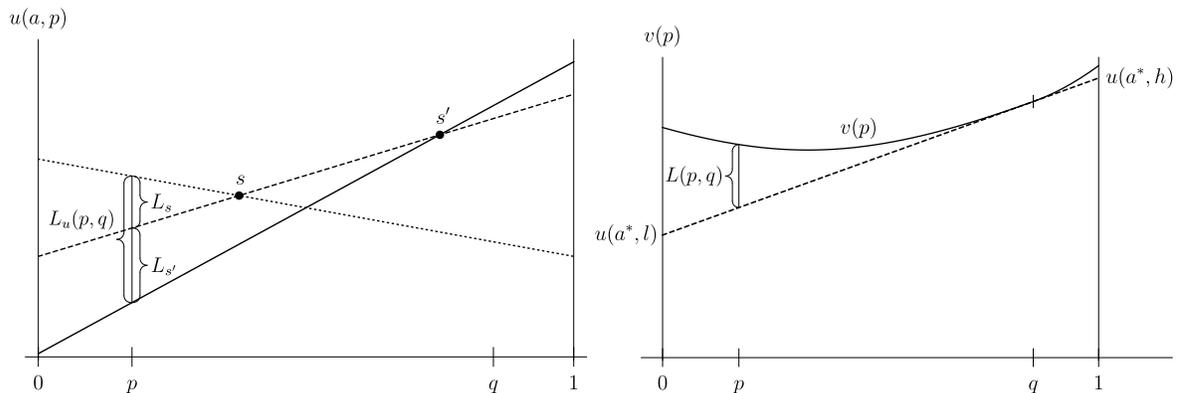


Figure 1: Left graph: Loss from misperceiving  $p$  as  $q$  for a realized utility function. Right graph: Ex ante expected loss from misperceiving  $p$  as  $q$ .

be easily computed directly, but the value function  $v$  can. This is the case for instance when the joint payoff distribution is Gaussian since then the value function admits an analytical characterization; see section 5.

In section 4, we allow the agent to control the distribution of perception errors, and show how monotonicity properties of  $v''$  translate into the direction of second-best perception biases. Intuitively, higher values of  $v''(q)$  correspond to higher marginal sensitivity of the loss  $L(p, q)$  to a marginal variation in  $q$ .

We illustrate Lemma 1 in figure 1. The left graph represents  $u(a, p)$  for all actions  $a$  under a particular realization of the payoff function  $u$ . The agent with perception  $q$  chooses the action associated with the full line, whereas the optimal choice at the objective probability  $p$  is the action associated with the dotted line, which entails the loss  $L_u(p, q) = u(a_{p,u}^*, p) - u(a_{q,u}^*, p)$  under the payoff function  $u$ . If we let  $v_u(p) = \max_a u(a, p)$ , then the loss  $L_u(p, q)$  appears as the difference at  $p$  between the function  $v_u$  and its linear approximation by its tangent at  $q$ . When averaging among all payoff functions  $u$ , the right graph depicts the expected loss as the error of the linear approximation of the expected value function.

### 3.1 The role of $v''$

The second derivative of the value function plays a central role in our analysis. Here, we offer intuitive explanations of its role in our results.

Let us start with an interpretation of  $v''(p)$ . For each realized utility function  $u$ , the

value function  $v_u(p)$  is linear in the neighborhood of  $p$  unless  $v_u(p)$  has a kink (or non-differentiability point) near  $p$ . This implies that  $v''(p)$  is determined by the stochastic occurrence of these kinks. Assume  $u$  has a density  $\rho$  over  $\mathbb{R}^{2n}$ . We define the intensity of the kink of  $v_u$  at  $p$  as the difference between the right and left derivatives  $v_u'^+(p)$  and  $v_u'^-(p)$  of  $v_u$  at  $p$ :  $k_u(p) = v_u'^+(p) - v_u'^-(p)$ . Let  $U(p, q)$  be the set of utility functions  $u$  that have a non-differentiability point between  $p$  and  $q$ . Then, for a small  $\varepsilon$ ,

$$v'(p + \varepsilon) - v'(p) \approx v''(p)\varepsilon \approx \int_{U(p, p+\varepsilon)} k_u(p_u) d\rho(u),$$

where  $p_u$  is the non-differentiability point of  $v_u(p)$  between  $p$  and  $p + \varepsilon$ .<sup>6</sup> The above expression shows that  $v''(p)$  admits a natural interpretation as the product of the expected intensity of a kink at  $p$  times the “density” of the kinks at  $p$ .

We point out a geometric representation of the loss  $L_u(p, q)$  under payoff function  $u$ . This loss is positive when the optimal choices at  $p$  and  $q$  differ, and hence when the agent is indifferent between the first-best and the second-best action at some  $s$  between  $p$  and  $q$ . Figure 1 illustrates a situation in which such indifference arises at two values  $s$  and  $s'$  between  $p$  and  $q$ . In this case, the loss  $L_u(p, q)$  is the sum of two parts,  $L_s$  and  $L_{s'}$ , where each term is identified with a tie arising at probabilities  $s$  and  $s'$ . Each of these terms equals the distance of the associated tie ( $s$  or  $s'$ ) from the true  $p$ , multiplied by the intensity of the corresponding kink in  $v_u$ . The ex ante expected loss  $L(p, q)$  is thus an integral  $\int_p^q v''(s)(s - p)ds$ , where the weight  $v''(s)$  captures the likelihood of  $u(\cdot, s)$  having a tie, compounded by the intensity of the corresponding kink of  $v_u$  at  $s$ .

### 3.2 General characterization of $v''$

We now offer a characterization of  $v''(p)$  for general payoff distributions that formalizes the intuitive discussion of  $v''(p)$  from the previous subsection. For each action  $a$ , we let  $u_a$  be the vector  $(u(a, 0), u(a, 1)) \in \mathbb{R}^2$  and recall that for  $p \in [0, 1]$ ,  $u(a, p) = pu(a, 1) + (1 - p)u(a, 0)$ . We let  $u = (u_a)_{a \in A} \in \mathbb{R}^{2n}$  denote the vector of all utility values, and for any pair of actions  $a, b$  we let  $u_{-ab} = (u_{a'})_{a' \neq a, b} \in \mathbb{R}^{2(n-2)}$ . Given a utility level  $\nu \in \mathbb{R}$ , we let  $D_{ab}(\nu) = \{u_{-ab} : \forall a' \neq a, b, u(a', p) < \nu\}$ , and introduce vectors  $w_p = (p, p - 1)$ ,  $e = (1, 1)$ . Finally,  $\|\cdot\|$  denotes the Euclidian norm in  $\mathbb{R}^2$ . We assume that  $u$  admits a density  $\rho$  and

---

<sup>6</sup>The second inline approximation relates to the possibility that more than one non-differentiability point in  $(p, p + \varepsilon)$  occurs. The probability of this event is of order  $\varepsilon^2$ .

let  $\rho_{ab}$  denote the marginal density of  $(u_a, u_b) \in \mathbb{R}^4$ .

To make sure that  $v''(p)$  is finite, we impose a bound on the density of  $u$ . We say that the density  $\rho$  satisfies the tail condition if there exists a mapping  $\xi: \mathbb{R}_+ \rightarrow [0, \bar{\xi}]$ ,  $\bar{\xi} > 0$ , such that,

1. for every  $u_a, u_b$ ,  $\rho_{ab}(u_a, u_b) \leq \xi(\|u_a\|)\xi(\|u_b\|)$ , and
2.  $\int_0^\infty x^2 \xi(x) dx$  is finite.

**Proposition 1.** *If the vector  $u$  of the utility values admits a density  $\rho$  and  $\rho$  satisfies the tail condition, then  $v$  is twice differentiable and for every  $p \in [0, 1]$ :*

$$v''(p) = \sum_{a,b:a>b} \int_{\nu,x,t \in \mathbb{R}} \int_{u_{-ab} \in D_{ab}(\nu)} (t-x)^2 \rho(u_{-ab}, \nu e + xw_p, \nu e + tw_p) du_{-ab} d\nu dx dt. \quad (2)$$

Expression (2) formalizes the intuition from the previous subsection. It integrates over actions pairs  $a, b$  and utility realizations such that the tie arises between  $a$  and  $b$  at  $p$ ,  $u(a, p) = u(b, p) = \nu$ , and all other actions  $a' \neq a, b$  are inferior at  $p$ ,  $u(a', p) < \nu$ . Each such instance contributes to  $v''(p)$  the square of the kink intensity  $(u(a, 1) - u(a, 0) - u(b, 1) + u(b, 0))^2$ , which equals  $(t-x)^2$  in the integral.<sup>7</sup>

## 4 Second-Best Perceptions

The first-best perception strategy, which minimizes losses, sets  $q = p$  for all  $p$ . We assume that perception is constrained to be noisy and study second-best perception strategies.

To model the noise, a perception strategy associates to each objective probability  $p$  a distribution over subjective perceptions  $q$ . A strategy assigns a message  $m(p)$  to each objective probability  $p \in [0, 1]$ , where we interpret the message as the physiological stimulus encoding the probability  $p$ . The message  $m$  triggers a perception  $q = m + \varepsilon$ , where  $\varepsilon$  is a random variable with atomless density  $g(\varepsilon)$  taking values in  $[-\sigma, \sigma]$  for some parameter  $\sigma \in (0, 1/2)$ . Thus, a *perception strategy* is a map  $m: [0, 1] \rightarrow [\sigma, 1 - \sigma]$ . We normalize the error term  $\varepsilon$  to have zero mean. Since the simple additive error specification is not well suited for modeling perception of probabilities near the boundaries of the probability

---

<sup>7</sup>Equation (2) integrates over the square of the kink intensity because the kink intensity contributes to  $v''(p)$  both directly and also via its impact on the density of the kinks in the neighborhood of  $p$ .

interval, we ensure that every perception  $q$  is in the  $[0, 1]$  interval by restricting  $m$  to  $[\sigma, 1 - \sigma]$ .

A particular example of a perception strategy is the unbiased perception  $m(p) = p$  for  $p \in [\sigma, 1 - \sigma]$ . This strategy does not exhibit a systematic error since  $E[q | p] = p$ . The unbiased strategy is a benchmark to which we compare the optimal perception strategy  $m^*(p)$  that solves a *naive perception problem* for each  $p$ ,<sup>8</sup>

$$m^*(p) \in \arg \min_{m \in [\sigma, 1 - \sigma]} E_\varepsilon L(p, m + \varepsilon). \quad (3)$$

Note that (3) is equivalent to the maximization of the expected payoff  $E_\varepsilon V(p, q)$ .

We summarize the two-stage decision process as follows. The agent observes the objective probability  $p$  and memorizes a probability  $m(p)$  before she learns the payoff function  $u$ . At the choice stage, the agent recalls probability  $q = m(p) + \varepsilon$  (but does not recall  $p$  and  $\varepsilon$ ), and chooses an optimal action  $a_{q,u}^*$  under her perception  $q$  and payoff  $u$ .

The next result presents a simple first-order condition. It implies that the second-best perception strategy is unbiased in a generalized sense once the perception error  $q - p$  is weighted by  $v''(q)$ . Compared to the naive unbiased perception strategy  $m(p) = p$  that minimizes the common but ad hoc mean-square-error loss criterion, the optimality condition weights errors by their expected impact on the agent's performance, where all these considerations are summarized by the weight  $v''(q)$ . In section 5, we provide four illustrative applications in which  $v''(q)$  has intuitive monotonicity properties that allow for qualitative predictions of the second-best biases.

**Proposition 2.** *Let random variable  $q = m^*(p) + \varepsilon$  be the stochastic perception when  $m^*(p)$  solves the naive perception problem (3) and the true probability is  $p$ . If  $m^*(p) \in (\sigma, 1 - \sigma)$ , then*

$$E_\varepsilon [v''(q)(q - p)] = 0. \quad (4)$$

To gain some intuition about this result, consider a message  $m \in [p - \sigma, p + \sigma]$ .<sup>9</sup> A marginal increase in  $m$  translates to a marginal increase of the perception  $q = m + \varepsilon$ . This marginal change in  $q$  affects the agent's payoff only if it affects her choice—that is, when a tie arises in between the first-best and the second-best action at  $q$ . The first-order

<sup>8</sup>A solution to the naive perception problem exists for each  $p$  since the objective is continuous in the message  $m$  and the message space is compact.

<sup>9</sup>Any message  $m > p + \sigma$  is dominated by  $m = p + \sigma$ , since for every  $\varepsilon \in [-\sigma, \sigma]$ ,  $L(p, m + \varepsilon) - L(p, p + \sigma + \varepsilon) = \int_{p+\sigma+\varepsilon}^{m+\varepsilon} v''(s)(s - p)dp \geq 0$ . Similarly any  $m < p - \sigma$  is dominated by  $m = p - \sigma$ .

condition thus minimizes the expected perception error weighted by the likelihood of a tie at  $q = m + \varepsilon$ , adjusted for the intensity of the kink, that is, weighted by  $v''(q)$ .

Note that in the very particular case where  $v$  is quadratic and  $v''$  is constant, the condition of Proposition 2 is satisfied by the unbiased perception strategy  $m(p) = p$ . In the general case, we derive the direction of the second-best perception bias from the monotonicity properties of  $v''$ . The result formalizes the intuition that the agent unable to avoid errors altogether avoids the costlier ones.

**Proposition 3.** *Suppose that  $g$  is symmetric around 0, and let  $p \in (\sigma, 1 - \sigma)$ .*

1. *If, for all  $\delta \in (0, \sigma)$ ,*

$$v''(p + \delta) < v''(p - \delta), \tag{5}$$

*then  $m^*(p) > p$ .*

2. *If, for all  $\delta \in (0, \sigma)$ ,*

$$v''(p + \delta) > v''(p - \delta),$$

*then  $m^*(p) < p$ .*

The agent considered in this section is naive in that she does not fully use the information retained up to the choice stage. A sophisticated agent could improve her performance by forming the bayesian posterior  $E[p | q]$  based on her perception  $q$ , and choose an optimal action under this posterior belief. In section 6, we study such a sophisticated agent in an example. We show there that the optimal sophisticated perception continues to be closely related to the monotonicity of  $v''$ , and that the qualitative results about the direction of the bias derived for the naive agent extend to the sophisticated one.

## 5 Applications

In this section we use our model to provide microfoundations to behavioural stylized facts. In each of our four applications, we specify simple distributional assumptions on the payoffs  $u$ , derive  $v''$  and use its monotonicity properties to find the direction of the second-best perception biases. Most, though not all, of our results hold for the Gaussian environment since it facilitates analytical derivation of the value function  $v$ .

## 5.1 Illusion of control

The term *Illusion of control*, introduced by the psychologist Langer (1975), refers to the overestimation of one's ability to impact one's own well-being. As an extreme example, casino visitors may overestimate the relevance of their choices over payoff-equivalent lotteries.

In this example, the agent is either of the low type,  $\theta = 0$ , or of the high type,  $\theta = 1$ . At the choice stage, she knows the payoffs  $u(a, \theta)$  assigned to all available actions  $a \in A$  and both types  $\theta \in \{0, 1\}$ . She is uncertain of her type, and instead of the objective probability  $p$ , she attaches the subjective probability  $q$  to being the high type.

We assume that the choice of the high type has a greater impact on her well-being than the choice of the low type in most (but not all) of the decision problems that the agent encounters in her environment. This may be the case if the high type stands for a higher talent, and actions of talented people have relatively larger consequences in typical (but not all) life situations. The assumption may also hold if the payoff consequence of each action of the low type is influenced by luck to a larger extent than the payoff of the high type, so that the low type has relatively low ability to affect her expected well-being.<sup>10</sup> We capture the relatively higher control of the high type over her well-being by assuming that the payoff function is generated by the following process:

$$u(a, \theta) = \tilde{u}_a + \tau_\theta \eta_a,$$

where  $\tilde{u}_a$  and  $\eta_a$  are independent normal random variables, iid. across actions, and  $\tau_1 > \tau_0 > 0$  are two fixed parameters. Then, the payoff differences  $u(a, 1) - u(a', 1)$  tend to be larger for the high type 1 than differences  $u(a, 0) - u(a', 0)$  for the low type 0. Accordingly, we say that the high type has *more control* than the low one. See appendix B for an extension of this application to general distributions.

The following lemma relates control in each state with the direction of monotonicity of  $v''$ .

**Lemma 2.** *If the high type has more control than the low type, then  $v''$  is decreasing.*

In this example, we are able to compute directly the analytical form of the value function  $v$ , which makes the derivation of  $v''$  simple.

---

<sup>10</sup>See examples 1-3 in the working-paper version of this paper, Gossner and Steiner (2016), for formalizations.

*Proof of Lemma 2.* Let  $\tilde{u}_a \sim \mathcal{N}(\mu, \omega^2)$  and  $\eta_a \sim \mathcal{N}(\nu, \omega'^2)$ . Then

$$u(a, p) \sim \mathcal{N}(\mu + \tau(p)\nu, \omega^2 + \tau^2(p)\omega'^2),$$

where  $\tau(p) = (1 - p)\tau_0 + p\tau_1$ . The value  $v(p)$  is the expected maximum over  $n$  iid. draws from this distribution. The expectation of the maximum of  $n$  iid. draws from  $\mathcal{N}(\tilde{\mu}, \sigma^2)$  equals:

$$\int_{-\infty}^{+\infty} x f\left(\frac{x - \tilde{\mu}}{\sigma}\right) \frac{1}{\sigma} F^{n-1}\left(\frac{x - \tilde{\mu}}{\sigma}\right) dx = \tilde{\mu} + \sigma \int_{-\infty}^{+\infty} \tilde{x} f(\tilde{x}) F^{n-1}(\tilde{x}) d\tilde{x},$$

where  $f$  and  $F$  are the standard normal density and distribution. Therefore,  $v(p) = \mu + \tau(p)\nu + c\sigma(p)$  where  $\sigma(p) = \sqrt{\omega^2 + \tau^2(p)\omega'^2}$  is the standard deviation of  $u(a, p)$  and  $c$  is the latter integral. The linear term does not affect  $v''$ , so we have that  $v''(p) = c\sigma''(p)$ . Using the equality  $\sigma(p) = \sqrt{c_1 + c_2p + c_3p^2}$  for some constants  $c_1$ ,  $c_2$ , and  $c_3$ , we obtain:

$$v''(p) = c \frac{d^2}{dp^2} \sqrt{c_1 + c_2p + c_3p^2} = c \frac{-c_2^2 + 4c_1c_3}{4} \sigma^{-3}(p).$$

Since the value function is the expectation over the upper envelopes of linear functions and thus convex, and since  $\sigma(p)$  is positive, it must be the case that  $c \frac{-c_2^2 + 4c_1c_3}{4}$  is positive. By the assumption that  $\tau_1 > \tau_0$ , the function  $\tau(p)$  and hence also  $\sigma(p)$  are increasing. Thus,  $v''(p)$  decreases in  $p$ .  $\square$

When  $v''$  is decreasing, the first point of Proposition 3 applies, and the agent systematically overestimates her control over her own well-being.

**Corollary 1.** *If the high type has more control than the low type, then the second-best perception is biased upwards:  $m^*(p) > p$  for all  $p \in (0, 1 - \sigma)$ .*

A misperception of  $p$  as  $q$  causes a loss only for draws of the payoff function  $u$  under which a tie arises in-between  $p$  and  $q$ . In this application, the variance of  $u(a, s)$  is increasing in the belief  $s$ . Larger values of this variance correspond to more spread between payoffs, hence to lower likelihoods of a tie at probability  $s$ . Therefore, upward perception errors lead to less frequent losses than downward errors of the same magnitude.<sup>11</sup>

<sup>11</sup>In the formal derivation of the optimal perception bias we account for not only the frequencies of the errors but also the magnitudes of the payoff losses.

## 5.2 Order effects

Psychologist Baron (2000) defines an *order effect* as order-dependent weighting of the observed pieces of evidence when the order of the evidence presentation is normatively uninformative. *Primacy effect* arises when the agent overweights her first impression and the *recency effect* refers to overweighting the last impression. Page and Page (2010) document the order effect in a field study of talent judgement and they argue for the relevance of the effect in hiring practices. Here we offer a stylized model in which the order effect arises as a second-best perception strategy for an agent with imperfect memory.

The agent chooses an object  $a \in \{1, 2\}$ . The quality of each object  $a$  is a sum  $x_1^a + x_2^a$  of two components,  $x_1^a$  and  $x_2^a$ . The agent observes the first components of both objects in round 1 and the second components in round 2. For instance, the agent may be a firm that chooses one job candidate out of two applicants; it screens each applicant in two tests, each test reveals one of the two applicant's qualities, and the firm's goal is to choose the applicant with the higher sum of their qualities.

We let  $\Delta_t = x_t^2 - x_t^1$  denote the quality difference in round  $t$ , and refer to  $\Delta_1$  as the first impression. The pair of second components  $(x_2^1, x_2^2)$  is drawn from a joint density  $\varrho(x_2^1, x_2^2)$ , independently of the first components. We assume that the density  $\psi(\Delta_2)$  of the difference  $\Delta_2 = x_2^2 - x_2^1$ , expressed by the formula

$$\psi(\Delta_2) = \int_{-\infty}^{\infty} \varrho(x, x + \Delta_2) dx,$$

is single-peaked and symmetric around zero. This is the case, for instance, if  $x_2^1$  and  $x_2^2$  are drawn iid. from a Gaussian distribution, but also for many other symmetric densities  $\varrho$ . The distribution of the first components plays no role in the analysis.

The agent chooses an object at the end of round 2, after she has observed all four components  $x_t^a$ . She flawlessly perceives the second-round components, but has a distorted recollection of the first impression  $\Delta_1$ . Accordingly, we distinguish between the objective value  $\Delta_1$  and the perceived value  $\hat{\Delta}_1$ . The agent chooses object 2 if  $\hat{\Delta}_1 + \Delta_2 > 0$ , and object 1 otherwise.

Let us consider the loss from misperceiving  $\Delta_1$  as  $\hat{\Delta}_1$ , as a function of  $\Delta_2$ . If both  $\Delta_1 + \Delta_2$  and  $\hat{\Delta}_1 + \Delta_2$  have a same sign, a same object is chosen in both cases and the loss is null. If  $\Delta_1 + \Delta_2$  and  $\hat{\Delta}_1 + \Delta_2$  have opposite signs, the wrong object is chosen under  $\hat{\Delta}_1$ , and the corresponding loss is  $|\Delta_1 + \Delta_2|$ . The expected loss of perceiving  $\Delta_1$  as  $\hat{\Delta}_1$  is

thus independent of the particular realizations of the first components  $(x_1^1, x_1^2)$  such that  $\Delta_1 = x_1^2 - x_1^1$ , and can be expressed as:

$$\tilde{L}(\Delta_1, \hat{\Delta}_1) = \mathbb{E}_{\Delta_2} \left[ (\Delta_1 + \Delta_2) \left( 1_{\Delta_1 + \Delta_2 > 0} - 1_{\hat{\Delta}_1 + \Delta_2 > 0} \right) \right].$$

The next lemma characterizes this loss through an integral formula that is similar to the characterization in Lemma 1.

**Lemma 3.** *The expected loss from misperceiving the first impression  $\Delta_1$  as  $\hat{\Delta}_1$  is:*

$$\tilde{L}(\Delta_1, \hat{\Delta}_1) = \int_{\Delta_1}^{\hat{\Delta}_1} \psi(s) (s - \Delta_1) ds. \quad (6)$$

The proof in the appendix consists of mapping the setting of this example with continuous payoff state  $\Delta_1$  to our main setting with binary payoff state  $\theta$ . We achieve this by a simple rescaling of the problem. The substantial part of the proof consists of showing that the relevant weight  $v''$  is, after rescaling, equal to the density  $\psi$ .

We provide a simple heuristic for the loss expression in (6). A misperception of  $\Delta_1$  for  $\hat{\Delta}_1$  leads to a loss if it reverts choice, which happens if for some  $\tilde{\Delta}_1$  in between  $\Delta_1$  and  $\hat{\Delta}_1$ , the agent is indifferent among the two objects. Given  $\tilde{\Delta}_1$ , such an indifference occurs when  $\Delta_2 = -\tilde{\Delta}_1$ , hence with likelihood  $\psi(-\tilde{\Delta}_1) = \psi(\tilde{\Delta}_1)$ . The loss arising from a mistake is  $|\Delta_1 + \Delta_2| = |\Delta_1 - \tilde{\Delta}_1|$ . Hence, the expected loss from the misperception is given by the integral in (6).

Equipped with the loss formula of Proposition 3, we now study second-best noisy perception. We assume that the subjective perception  $\hat{\Delta}_1$  is formed essentially through the same process as in our main model: The agent observes the true first impression  $\Delta_1$  in round 1, memorizes a value  $\tilde{m}(\Delta_1)$ , and recalls  $\hat{\Delta}_1 = \tilde{m}(\Delta_1) + \varepsilon$ , where  $\varepsilon$  is drawn from a density  $g(\varepsilon)$  that is symmetric around 0 and has support on  $[-\sigma, \sigma]$ . Let  $\tilde{m}^*(\Delta_1)$  be the optimal perception strategy that minimizes the expected loss, i.e.:

$$\tilde{m}^*(\Delta_1) \in \arg \min_m \mathbb{E}_\varepsilon \tilde{L}(\Delta_1, m + \varepsilon).$$

Our next result shows that the second-best perception avoids the relatively costly types of error by a systematic exaggeration of the first impression.

**Proposition 4.** *The second-best perception exaggerates the first impression:  $\tilde{m}^*(\Delta_1) > \Delta_1$  if  $\Delta_1 > 0$ , and  $\tilde{m}^*(\Delta_1) < \Delta_1$  if  $\Delta_1 < 0$ .*

The proof in the appendix consists of applying Proposition 3 to the rescaled problem. The direction of the bias is intuitive. Ties between the two objects arise relatively often, across all  $\Delta_2$ , when the first impression  $\Delta_1$  does not strongly favour any of the two objects. Thus, an underestimation of the absolute value of the first impression is more likely to lead to the suboptimal choice than the overestimation of the same size. The primacy effect arising in Proposition 4 is the consequence of our assumption that the agent recalls the early evidence imperfectly and the late evidence errorlessly. Under the opposite assumption, a recency effect would arise.

### 5.3 Overprecision

Moore and Healy (2008) define *overprecision* as an overvaluation of the precision of one's own belief. Daniel et al. (1998) is an early study of the implications of overprecision to financial markets. In our example, the agent bets on a binary event. Each bet  $a \in \{1, \dots, n\}$  pays a reward  $u(a, \theta)$  in state  $\theta \in \{0, 1\}$ . The agent observes the  $2n$  rewards  $(u(a, \theta))_{a, \theta}$  and chooses a bet  $a^* \in \{1, \dots, n\}$  that maximizes the perceived expected payoff  $u(a, q)$  under the subjective probability  $q$  of the state being  $\theta = 1$ . All rewards  $u(a, \theta)$  are independently drawn from the standard normal distribution.

The further away  $q$  is from  $1/2$ , the more confident the agent is of her ability to predict the correct state. We show that the second-best perception systematically exaggerates this ability. To this end, we first establish monotonicity property of  $v''$ .

**Lemma 4.**  *$v''$  is single-peaked and symmetric around  $1/2$ .*

The proof relies on an explicit computation of  $v$  and  $v''$ , which the Gaussian framework makes possible.

*Proof of Lemma 4.* The expected payoff  $u(a, p)$  for each action  $a$  at probability  $p$  is an iid. draw from  $\mathcal{N}\left(0, \sqrt{p^2 + (1-p)^2}\right)$ , and  $v(p)$  is the expected maximum over these draws. As in the proof of Lemma 2,  $v(p)$  is proportional to the standard deviation of each draw:

$$v(p) = c\sqrt{p^2 + (1-p)^2},$$

for some positive constant  $c$  that depends on  $n$  only. The second derivative is

$$v''(p) = c\frac{1}{\sqrt{p^2 + (1-p)^2}^3},$$

which has the properties stated in the lemma.  $\square$

Proposition 3 together with Lemma 4 imply the direction of the bias.

**Corollary 2.** *The second-best perception exaggerates the agent’s ability to predict the state:  $m^*(p) < p$  for  $p \in (\sigma, 1/2)$ , and  $m^*(p) > p$  for  $p \in (1/2, 1 - \sigma)$ .*

Since the variance of the expected payoff  $u(a, p)$  for each  $a$  increases with the distance of  $p$  from  $1/2$ , ties between actions arise more often at beliefs close to  $1/2$  than at beliefs further away from  $1/2$ . Therefore, an error in the perception of  $p$  in the direction towards  $1/2$ —the undervaluation of one’s ability to predict the state—is costlier than an opposite error of the same size. The second-best perception avoids the costlier errors by systematically overvaluing one’s information about the state.

## 5.4 Overweighting of small probabilities

We revisit here Steiner and Stewart (2016), who provide a microfoundation for overweighting of small probabilities akin to the one in prospect theory (Kahneman and Tversky, 1979). The purpose of the revisit is twofold. First, we show how the problem from Steiner and Stewart can be solved by the general method from this paper. Second, we contrast their model with the setting from the previous subsection in which underestimation of small probabilities arises, and clarify the forces driving the opposite biases in these two frameworks.

As in Steiner and Stewart (2016), we consider an agent who chooses  $a \in \{1, 2\}$ , where the payoff for action 1 is a lottery that pays  $u(1, 1)$  with probability  $p$ , and  $u(1, 0)$  with probability  $1 - p$ . Action 2 is an outside option that pays a fixed payoff  $u(2, \theta) = \tilde{u} \in \mathbb{R}$ . The agent misperceives the probability value  $p$  of state 1 for  $q$  and chooses the lottery if and only if  $qu(1, 1) + (1 - q)u(1, 0) > \tilde{u}$ . The lottery rewards  $u(1, 1)$  and  $u(1, 0)$  are iid. draws from the standard normal distribution.<sup>12</sup> Steiner and Stewart interpret the outside option value  $\tilde{u}$  as a maximum over many alternatives, and thus focus on values of  $\tilde{u}$  larger than typical draws of the lottery rewards.

Steiner and Stewart solved this problem by explicitly computing the loss  $L(p, q)$  and by optimizing over the messages directly. Here, we apply the methodology of this paper. We compute  $v(p)$ , characterize monotonicity properties of its second derivative, and use Proposition 3 to derive the direction of bias.

---

<sup>12</sup>See Steiner and Stewart (2016) for an extension to general distributions.

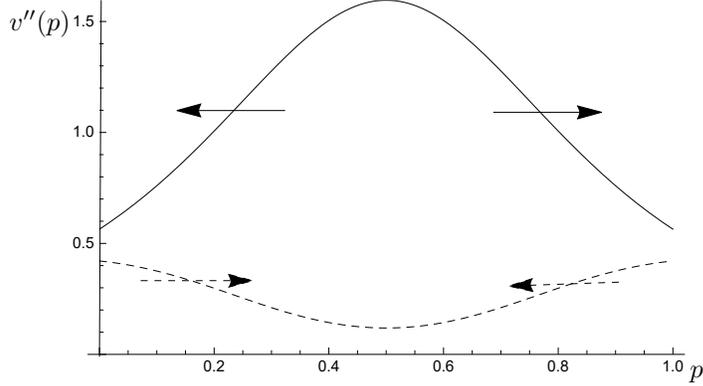


Figure 2:  $v''$  and the direction of the bias. Solid line: the overprecision setting. Dashed line: the setting of this section.

Our first step is to establish that, for values of the outside option  $\tilde{u}$  larger than a certain threshold, the function  $v''$  has the opposite monotonicity property than that in the previous application.

**Lemma 5.**  *$v''$  is symmetric around  $1/2$ . If  $|\tilde{u}| > 3^{1/4}$ , then  $v''$  is U-shaped.*

The proof, shown in the appendix, relies on a computation of  $v''$  from an explicit derivation of  $v$ .

Proposition 3 implies that overvaluation of small probabilities arises in this setting.

**Corollary 3.** *If  $|\tilde{u}| > 3^{1/4}$ , then the second-best perception exaggerates small probabilities and undervalues large probabilities:  $m^*(p) > p$  for  $p \in (0, 1/2)$ , and,  $m^*(p) < p$  for  $p \in (1/2, 1)$ .*

Why does the agent from the previous application undervalue small probabilities, whereas the agent of this application overvalues them? The difference is caused by the distinct patterns of the ties. Figure 2 depicts  $v''$  for both settings. In the overprecision setting, a tie between lotteries is likely when the lottery values have a *low* variance, which happens when  $p$  is close to half. This contrasts with the prospect-theory setting, where a tie between the lottery and the high outside option  $\tilde{u}$  is likely when the lottery value has a *high* variance, and hence  $v''$  is U-shaped. Consequently, the perceptions exhibit the opposite biases. The two results suggest that the bias in the perception of probabilities depends on whether the perceived probability affects payoffs to all actions, or idiosyncratically affects the payoff only to one action.

## 6 A Sophisticated Agent—an Example

The agent from section 4 is naive in that she does not fully utilize the information available to her at the choice stage. This section studies the sophisticated perception strategies of an agent who loses some information during the decision process, but who is then fully capable of utilizing the information retained until the choice.

The agent assigns prior probability  $1/2$  to the state  $\theta = 1$ . At the perception stage, she receives an informative signal and updates the probability of  $\theta = 1$  according to the Bayes rule to value  $p$ . The agent is able to retain only some of this information until the choice stage. In particular, she commits to a *perception function*  $m(p)$ ,  $m : [0, 1] \rightarrow M$ , where  $M = \{0, 1\}$  is a binary message space before she observes the payoff function  $u$ . At the choice stage, the agent chooses an action while informed of the payoffs  $u = (u(a, \theta))_{a, \theta}$  and remembering the message  $m(p)$ . Her behaviour at this stage is therefore described by a *choice rule*  $c : \mathbb{R}^{2n} \times M \rightarrow A$ . The *sophisticated perception problem* is:

$$\max_{m(\cdot), c(\cdot, \cdot)} \mathbb{E}_{u, p} u(c(u, m(p)), p).$$

The next lemma states that the optimal perception function  $m(\cdot)$  generates a binary partition  $\mathcal{I}(p^*) = \{[0, p^*), [p^*, 1]\}$  of the probability interval. The agent at the choice stage remembers that  $p$  belongs to  $I \in \mathcal{I}(p^*)$ , forms a Bayesian posterior  $q = \mathbb{E}[p \mid p \in I]$ , and chooses the optimal action  $a_{q, u}^*$  under the posterior belief  $q$ .<sup>13</sup>

**Lemma 6.** *There exists  $p^*$  such that  $m(p)$  is an indicator function of one of the two intervals of the partition  $\mathcal{I}(p^*)$  for almost all  $p$ ; and  $c(u, m) = a_{q(m), u}^*$  for almost all  $u$ , where  $q(m)$  is the posterior belief  $\mathbb{E}[p \mid m(p) = m]$ .*

Let us denote the density of the belief  $p$  by  $\varphi$  and assume that it has support on  $[0, 1]$  and is symmetric around half:  $\varphi(p) = \varphi(1 - p)$  for all  $p \in [0, 1]$ . We interpret the symmetry as a neutrality assumption abstracting from biases driven by an asymmetry of  $\varphi$ .

Denote the two Bayesian posteriors by  $\underline{q}(p^*) = \mathbb{E}[p \mid p < p^*]$ , and  $\bar{q}(p^*) = \mathbb{E}[p \mid p \geq p^*]$ .

---

<sup>13</sup>The partition of the probability interval is reminiscent of the coarse reasoning model in Jehiel (2005). Other models that assume biased beliefs in strategic interactions include Eyster and Rabin (2005) with their application to the winner's curse bias, and Pavan (2014) who derives a compromise effect in a model with bounded recall. One distinction is that our agent optimizes over the information loss in a range of nonstrategic decision problems.

**Proposition 5** (sophisticated perception). *If*

$$v''(p) > v''(1-p) \text{ for all } p < 1/2, \tag{7}$$

*and  $\varphi(p)$  is symmetric around  $1/2$ , then the optimal threshold probability  $p^{**}$  is less than  $1/2$  and the agent attains the high posterior  $\bar{q}(p^{**})$  with a higher probability than the low posterior  $\underline{q}(p^{**})$ .*

The symmetric result holds when  $v''(p) < v''(1-p)$  for all  $p < 1/2$ ; in this case  $p^{**} > 1/2$ .

To obtain some intuition for the last result, observe that  $v''(p)$  can be interpreted as a local value of information. To see this, consider an agent with a belief  $p$  who receives a weakly informative signal upon which she updates her belief to  $\pi$  in a neighborhood of  $p$ . The Bayesian constraint  $E\pi = p$  and the second-order Taylor expansion of  $v$  imply that the value of the received information is  $E v(\pi) - v(p) \approx v''(p)Var(\pi)/2$ . By assumption (7), the local value of information is higher in the lower half of the probability interval than in the upper one. The optimal binary partition then discriminates between the probabilities in the lower half of the probability interval at the expense of not distinguishing among the probabilities in the upper half.

Since the studied agent is Bayesian, the expected terminal belief,  $E q$ , is necessarily equal to the prior belief  $1/2$ . Bayesian rationality thus implies that the relatively frequent occurrence of the high terminal belief is compensated by adjustments of the values of the two terminal beliefs. Proposition 5 has implications to perception in the following sense. Assume, for instance, that the distribution of payoffs is as specified in subsection 5.1. Let half of the population have the high control over their own well-being, and the other half the low control. Each agent receives incomplete information about her own type. Further assume that the agents are unable to perceive a continuum of probabilities, and can only distinguish a “high” and “low” probability of being the high type. Since  $v''$  is decreasing in this case, we predict that the agents who perceive themselves as “highly” likely to be of the high type are more common than those who evaluate the chance of being the high type as “low”.

## 7 Discussion

We have studied the cost of misperception in a model of decision making under uncertainty. Our main technical result, Lemma 1, shows that the loss function can be entirely characterized from the value function of the underlying decision problem. In our applications, the decision problem is specified by a random payoff function, which makes our environment rich enough. Note however that our loss formula applies to all settings, whether or not the payoff function is random, as long as the value function admits a second derivative. We presented a series of applications of this loss formula to settings in which the agent has imperfect memory, and hence faces an error-management problem; these examples allowed us to microfound several well-known behavioural biases.

Microfoundations for behavioural biases can contribute to normative discussions of debiasing. Our model suggests that a bias relative to the precision-maximizing perception may be second-best optimal and thus presence of a bias is not enough to justify an intervention without further arguments. As the standard maladaptation argument goes, an intervention into the decision process may be beneficial if the agent’s environment has changed since the biases have evolved. Interestingly, the model suggests that an intervention may be justified even if the agent’s environment has not changed since adaptation took place. This is the case if an intervening outsider knows more about the agent’s decision problem than the perception designer—evolution—has known.

## References

- Alaoui, L. and A. Penta (2016). Cost-benefit analysis in reasoning. mimeo.
- Baron, J. (2000). *Thinking and deciding*. Cambridge University Press.
- Brunnermeier, M. K. and J. A. Parker (2005). Optimal expectations. *American Economic Review* 95(3), 1092–1118.
- Caplin, A. and J. Leahy (2001). Psychological expected utility theory and anticipatory feelings. *The Quarterly Journal of Economics* 116(1), 55–79.
- Compte, O. and A. Postlewaite (2012). Belief formation. PIER Working Paper 12-027, University of Pennsylvania.

- Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998). Investor psychology and security market under- and overreactions. *The Journal of Finance* 53(6), 1839–1885.
- De Lara, M. and O. Gossner (2017). An instrumental approach to the value of information. Technical report.
- Dow, J. (1991). Search decisions with limited memory. *The Review of Economic Studies* 58(1), 1–14.
- Eyster, E. and M. Rabin (2005). Cursed equilibrium. *Econometrica* 73(5), 1623–1672.
- Gigerenzer, G. and P. Todd (1999). *Simple heuristics that make us smart*. Oxford University Press.
- Gossner, O. and J. Steiner (2016). Optimal illusion of control and related perception biases. CEPR Discussion Paper 11478.
- Herold, F. and N. Netzer (2010). Probability weighting as evolutionary second-best. University of Zurich Socioeconomic Institute Working Paper 1005.
- Hirshleifer, D. and I. Welch (2002). An economic approach to the psychology of change: Amnesia, inertia, and impulsiveness. *Journal of Economics & Management Strategy* 11(3), 379–421.
- Jehiel, P. (2005). Analogy-based expectation equilibrium. *Journal of Economic Theory* (123), 81–104.
- Johnson, D. D., D. T. Blumstein, J. H. Fowler, and M. G. Haselton (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology & Evolution* 28(8), 474–481.
- Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica* 47(2), 263–291.
- Kőszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association* 4(4), 673–707.
- Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology* 32(2), 311–328.

- Lipman, B. L. (1995). Information processing and bounded rationality: A survey. *Canadian Journal of Economics*, 42–67.
- Milgrom, P. R. and R. J. Weber (1982). A theory of auctions and competitive bidding. *Econometrica* 50(5), 1089–1122.
- Moore, D. A. and P. J. Healy (2008). The trouble with overconfidence. *Psychological Review* 115(2), 502–517.
- Mullainathan, S. (2002). A memory-based model of bounded rationality. *The Quarterly Journal of Economics* 117(3), 735–774.
- Netzer, N. (2009). Evolution of time preferences and attitudes toward risk. *American Economic Review* 99(3), 937–955.
- Page, L. and K. Page (2010). Last shall be first: A field study of biases in sequential performance evaluation on the idol series. *Journal of Economic Behavior & Organization* 73(2), 186–198.
- Pavan, A. (2014). Attention, coordination and bounded recall. Technical report, Center for Mathematical Studies in Economics and Management Science.
- Rayo, L. and G. S. Becker (2007). Evolutionary efficiency and happiness. *Journal of Political Economy* 115(2), 302–337.
- Robson, A. J. (2001). Why would nature give individuals utility functions? *Journal of Political Economy* 109(4), 900–914.
- Robson, A. J. and L. Samuelson (2010). The evolutionary foundations of preferences. In *Handbook of Social Economics*, Volume 1, pp. 221–310. Elsevier, Amsterdam: North Holland.
- Schwartz, L. (1957). *Théorie des distributions*. Hermann Paris.
- Shannon, C. (1959). Coding theorems for a discrete source with a fidelity criterion. *Institute of Radio Engineers National Convention Record part 4*, 142–163.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal* 27, 3–55.

- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics* 69(1), 99–118.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review* 63(2), 129.
- Sims, C. A. (1998). Stickiness. In *Carnegie-Rochester Conference Series on Public Policy*, Volume 49, pp. 317–356. Elsevier.
- Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics* 50(3), 665–690.
- Steiner, J. and C. Stewart (2016). Perceiving prospects properly. *American Economic Review* 106(7), 1601–31.
- Taylor, S. E. and J. D. Brown (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin* 103(2), 193.
- Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* 13(4), 600–612.
- Wilson, A. (2014). Bounded memory and biases in information processing. *Econometrica* 82(6), 2257–2294.
- Woodford, M. (2012a). Inattentive valuation and reference-dependent choice. Columbia University, New York.
- Woodford, M. (2012b). Prospect theory as efficient perceptual distortion. *American Economic Review* 102(3), 41–46.

## A Proofs

*Proof of Lemma 1.* Note that  $v(p) = \max_q V(p, q) = V(p, p)$ . Applying the Envelope theorem to

$$V(p, q) = \mathbb{E}_u [(1-p)u(a_{q,u}^*, 0) + pu(a_{q,u}^*, 1)]$$

gives

$$v'(p) = \left. \frac{\partial}{\partial p} V(p, q) \right|_{q=p} = \mathbb{E}_u [u(a_{p,u}^*, 1) - u(a_{p,u}^*, 0)]. \quad (8)$$

Using the linearity of the expected payoff with respect to probability and (8), we obtain

$$V(p, q) = v(q) - v'(q)(q - p),$$

and thus,

$$L(p, q) = v(p) - V(p, q) = v(p) - v(q) + v'(q)(q - p)$$

is the error of the linear approximation of the value function  $v(p)$  when it is approximated at  $q$ . Expression (1) follows once we integrate by parts:

$$\int_p^q v''(s)(s-p)ds = [v'(s)(s-p)]_p^q - \int_p^q v'(s)ds = L(p, q).$$

□

The next result is an auxiliary lemma used in the proof of Proposition 1. For any action  $a$ , we let  $u_{-a} = (u_{a'})_{a' \neq a} \in \mathbb{R}^{2(n-1)}$ . Given  $a$  and  $u_a$ , we let the measure  $Q_{u_a}$  on  $\mathbb{R}^{2(n-1)}$  be given by  $Q_{u_a}(S) = \int_S \rho(u_{-a}, u_a) du_{-a}$  for every borel set  $S \subseteq \mathbb{R}^{2(n-1)}$ , and let  $U_a^p = \{u_{-a} : \forall b \neq a, u(b, p) < u(a, p)\}$ .

**Lemma 7.**

$$\frac{d}{dp} Q_{u_a}(U_a^p) = \sum_{b \neq a} \int_{u_{-ab} \in D_{ab}(u(a, p))} \int_{s \in \mathbb{R}} s \rho(u_{-ab}, u_a + sw_p, u_a) ds du_{-ab}.$$

*Proof of Lemma 7.* Fix action  $a$  and  $u_a$ . Let the sets  $B_b^p = \{u_{-a} : \forall a' \neq a, b, u(a', p) < u(b, p)\}$ . Given  $p, p_0 \in [0, 1]$ , since the sets  $(B_b^{p_0})_{b \neq a}$  partition  $\mathbb{R}^{2(n-1)}$  up to a set of  $Q_{u_a}$ -measure 0,

we have:

$$Q_{u_a}(U_a^p) = \sum_{b \neq a} Q_{u_a}(U_a^p \cap B_b^{p_0}).$$

Hence,

$$\left. \frac{d}{dp} \right|_{p=p_0} Q_{u_a}(U_a^p) = \sum_{b \neq a} \left. \frac{d}{dp} \right|_{p=p_0} Q_{u_a}(U_a^p \cap B_b^{p_0}).$$

In order to evaluate each term of the sum we first write:

$$Q_{u_a}(U_a^p \cap B_b^{p_0}) = \int_{u_{-ab} \in \mathbb{R}^{2(n-2)}} \int_{u_b: u(a,p) > u(b,p)} 1_{B_b^{p_0}} \rho(u_{-ab}, u_b, u_a) du_b du_{-ab}.$$

We make a change of variables:  $u_b = u_a - r(\cos \theta, \sin \theta)$ , and observe that  $u(a, p) > u(b, p)$  is equivalent to  $\theta_p < \theta < \theta_p + \pi$  where  $\theta_p = \arctan\left(-\frac{1-p}{p}\right)$ . This leads to:

$$Q_{u_a}(U_a^p \cap B_b^{p_0}) = \int_{u_{-ab}} \int_{r>0} \int_{\theta=\theta_p}^{\theta_p+\pi} 1_{B_b^{p_0}} r \rho(u_{-ab}, u_a - r(\cos \theta, \sin \theta), u_a) d\theta dr du_{-ab}.$$

We differentiate this expression with respect to  $p$  at  $p = p_0$ , and use the fact that on the boundary, for  $\theta \in \{\theta_{p_0}, \theta_{p_0} + \pi\}$ , one has  $u(a, p_0) = u(b, p_0)$ , and thus  $1_{B_b^{p_0}} = 1_{D_{ab}(u(a, p_0))}$ :

$$\begin{aligned} \left. \frac{d}{dp} \right|_{p=p_0} Q_{u_a}(U_a^p \cap B_b^{p_0}) &= \int_{u_{-ab}} 1_{D_{ab}(u(a, p_0))} \left( \int_{r>0} r \rho(u_{-ab}, u_a + r(\cos \theta_{p_0}, \sin \theta_{p_0}), u_a) dr \right. \\ &\quad \left. - \int_{r>0} r \rho(u_{-ab}, u_a - r(\cos \theta_{p_0}, \sin \theta_{p_0}), u_a) dr \right) du_{-ab} \left. \frac{d}{dp} \right|_{p=p_0} \theta_p \\ &= \int_{u_{-ab} \in D_{ab}(u(a, p_0))} \int_{r \in \mathbb{R}} r \rho(u_{-ab}, u_a + r(\cos \theta_{p_0}, \sin \theta_{p_0}), u_a) dr du_{-ab} \\ &\quad \times \left. \frac{d}{dp} \right|_{p=p_0} \theta_p. \end{aligned}$$

Using  $\frac{d}{dp}\Big|_{p=p_0} \theta_{p_0} = \frac{1}{p_0^2 + (1-p_0)^2}$ , noting that  $(\cos \theta_p, \sin \theta_p) = \frac{w_p}{\sqrt{p^2 + (1-p)^2}}$ , and introducing  $\sigma_p = \sqrt{p^2 + (1-p)^2}$  we obtain:

$$\frac{d}{dp}\Big|_{p=p_0} Q_{u_a} (U_a^p \cap B_b^{p_0}) = \int_{u_{-ab} \in D_{ab}(u(a, p_0))} \int_{r \in \mathbb{R}} \frac{r}{\sigma_p^2} \rho \left( u_{-ab}, u_a + \frac{r w_p}{\sigma_p}, u_a \right) dr du_{-ab}.$$

Finally, we use a last change of variable,  $s = r/\sigma_p$  to derive:

$$\frac{d}{dp}\Big|_{p=p_0} Q_{u_a} (U_a^p \cap B_b^{p_0}) = \int_{u_{-ab} \in D_{ab}(u(a, p_0))} \int_{s \in \mathbb{R}} s \rho (u_{-ab}, u_a + s w_p, u_a) ds du_{-ab}.$$

Summation over  $b \neq a$  concludes the proof of the lemma.  $\square$

*Proof of Proposition 1.* Recall that

$$v(p) = \sum_{a=1}^n \mathbb{E} [u(a, p) 1_{\{\forall b \neq a, u(a, p) > u(b, p)\}}].$$

Since the agent optimizes over actions, we can apply the Envelope theorem to get (see De Lara and Gossner (2017) for related work)

$$\begin{aligned} v'(p) &= \sum_a \mathbb{E} \left[ \left( \frac{d}{dp} u(a, p) \right) 1_{\{\forall b \neq a, u(a, p) > u(b, p)\}} \right] \\ &= \sum_a \mathbb{E} [(u(a, 1) - u(a, 0)) 1_{\{\forall b \neq a, u(a, p) > u(b, p)\}}] \\ &= \sum_a \int_{u_a \in \mathbb{R}^2} (u(a, 1) - u(a, 0)) Q_{u_a} (U_a^p) du_a. \end{aligned}$$

By applying Lemma 7 we obtain:

$$v''(p) = \sum_{a, b: a \neq b} \int_{u_a} \int_{u_{-ab} \in D_{ab}(u(a, p))} \int_s (u(a, 1) - u(a, 0)) s \rho (u_{-ab}, u_a + s w_p, u_a) ds du_{-ab} du_a.$$

A change of variables,  $(\nu, t) = (u(a, p), u(a, 0) - u(a, 1))$  gives:

$$v''(p) = \sum_{a, b: a \neq b} \int_{\nu, s, t \in \mathbb{R}} \int_{u_{-ab} \in D_{ab}(\nu)} -st \rho (u_{-ab}, \nu e + (s+t) w_p, \nu e + t w_p) du_{-ab} d\nu ds dt.$$

With a new change of variable  $x = s + t$  we obtain:

$$v''(p) = \sum_{a,b:a \neq b} \int_{\nu,x,t \in \mathbb{R}} \int_{u_{-ab} \in D_{ab}(\nu)} (t-x)t\rho(u_{-ab}, \nu e + xw_p, \nu e + tw_p) du_{-ab} d\nu dx dt.$$

We group the terms  $a, b$  and  $b, a$  in the sum to get:

$$v''(p) = \sum_{a,b:a > b} \int_{\nu,x,t \in \mathbb{R}} \int_{u_{-ab} \in D_{ab}(\nu)} (t-x)^2 \rho(u_{-ab}, \nu e + xw_p, \nu e + tw_p) du_{-ab} d\nu dx dt.$$

It remains to prove that  $v''(p)$  is finite. For  $a > b$ , let  $v''_{ab}(p)$  be the term:

$$\begin{aligned} v''_{ab}(p) &= \int_{\nu,x,t \in \mathbb{R}} \int_{u_{-ab} \in D_{ab}(\nu)} (t-x)^2 \rho(u_{-ab}, \nu e + xw_p, \nu e + tw_p) du_{-ab} d\nu dx dt \\ &\leq \int_{\nu,x,t \in \mathbb{R}} \int_{u_{-ab} \in \mathbb{R}^{2(n-2)}} (t-x)^2 \rho(u_{-ab}, \nu e + xw_p, \nu e + tw_p) du_{-ab} d\nu dx dt \\ &= \int_{\nu,x,t \in \mathbb{R}} (t-x)^2 \rho_{ab}(\nu e + xw_p, \nu e + tw_p) d\nu dx dt. \end{aligned}$$

With  $w_p^\perp = ((1-p), p)$  and  $\sigma_p = \sqrt{p^2 + (1-p)^2}$  we rewrite this as:

$$\begin{aligned} v''_{ab}(p) &\leq \int_{\nu,x,t \in \mathbb{R}} (t-x)^2 \rho_{ab}(\nu(w_p + w_p^\perp) + xw_p, \nu(w_p + w_p^\perp) + tw_p) d\nu dx dt \\ &= \int_{\nu,x,t \in \mathbb{R}} (t-x)^2 \rho_{ab}(\nu w_p^\perp + xw_p, \nu w_p^\perp + tw_p) d\nu dx dt \\ &\leq \int_{\nu,x,t \in \mathbb{R}} (t-x)^2 \xi(\sigma_p(|\nu| + |x|)) \xi(\sigma_p(|\nu| + |t|)) d\nu dx dt \\ &\leq \int_{\nu,x,t \geq 0} 8(t+x)^2 \xi(\sigma_p(\nu+x)) \xi(\sigma_p(\nu+t)) d\nu dx dt \\ &\leq \int_{\nu,x,t \geq 0} 8((t+\nu) + (x+\nu))^2 \xi(\sigma_p(\nu+x)) \xi(\sigma_p(\nu+t)) d\nu dx dt. \end{aligned}$$

With a change of variables,  $X = \sigma_p(x + \nu)$ ,  $Y = \sigma_p(t + \nu)$ , this gives:

$$\begin{aligned}
v''_{ab}(p) &\leq \int_{\nu \geq 0} 8\sigma_p^{-4} \int_{X, Y \geq \nu} (X + Y)^2 \xi(X) \xi(Y) dX dY d\nu \\
&= 8\sigma_p^{-4} \int_{\nu \geq 0} \left( \int_{X \geq \nu} X^2 \xi(X) dX \int_{Y \geq \nu} \xi(Y) dY \right) d\nu \\
&\quad + 8\sigma_p^{-4} \int_{\nu \geq 0} \left( \int_{Y \geq \nu} Y^2 \xi(Y) dY \int_{X \geq \nu} \xi(X) dX \right) d\nu \\
&\quad + 16\sigma_p^{-4} \int_{\nu \geq 0} \left( \int_{X \geq \nu} X \xi(X) dX \int_{Y \geq \nu} Y \xi(Y) dY \right) d\nu \\
&\leq 16\sigma_p^{-4} \int_{X \geq 0} X^2 \xi(X) dX \int_{\nu \geq 0} \int_{Y \geq \nu} \xi(Y) dY d\nu \\
&\quad + 16\sigma_p^{-4} \int_{X \geq 0} X \xi(X) dX \int_{\nu \geq 0} \int_{Y \geq \nu} Y \xi(Y) dY d\nu \\
&= 16\sigma_p^{-4} \int_{X \geq 0} X^2 \xi(X) dX \int_{Y \geq 0} Y \xi(Y) dY \\
&\quad + 16\sigma_p^{-4} \int_{X \geq 0} X \xi(X) dX \int_{Y \geq 0} Y^2 \xi(Y) dY, \\
&= 32\sigma_p^{-4} \int_{X \geq 0} X^2 \xi(X) dX \int_{X \geq 0} X \xi(X) dX,
\end{aligned}$$

which is finite under the tail condition since  $X\xi(X)$  is bounded by  $\bar{\xi}$  on  $(0, 1)$ , it is bounded by  $X^2\xi(X)$  for  $X > 1$ , and  $\int_{X \geq 0} X^2\xi(X) dX$  is finite by assumption.  $\square$

*Proof of Proposition 2.* By Lemma 1,

$$E_\varepsilon L(p, m + \varepsilon) = E_\varepsilon \int_p^{m+\varepsilon} v''(s)(s - p) ds.$$

Since

$$\frac{d}{dm} E_\varepsilon \int_p^{m+\varepsilon} v''(s)(s - p) ds = E_\varepsilon [v''(m + \varepsilon)(m + \varepsilon - p)],$$

the first-order condition for the interior optimal  $m$  gives (4).  $\square$

*Proof of Proposition 3.* We prove point 1, point 2 being symmetric to it. For  $p \in (\sigma, 1 - \sigma)$ , we recall from footnote 9 that any message  $m < p - \sigma$  is dominated by  $p - \sigma$ . We now show that any message of the form  $m = p - \eta$  with  $\eta \in (0, \sigma)$  is dominated by  $p + \eta$ . Thus

we want to show that:

$$\mathbb{E}_\varepsilon L(p, p - \eta + \varepsilon) > \mathbb{E}_\varepsilon L(p, p + \eta + \varepsilon).$$

Given the symmetry assumption on the distribution  $g$  of  $\varepsilon$ , this is equivalent to:

$$\mathbb{E}_\varepsilon [L(p, p - \eta + \varepsilon) + L(p, p - \eta - \varepsilon)] > \mathbb{E}_\varepsilon [L(p, p + \eta + \varepsilon) + L(p, p + \eta - \varepsilon)].$$

It is sufficient to prove that, for every  $\varepsilon \in (0, \sigma)$ ,

$$L(p, p - \eta + \varepsilon) + L(p, p - \eta - \varepsilon) > L(p, p + \eta + \varepsilon) + L(p, p + \eta - \varepsilon).$$

Let us first examine the case in which  $\varepsilon < \eta$ . By applications of Lemma 1 we have:

$$\begin{aligned} & L(p, p - \eta + \varepsilon) + L(p, p - \eta - \varepsilon) - L(p, p + \eta + \varepsilon) - L(p, p + \eta - \varepsilon) \\ &= \int_p^{p-\eta+\varepsilon} v''(s)(s-p)ds + \int_p^{p-\eta-\varepsilon} v''(s)(s-p)ds \\ &\quad - \int_p^{p+\eta+\varepsilon} v''(s)(s-p)ds - \int_p^{p+\eta-\varepsilon} v''(s)(s-p)ds \\ &= \int_0^{\eta-\varepsilon} v''(p-\delta)\delta d\delta + \int_0^{\eta+\varepsilon} v''(p-\delta)\delta d\delta \\ &\quad - \int_0^{\eta+\varepsilon} v''(p+\delta)\delta d\delta - \int_0^{\eta-\varepsilon} v''(p+\delta)\delta d\delta \\ &> 0, \end{aligned}$$

where the final inequality uses assumption (5) and the fact that both  $\eta - \varepsilon$  and  $\eta + \varepsilon$  are positive.

Let us now consider the case in which  $\varepsilon \geq \eta$ . In this case, Lemma 1 gives:

$$\begin{aligned} & L(p, p - \eta + \varepsilon) + L(p, p - \eta - \varepsilon) - L(p, p + \eta + \varepsilon) - L(p, p + \eta - \varepsilon) \\ &= - \int_{p-\eta+\varepsilon}^{p+\eta+\varepsilon} v''(s)(s-p)ds - \int_{p-\eta-\varepsilon}^{p+\eta-\varepsilon} v''(s)(s-p)ds \\ &= - \int_{\varepsilon-\eta}^{\varepsilon+\eta} v''(p+\delta)\delta d\delta + \int_{\varepsilon-\eta}^{\varepsilon+\eta} v''(p-\delta)\delta d\delta \\ &> 0, \end{aligned}$$

where again the final inequality follows from (5).

We have proved that no message  $m < p$  can be optimal. It remains to show that  $m = p$  is not optimal either. Using the symmetry of  $g$  and (5), we obtain:

$$\begin{aligned} \mathbb{E}_\varepsilon [v''(p + \varepsilon)\varepsilon] &= \int_{-\sigma}^0 v''(p + \varepsilon)\varepsilon g(\varepsilon)d\varepsilon + \int_0^\sigma v''(p + \varepsilon)\varepsilon g(\varepsilon)d\varepsilon \\ &= - \int_0^\sigma v''(p - \varepsilon)\varepsilon g(\varepsilon)d\varepsilon + \int_0^\sigma v''(p + \varepsilon)\varepsilon g(\varepsilon)d\varepsilon \\ &< 0. \end{aligned}$$

Thus, the message  $m = p$  cannot satisfy the first-order condition (4) that is necessary for optimality. Since the optimal message exists, it must be that  $m^*(p) > p$ .  $\square$

*Proof of Lemma 3.* We first relate the current setting, which has a continuous payoff state  $\Delta_1 \in \mathbb{R}$ , to our main model with binary payoff state  $\theta \in \{0, 1\}$ . We do this by rescaling the problem. For each  $\Delta_1$  and  $\hat{\Delta}_1$ , we fix  $\underline{\Delta}(\Delta_1, \hat{\Delta}_1) < \min\{\Delta_1, \hat{\Delta}_1\}$ , and  $\overline{\Delta}(\Delta_1, \hat{\Delta}_1) > \max\{\Delta_1, \hat{\Delta}_1\}$ . For brevity we omit the arguments of  $\underline{\Delta}$  and  $\overline{\Delta}$  and let  $p, q \in (0, 1)$  be given by  $p = \frac{\Delta_1 - \underline{\Delta}}{\overline{\Delta} - \underline{\Delta}}$  and  $q = \frac{\hat{\Delta}_1 - \underline{\Delta}}{\overline{\Delta} - \underline{\Delta}}$ . We define the utility function  $u$  by:

$$u(a, \theta) = \begin{cases} x_1^1 + x_2^1 & \text{if } a = 1, \\ x_1^1 + \underline{\Delta} + x_2^2 & \text{if } a = 2 \text{ and } \theta = 0, \\ x_1^1 + \overline{\Delta} + x_2^2 & \text{if } a = 2 \text{ and } \theta = 1, \end{cases} \quad (9)$$

and, as in our main model, let  $L(p, q)$  be the loss from the misperception of the objective probability  $p$  of the state  $\theta = 1$  for the subjective value  $q$ . The definition of  $L(p, q)$  does not depend on  $x_1^1$  since it enters payoffs for both actions additively.

We show that both loss functions are isomorphic in the sense that  $L(p, q) = \tilde{L}(\Delta_1, \hat{\Delta}_1)$ . Recall that  $L(p, q) = \mathbb{E}_u L_u(p, q)$ , let

$$\tilde{L}_{\Delta_2}(\Delta_1, \hat{\Delta}_1) = (\Delta_1 + \Delta_2) \left( \mathbf{1}_{\Delta_1 + \Delta_2 > 0} - \mathbf{1}_{\hat{\Delta}_1 + \Delta_2 > 0} \right),$$

and observe that  $\tilde{L}(\Delta_1, \hat{\Delta}_1) = \mathbb{E}_{\Delta_2} \tilde{L}_{\Delta_2}(\Delta_1, \hat{\Delta}_1)$ . Note first that  $L_u(p, q) = 0$  if and only if  $pu(a, 1) + (1 - p)u(a, 0)$  and  $qu(a, 1) + (1 - q)u(a, 0)$  are maximized for the same  $a \in \{1, 2\}$ , i.e. if and only if  $p\overline{\Delta} + (1 - p)\underline{\Delta} + \Delta_2$  and  $q\overline{\Delta} + (1 - q)\underline{\Delta} + \Delta_2$  have a same sign, which happens iff  $\Delta_1 + \Delta_2$  and  $\hat{\Delta}_1 + \Delta_2$  have the same sign, i.e. iff  $\tilde{L}_{\Delta_2}(\Delta_1, \hat{\Delta}_1) = 0$ . When

$L_u(p, q) > 0$ , then

$$\begin{aligned}
L_u(p, q) &= |pu(1, 1) + (1 - p)u(1, 0) - pu(2, 1) + (1 - p)u(2, 0)| \\
&= |p\bar{\Delta} + (1 - p)\underline{\Delta} + \Delta_2| \\
&= |\Delta_1 + \Delta_2| \\
&= \hat{L}_{\Delta_2}(\Delta_1, \hat{\Delta}_1).
\end{aligned}$$

We have thus shown that, for every  $u, p, q$ :  $L_u(p, q) = \hat{L}_{\Delta_2}(\Delta_1, \hat{\Delta}_1)$  and it follows that  $L(p, q) = \tilde{L}(\Delta_1, \hat{\Delta}_1)$ .

We now compute the value function  $v$  of the rescaled decision problem with random utilities  $(u(a, \theta))_{a, \theta}$ . Let us denote  $\Delta_1(s) = s\bar{\Delta} + (1 - s)\underline{\Delta}$  for  $s \in [0, 1]$ . We have:

$$\begin{aligned}
v(s) &= \int_{-\infty}^{\infty} \int_{-\infty}^{x_2^1 - \Delta_1(s)} (x_1^1 + x_2^1) \varrho(x_2^1, x_2^2) dx_2^2 dx_2^1 \\
&\quad + \int_{-\infty}^{\infty} \int_{x_2^1 - \Delta_1(s)}^{\infty} (x_1^1 + \Delta_1(s) + x_2^2) \varrho(x_2^1, x_2^2) dx_2^2 dx_2^1,
\end{aligned}$$

where the first summand stands for the contingencies in which the agent optimally chooses the first object and receives  $x_1^1 + x_2^1$ , and the second summand stands for contingencies in which the agent optimally chooses the second object and receives  $x_1^2 + x_2^2 = x_1^1 + \Delta_1(s) + x_2^2$ .

It follows that:

$$v'(s) = (\bar{\Delta} - \underline{\Delta}) \int_{-\infty}^{\infty} \int_{x_2^1 - \Delta_1(s)}^{\infty} \varrho(x_2^1, x_2^2) dx_2^2 dx_2^1,$$

and

$$\begin{aligned}
v''(s) &= (\bar{\Delta} - \underline{\Delta})^2 \int_{-\infty}^{\infty} \varrho(x_2^1, x_2^1 - \Delta_1(s)) dx_2^1 \\
&= (\bar{\Delta} - \underline{\Delta})^2 \psi(-\Delta_1(s)) \\
&= (\bar{\Delta} - \underline{\Delta})^2 \psi(\Delta_1(s)),
\end{aligned}$$

where the last equality follows from the assumed symmetry of  $\psi$ .

Finally, we use Lemma 1 to compute the loss in the original decision problem:

$$\begin{aligned}
\tilde{L}(\Delta_1, \hat{\Delta}_1) &= \int_p^q v''(s)(s-p)ds \\
&= \int_p^q (\bar{\Delta} - \underline{\Delta})^2 \psi(\Delta_1(s))(s-p)ds \\
&= \int_{\Delta_1}^{\hat{\Delta}_1} \psi(\tilde{s})(\tilde{s} - \Delta_1) d\tilde{s},
\end{aligned}$$

where we used the change of variable  $\tilde{s} = \Delta_1(s)$  to establish the last equality.  $\square$

*Proof of Proposition 4.* We use the same rescaling as in the proof of Lemma 3. For a fixed  $\Delta_1$ , we choose sufficiently distant  $\underline{\Delta}(\Delta_1) < \Delta_1$  and  $\bar{\Delta}(\Delta_1) > \Delta_1$ . Again we omit the argument of  $\underline{\Delta}$  and  $\bar{\Delta}$ , let  $p = \frac{\Delta_1 - \underline{\Delta}}{\bar{\Delta} - \underline{\Delta}}$ , and let  $u(a, \theta)$  be as specified in (9). We also rescale the errors by letting  $\tilde{\varepsilon} = \varepsilon / (\bar{\Delta} - \underline{\Delta})$ . Let  $m^*(p)$  be the optimal perception message in the rescaled problem. By construction,  $(m^*(p) - p) (\bar{\Delta} - \underline{\Delta}) = \tilde{m}^*(\Delta_1) - \Delta_1$ .

As we have shown in the proof of Lemma 3, in the rescaled problem, the value function  $v$  satisfies  $v''(s) = (\bar{\Delta} - \underline{\Delta})^2 \psi(\Delta_1(s))$ , where  $\Delta_1(s) = s\bar{\Delta} + (1-s)\underline{\Delta}$ . Thus, if  $\Delta_1 > 0$  then  $v''(p + \alpha) < v''(p - \alpha)$  for all positive  $\alpha$ , and symmetrically, when  $\Delta_1 < 0$ , then  $v''(p + \alpha) > v''(p - \alpha)$ . Then, by Proposition 3, for  $\Delta_1 > 0$  we have  $m^*(p) > p$ , and symmetrically, for  $\Delta_1 < 0$  we have  $m^*(p) < p$ , as needed.  $\square$

*Proof of Lemma 5.* Recall that  $F$  and  $f$  are the cdf and the pdf of the standard normal random variable, respectively. Then,

$$\begin{aligned}
v(p) &= E_u \max \{ pu(1, 1) + (1-p)u(1, 0), \tilde{u} \} \\
&= \tilde{u}F\left(\frac{\tilde{u}}{\sigma_p}\right) + \int_{\tilde{u}}^{\infty} xf\left(\frac{x}{\sigma_p}\right) \frac{dx}{\sigma_p} \\
&= \tilde{u}F\left(\frac{\tilde{u}}{\sigma_p}\right) + \sigma_p \int_{\frac{\tilde{u}}{\sigma_p}}^{\infty} zf(z)dz,
\end{aligned}$$

where  $\sigma_p = \sqrt{p^2 + (1-p)^2}$ . We have used that the lottery value  $pu(1, 1) + (1-p)u(1, 0)$  is  $\mathcal{N}(0, \sigma_p^2)$  in the second line. There, the first summand stands for contingencies in which the lottery value is below  $\tilde{u}$  and the agent obtains  $\tilde{u}$ . The second summand stands for the choice of the lottery. The last line follows from the substitution  $z = x/\sigma_p$ .

The first derivative of the value function is

$$\begin{aligned} v'(p) &= \left( -f\left(\frac{\tilde{u}}{\sigma_p}\right) \frac{\tilde{u}^2}{\sigma_p^2} + \int_{\frac{\tilde{u}}{\sigma_p}}^{\infty} z f(z) dz + f\left(\frac{\tilde{u}}{\sigma_p}\right) \frac{\tilde{u}^2}{\sigma_p^2} \right) \sigma'_p \\ &= \sigma'_p \int_{\frac{\tilde{u}}{\sigma_p}}^{\infty} z f(z) dz, \end{aligned}$$

where  $\sigma'_p$  is the derivative of  $\sigma_p$  with respect to  $p$  (similarly,  $\sigma''_p$  appearing below is the second derivative).

Finally, the second derivative is

$$\begin{aligned} v''(p) &= \sigma_p'^2 \frac{\tilde{u}^2}{\sigma_p^3} f\left(\frac{\tilde{u}}{\sigma_p}\right) + \sigma_p'' \int_{\frac{\tilde{u}}{\sigma_p}}^{\infty} z f(z) dz \\ &= \sigma_p'^2 \frac{\tilde{u}^2}{\sigma_p^3} f\left(\frac{\tilde{u}}{\sigma_p}\right) + \sigma_p'' f\left(\frac{\tilde{u}}{\sigma_p}\right) \\ &= \left( \frac{\tilde{u}^2 (2p-1)^2}{\sigma_p^5} + \frac{1}{\sigma_p^3} \right) f\left(\frac{\tilde{u}}{\sigma_p}\right), \end{aligned}$$

where we have used that  $\int_a^{\infty} z f(z) dz = f(a)$  for the standard normal density  $f$ . The symmetry is immediate. For the derivation of the monotonicity property of this explicit functional form, see Steiner and Stewart (2016), proof of Lemma 2 thereof.<sup>14</sup>  $\square$

*Proof of Lemma 6.* For any given  $m(\cdot)$ , the payoff optimization implies that  $c(u, m) = a_{q(m), u}^*$ , where  $q(m) = \mathbb{E}[p \mid m(p) = m]$ . For any two posteriors  $0 < \underline{q} < \bar{q} < 1$ , the difference between the losses across the two posteriors is

$$L(p, \bar{q}) - L(p, \underline{q}) = \int_{\underline{q}}^{\bar{q}} v''(s)(s - p) ds.$$

Since  $v''$  is positive, the last expression is decreasing in  $p$ , positive at  $p = 0$ , and negative at  $p = 1$ . Therefore, there exists  $p^* \in [0, 1]$  such that the posterior  $\underline{q}$  is loss minimizing below  $p^*$  and  $\bar{q}$  is loss minimizing above  $p^*$ .  $\square$

<sup>14</sup>This functional form was obtained by Steiner and Stewart (2016) as a relevant weighting function for their perception problem. Its derivation through the value function  $v$  is novel and a simplification over their method.

*Proof of Proposition 5.* Let

$$q(p; p^*) = \begin{cases} \bar{q}(p^*) & \text{if } p \geq p^*, \\ \underline{q}(p^*) & \text{if } p < p^*, \end{cases}$$

denote the Bayesian posterior formed at  $p$  when the threshold probability is  $p^*$ .

The agent's problem is

$$\max_{p^*} \mathbb{E}_p v(q(p; p^*)).$$

To utilize our characterization of the loss  $L(p, q)$ , we express the objective in two equivalent forms that both feature the loss function.

In step 1, we show that the agent's problem is equivalent to

$$\max_{p^*} \mathbb{E}_p L(q(p; p^*), 1/2). \quad (10)$$

The expression  $\mathbb{E}_p L(q(p; p^*), 1/2)$  is the difference in expected payoffs between the partially informed agent and an uninformed agent. The more useful the partial information retained by the forgetful agent is, the higher is this difference. To see that (10) is equivalent to the agent's problem, write

$$\begin{aligned} \mathbb{E}_p v(q(p; p^*)) &= v(1/2) + \mathbb{E}_p [v(q(p; p^*)) - v(1/2)] \\ &= v(1/2) + \mathbb{E}_p [v(q(p; p^*)) - V(q(p; p^*), 1/2)] \\ &= v(1/2) + \mathbb{E}_p L(q(p; p^*), 1/2), \end{aligned}$$

where in the second line, we have used  $v(1/2) = V(1/2, 1/2) = \mathbb{E}_p V(q(p; p^*), 1/2)$ . The last equality follows from the linearity of the function  $V(p, q)$  with respect to the first argument, and from  $\mathbb{E}_p q(p; p^*) = 1/2$  (since the posteriors  $q(p; p^*)$  satisfy the martingale property).

In step 2 we prove that the agent's problem is equivalent to

$$\min_{p^*} \mathbb{E}_p L(p, q(p; p^*)). \quad (11)$$

The expression  $\mathbb{E}_p L(p, q(p; p^*))$  represents the difference in expected payoffs between the fully informed agent and the partially informed agent. The more useful the partial infor-

mation is, the lower is this difference. To see this equivalence, write

$$\begin{aligned} \mathbb{E}_p v(q(p; p^*)) &= \mathbb{E}_p V(q(p; p^*), q(p; p^*)) \\ &= \mathbb{E}_p V(p, q(p; p^*)) \\ &= \mathbb{E}_p [v(p) - L(p, q(p; p^*))], \end{aligned}$$

where in the second line, we have again used the linearity of  $V$  with respect to the first argument, and that  $\mathbb{E}_p [p \mid q(p; p^*) = q] = q$  for both Bayesian posteriors  $q \in \{\underline{q}(p^*), \bar{q}(p^*)\}$ .

In step 3, we use the objective (10) to rule out optimality of  $p^* > 1/2$ . We show, using (10), that for any  $p^* > 1/2$ , the agent is better off using a threshold of  $1 - p^*$  rather than  $p^*$ . Let us rewrite the objective in (10) as

$$\mathbb{E}_p L(q(p; p^*), 1/2) = \frac{\bar{q}(p^*) - 1/2}{\bar{q}(p^*) - \underline{q}(p^*)} L(\underline{q}(p^*), 1/2) + \frac{1/2 - \underline{q}(p^*)}{\bar{q}(p^*) - \underline{q}(p^*)} L(\bar{q}(p^*), 1/2), \quad (12)$$

where the fractions are the expressions for  $\Pr(p < p^*)$  and  $\Pr(p \geq p^*)$ , respectively.

Substituting (1) for  $L(q, 1/2)$  into (12) gives

$$\mathbb{E}_p L(q(p; p^*), 1/2) = \int_0^1 v''(s) h(s, p^*) ds,$$

where

$$h(s, p^*) = \begin{cases} 0 & \text{if } s < \underline{q}(p^*), \\ \frac{\bar{q}(p^*) - 1/2}{\bar{q}(p^*) - \underline{q}(p^*)} (s - \underline{q}(p^*)) & \text{if } \underline{q}(p^*) \leq s < 1/2, \\ \frac{1/2 - \underline{q}(p^*)}{\bar{q}(p^*) - \underline{q}(p^*)} (\bar{q}(p^*) - s) & \text{if } 1/2 \leq s < \bar{q}(p^*), \\ 0 & \text{if } s > \bar{q}(p^*). \end{cases}$$

The function  $h(s, p^*)$  is continuous piecewise-linear in  $s$ , as depicted in figure 3.

Let us now prove that  $\mathbb{E}_p L(q(p; p^*), 1/2) < \mathbb{E}_p L(q(p; 1 - p^*), 1/2)$  when  $p^* > 1/2$ . First, note that the symmetry of the density of  $p$ ,  $\varphi$ , around  $1/2$  implies the relationships:

$$\begin{aligned} \bar{q}(1 - p^*) &= 1 - \underline{q}(p^*) \\ \underline{q}(1 - p^*) &= 1 - \bar{q}(p^*). \end{aligned}$$

Therefore, for any  $p^*, s \in (0, 1)$ :

$$\begin{aligned}
h(1-s, 1-p^*) &= \begin{cases} 0 & \text{if } 1-s < \underline{q}(1-p^*) \\ \frac{\bar{q}(1-p^*)-1/2}{\bar{q}(1-p^*)-\underline{q}(1-p^*)}((1-s)-\underline{q}(1-p^*)) & \text{if } \underline{q}(1-p^*) \leq 1-s < 1/2 \\ \frac{1/2-\underline{q}(1-p^*)}{\bar{q}(1-p^*)-\underline{q}(1-p^*)}(\bar{q}(1-p^*)-(1-s)) & \text{if } 1/2 \leq 1-s < \bar{q}(1-p^*) \\ 0 & \text{if } 1-s \geq \bar{q}(1-p^*) \end{cases} \\
&= \begin{cases} 0 & \text{if } s > \bar{q}(p^*) \\ \frac{1/2-\underline{q}(p^*)}{\bar{q}(p^*)-\underline{q}(p^*)}(\bar{q}(p^*)-s) & \text{if } 1/2 < s \leq \bar{q}(p^*) \\ \frac{\bar{q}(p^*)-1/2}{\bar{q}(p^*)-\underline{q}(p^*)}(s-\underline{q}(p^*)) & \text{if } \underline{q}(p^*) < s \leq 1/2 \\ 0 & \text{if } s \leq \underline{q}(p^*) \end{cases} \\
&= h(s, p^*).
\end{aligned}$$

Second, note that for  $p^* > 1/2$ ,

$$\begin{aligned}
\underline{q}(p^*) &= \mathbb{E}[p|p < p^*] \\
&= \Pr(p < 1-p^*|p < p^*) \mathbb{E}[p|p < 1-p^*] \\
&\quad + \Pr(1-p^* \leq p < p^*|p < p^*) \mathbb{E}[p|1-p^* < p \leq p^*] \\
&= \Pr(p < 1-p^*|p < p^*)\underline{q}(1-p^*) + \Pr(1-p^* \leq p < p^*|p < p^*)1/2 \\
&> \Pr(p < 1-p^*|p < p^*)(1-\bar{q}(p^*)) + \Pr(1-p^* \leq p < p^*|p < p^*)(1-\bar{q}(p^*)) \\
&= 1-\bar{q}(p^*).
\end{aligned}$$

We now compare  $h(s, p^*)$  with  $h(1-s, p^*)$  for  $p^* > \frac{1}{2}$ , as illustrated in figure 3. On the range  $0 \leq s \leq \frac{1}{2}$ ,  $h(s, p^*)$  is null for  $s$  below  $\underline{q}(p^*)$ , then increases linearly from 0 to its maximum  $h(\frac{1}{2}, p^*) = \frac{(\bar{q}(p^*)-1/2)(1/2-\underline{q}(p^*))}{\bar{q}(p^*)-\underline{q}(p^*)}$ . In comparison,  $h(s, 1-p^*)$  is null for  $s$  below  $\underline{q}(1-p^*)$  where  $\underline{q}(1-p^*) < \underline{q}(p^*)$ , then increases linearly from 0 to its maximum  $h(\frac{1}{2}, 1-p^*) = \frac{(\bar{q}(1-p^*)-1/2)(1/2-\underline{q}(1-p^*))}{\bar{q}(1-p^*)-\underline{q}(1-p^*)} = h(\frac{1}{2}, p^*)$ . This implies that for every  $p^* > \frac{1}{2}$  and  $s < 1/2$ ,

$$h(s, p^*) \leq h(1-s, p^*), \quad (13)$$

with strict inequality when  $\underline{q}(1-p^*) < s < \frac{1}{2}$ .

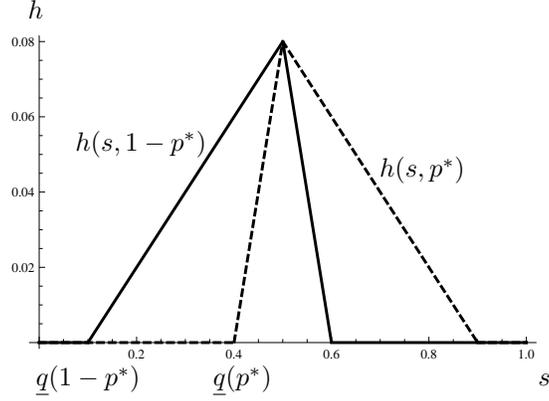


Figure 3: Comparison of  $h(s, p^*)$  and  $h(s, 1 - p^*)$  for  $p^* > 1/2$ . Dashed line:  $h(s, p^*)$ . Full line:  $h(s, 1 - p^*)$ .

Now, for any  $p^* > 1/2$ , we have:

$$\begin{aligned}
& \mathbb{E}_p [L(q(p; p^*), 1/2) - L(q(p; 1 - p^*), 1/2)] \\
&= \int_0^1 v''(s) (h(s, p^*) - h(s, 1 - p^*)) ds \\
&= \int_0^1 v''(s) (h(s, p^*) - h(1 - s, p^*)) ds \\
&= \int_0^{\frac{1}{2}} (v''(s) - v''(1 - s)) (h(s, p^*) - h(1 - s, p^*)) ds.
\end{aligned}$$

The first bracket in the last integral is positive by assumption. When  $p^* > 1/2$ , the second bracket is non-positive and strictly negative on some range of integration by (13). Therefore, for  $p^* > 1/2$ ,

$$\mathbb{E}_p L(q(p; 1 - p^*), 1/2) > \mathbb{E}_p L(q(p; p^*), 1/2),$$

which shows that  $p^* > 1/2$  cannot be optimal.

It remains for us to prove that  $p^* \neq 1/2$ , which we do in step 4.<sup>15</sup> Let us write the

---

<sup>15</sup>We are grateful to an anonymous referee who has improved this part of the proof.

objective in (11) as

$$\begin{aligned} \mathbb{E}_p L(p, q(p; p^*)) &= \int_0^{p^*} \varphi(p) L(p, \underline{q}(p^*)) dp + \int_{p^*}^1 \varphi(p) L(p, \bar{q}(p^*)) dp \\ &:= k(p^*, \underline{q}(p^*), \bar{q}(p^*)). \end{aligned}$$

The total derivative of this objective with respect to  $p^*$  is

$$\begin{aligned} \left. \frac{d}{dp^*} \mathbb{E}_p L(p, q(p; p^*)) \right|_{p^*=1/2} &= \varphi(1/2) (L(1/2, \underline{q}(1/2)) - L(1/2, \bar{q}(1/2))) \\ &+ \frac{\partial}{\partial \underline{q}} k(1/2, \underline{q}, \bar{q}) \frac{d}{dp^*} \underline{q}(1/2) \\ &+ \frac{\partial}{\partial \bar{q}} k(1/2, \underline{q}, \bar{q}) \frac{d}{dp^*} \bar{q}(1/2). \end{aligned}$$

Since the Bayesian posterior values  $\underline{q}(p^*)$  and  $\bar{q}(p^*)$  are optimal for the given threshold  $p^*$ , the second line is zero by the envelope theorem. When we replace the loss function with the formula of Lemma 1, we obtain

$$\left. \frac{d}{dp^*} \mathbb{E}_p L(p, q(p; p^*)) \right|_{p^*=1/2} = \varphi(1/2) \int_{\underline{q}(1/2)}^{\bar{q}(1/2)} v''(s) (1/2 - s) ds.$$

The symmetry of  $\varphi$  around  $1/2$  implies  $\underline{q}(1/2) = 1 - \bar{q}(1/2)$ . We then deduce:

$$\begin{aligned} \left. \frac{d}{dp^*} \mathbb{E}_p L(p, q(p; p^*)) \right|_{p^*=1/2} &= \varphi(1/2) \int_{\underline{q}(1/2)}^{1-\underline{q}(1/2)} v''(s) (1/2 - s) ds \\ &= \varphi(1/2) \int_{\underline{q}(1/2)}^{1/2} (v''(s) - v''(1-s)) (p^* - s) ds \\ &> 0, \end{aligned}$$

where the inequality follows from assumption (7). Thus,  $p^* = 1/2$  does not satisfy the first-order condition, so it cannot be optimal.  $\square$

## B Illusion of Control—General Distributions

We consider binary action sets in this section and utilities  $u(a, \theta) = \tilde{u}_a + \tau_\theta \eta_a$ ,  $\tau_1 > \tau_0 > 0$ ,  $a \in \{1, 2\}$ . We dispense with the assumption of Gaussian payoffs and provide sufficient conditions on general payoff distributions for the monotonicity of  $v''$ .

Let us introduce the differences  $\delta = \tilde{u}_2 - \tilde{u}_1$ , and  $\eta = \eta_2 - \eta_1$ ; we denote their densities by  $\chi(\delta)$  and  $\phi(\eta)$ . We allow for dependencies between  $\tilde{u}_1$  and  $\tilde{u}_2$  and between  $\eta_1$  and  $\eta_2$  but continue to assume that  $(\tilde{u}_1, \tilde{u}_2)$  and  $(\eta_1, \eta_2)$  are independent.

**Lemma 8.**

$$v''(p) = (\tau_1 - \tau_0)^2 \int_{-\infty}^{\infty} \eta^2 \phi(\eta) \chi(-\tau(p)\eta) d\eta. \quad (14)$$

The proof of the lemma is at the end of this subsection. Let us provide here a heuristic argument that emphasizes the connection between the likelihood of ties and the second derivative of the value function. Let  $\tau(p) = \tau_1 p + \tau_0(1 - p)$ , and let  $\Delta(p) = \delta + \tau(p)\eta$  be the payoff difference between the two actions, and  $p^*$  be the random belief at which the agent is indifferent between the two actions; that is for each  $\delta$  and  $\eta$ ,  $p^*$  solves  $\Delta(p^*) = 0$ . For each  $\eta$ , the conditional density of  $p^*$  is  $\chi(-\tau(p)\eta) \left| \frac{d\Delta}{dp} \right| = \chi(-\tau(p)\eta) |(\tau_1 - \tau_0)\eta|$ . Thus, the unconditional density of  $p^*$  is

$$\int_{-\infty}^{\infty} \chi(-\tau(p)\eta) |(\tau_1 - \tau_0)\eta| \phi(\eta) d\eta.$$

The intensity of the kink of the value function is  $|(\tau_1 - \tau_0)\eta|$ . The expression (14) for  $v''$  augments the last integral by the intensity of the kinks.

The last lemma allows us to establish the monotonicity of  $v''$  under a regularity condition without computing  $v$ .

**Lemma 9.** *If the density  $\chi(\delta)$  is single-peaked with its maximum at 0, then  $v''$  is decreasing.*

The difference  $\delta = \tilde{u}_2 - \tilde{u}_1$  has a single-peaked density when  $\tilde{u}_1$  and  $\tilde{u}_2$  are iid. draws from many common densities, including the normal or Pareto densities. Additionally, we provide a condition on the joint distribution of  $\tilde{u}_1$  and  $\tilde{u}_2$  that is sufficient for the single-peakedness of  $\chi(\delta)$ . Recall that a function  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$  is supermodular if for any  $x, y, x', y' \in \mathbb{R}^4$ :

$$h(\max\{x, x'\}, \max\{y, y'\}) + h(\min\{x, x'\}, \min\{y, y'\}) \geq h(x, y) + h(x', y').$$

Let  $g(\tilde{u}_1, \tilde{u}_2)$  denote the joint density of  $\tilde{u}_1$  and  $\tilde{u}_2$ .

**Lemma 10.** *If  $g$  is supermodular and  $g(\tilde{u}_1, \tilde{u}_2) = g(\tilde{u}_2, \tilde{u}_1)$  for every  $(\tilde{u}_1, \tilde{u}_2)$ , then  $\chi(\delta)$  is single-peaked with maximum at 0.*

The supermodularity assumption is related to the definition of affiliated random variables (Milgrom and Weber, 1982). Random variables are affiliated if the logarithm of their joint density is supermodular. Our assumption applies to the density instead of the log-density.

The following result concludes this section by summarizing conditions under which upward bias obtains.

**Corollary 4.** *If  $u(a, \theta) = \tilde{u}_a + \tau_\theta \eta_a$ ,  $\tau_1 > \tau_0 > 0$  and the joint density  $g$  is both symmetric and supermodular, then  $v''$  is decreasing and the solution of the naive perception problem (3) is biased upwards:  $m^*(p) > p$ .*

*Proof of Lemma 8.* Using insight (8) from the beginning of the proof of Lemma 1, we obtain:

$$\begin{aligned} v'(p) &= \mathbf{E}_u [u(a_{p,u}^*, 1) - u(a_{p,u}^*, 0)] \\ &= \mathbf{E}_u [\tilde{u}_{a_{p,u}^*} + \tau_1 \eta_{a_{p,u}^*} - \tilde{u}_{a_{p,u}^*} - \tau_0 \eta_{a_{p,u}^*}] \\ &= (\tau_1 - \tau_0) \mathbf{E}_u \eta_{a_{p,u}^*}. \end{aligned}$$

Recall that  $\tau(p) = p\tau_1 + (1-p)\tau_0$ . The agent chooses action 1 if  $-\tau(p)\eta > \delta$ , and chooses action 2 if  $-\tau(p)\eta < \delta$ . Thus, if we denote by  $\mu(\eta_1, \eta_2)$  the joint density of  $(\eta_1, \eta_2)$ ,

$$\begin{aligned} v'(p) &= \left( \int_{\{(\eta_1, \eta_2, \delta): -\tau(p)\eta > \delta\}} \eta_1 \chi(\delta) \mu(\eta_1, \eta_2) d\delta d\eta_1 d\eta_2 \right. \\ &\quad \left. + \int_{\{(\eta_1, \eta_2, \delta): -\tau(p)\eta < \delta\}} \eta_2 \chi(\delta) \mu(\eta_1, \eta_2) d\delta d\eta_1 d\eta_2 \right) \times (\tau_1 - \tau_0) \\ &= (\tau_1 - \tau_0) \int_{\mathbb{R}^2} \mu(\eta_1, \eta_2) \left( \int_{-\infty}^{-\tau(p)\eta} \eta_1 \chi(\delta) d\delta + \int_{-\tau(p)\eta}^{\infty} \eta_2 \chi(\delta) d\delta \right) d\eta_1 d\eta_2, \end{aligned}$$

and

$$\begin{aligned}
v''(p) &= (\tau_1 - \tau_0) \\
&\times \int_{\mathbb{R}^2} \mu(\eta_1, \eta_2) \left( -\eta_1 \chi(-\tau(p)\eta) (\tau_1 - \tau_0)\eta + \eta_2 \chi(-\tau(p)\eta) (\tau_1 - \tau_0)\eta \right) d\eta_1 d\eta_2 \\
&= (\tau_1 - \tau_0)^2 \int_{\mathbb{R}^2} \mu(\eta_1, \eta_2) \chi(-\tau(p)\eta) \eta^2 d\eta_1 d\eta_2 \\
&= (\tau_1 - \tau_0)^2 \int_{-\infty}^{\infty} \eta^2 \phi(\eta) \chi(-\tau(p)\eta) d\eta,
\end{aligned}$$

as stated in the lemma.  $\square$

*Proof of Lemma 9.* The result follows from (14): it follows from  $\tau_1 > \tau_0$  that  $\tau(p)$  is increasing in  $p$ , which implies that  $\chi(-\tau(p)\eta)$  is decreasing in  $p$  for all  $p$  and  $\eta$ .  $\square$

*Proof of Lemma 10.* The symmetry of  $g$  implies  $\chi(\delta) = \chi(-\delta)$  for all  $\delta$ . Thus, for arbitrary  $t \in \mathbb{R}$ ,

$$\chi(\delta) = \int_{-\infty}^{\infty} g(x+t, x+t+|\delta|) dx = \int_{-\infty}^{\infty} g(x+t, x+t-|\delta|) dx. \quad (15)$$

Consider  $\delta, \delta'$  such that  $|\delta'| > |\delta|$ , and let  $z = |\delta'| - |\delta| > 0$ . Using (15), we get:

$$\begin{aligned}
2(\chi(\delta) - \chi(\delta')) &= \int_{-\infty}^{\infty} g(u, u+|\delta|) + g(u-z, u-z-|\delta|) \\
&\quad - g(u, u-|\delta'|) - g(u-z, u-z+|\delta'|) du.
\end{aligned}$$

Since,

$$u + |\delta| = u - z + |\delta'| \quad \text{and} \quad u - |\delta'| = u - z - |\delta|,$$

the supermodularity of  $g$  implies that the last integrand is positive.  $\square$