amse
école d'économie d'aix-marseille
aix-marseille school of economics

Working Papers / Documents de travail

# The benefits of being misinformed

Marcus Roel
Manuel Staab

WP 2021 - Nr 08

# The benefits of being misinformed *

Marcus Roel
mcs.roel@gmail.com
Beijing Normal University

Manuel Staab
manuel.staab@univ-amu.fr
Aix-Marseille University, CNRS, AMSE

February 16, 2021

**Abstract**

In the spirit of Blackwell (1951), we analyze how two fundamental mistakes in information processing - incorrect beliefs about the world and misperception of information - affect the expected utility ranking of information experiments. We explore their individual and combined influence on welfare and provide necessary and sufficient conditions when mistakes alter and possibly reverse the ranking of information experiments. Both mistakes by themselves reduce welfare in a model where payoff relevant actions also generate informative signals. This is true for naive decision-makers, unaware of any errors, as well as for sophisticated decision-makers, who account for the possibility of mistakes. However, mistakes can interact in non-obvious ways and an agent might be better off suffering from both, rather than just one. We provide a characterization when such positive interactions are possible. Surprisingly, this holds true only for naive decision-makers and thus naivete can be beneficial. We discuss implications for information acquisition and avoidance, welfare-improving belief manipulation, and policy interventions in general.

**Keywords:** ranking of experiments, information acquisition, misperception, confirmation bias, overconfidence, underconfidence
**JEL Codes:** D03, D81, D83

*We would like to thank Andrew Ellis, Erik Eyster, Gilat Levy, Matthew Levy, Francesco Nava, Ronny Razin, Balázs Szentes, and seminar participants at the London School of Economics for helpful comments and suggestions. This work was partly supported by French National Research Agency Grant ANR-17-EURE-0020.

# 1. INTRODUCTION

We commonly encounter situations in which information is difficult to evaluate. In such circumstances, we can observe people taking actions in line with hypotheses that seem to be disfavored or even contradicted by empirical evidence. Furthermore, these actions often produce additional information that should highlight the initial mistake and lead to subsequent adjustments. Yet learning often appears limited at best. For example, a significant number of people refuse even essential vaccinations despite the very strong evidence of their benefit and despite the measurable increase in outbreaks of the related disease as a consequence of this refusal.[1] In light of these observations, we revisit Blackwell's (1951) comparison of experiments under the assumption that information processing is not always flawless but impeded by inaccuracies and systematic mistakes.

We analyze a simple model that captures the fundamentals of information acquisition and Bayesian updating: first an agent takes a payoff-relevant action that also provides information about the state of the world. The agent then takes a second action before payoffs are realized. In this setting, information processing can be imperfect in two ways: (1) an agent might hold incorrect beliefs about the world and (2) might misperceive information. Both imperfections are motivated by the psychological and experimental literature on beliefs and perception: while (1) captures concepts such as over- or underconfidence (Fischoff et al. (1977), Lichtenstein et al. (1982), Moore and Healy (2008)) or motivated beliefs (Epley and Gilovich's (2016)), (2) broadly covers directional mistakes, such as confirmation bias (Bruner and Potter (1964), Darley and Gross (1983), Fischoff et al. (1977) Lichtenstein et al. (1982)) or one-sided updating to protect one's ego utility or self-image (Mobius et al. (2014), Eil and Rao (2011)), as well as simple errors. Our model can thus be used to study a wide variety of imperfections in information processing; from random inaccuracies to systematic mistakes.

Incorrect beliefs about the world decrease the agent's welfare as they distort choices away from the optimal ones. Misperception reduces the information value of experiments and consequently welfare. This is true for agents who are aware of their tendency to misperceive information (sophisticated) as well as those who aren't (naive). For a given type of misperception, naive decision-makers tend to be worse off than sophisticated ones as they fail to adjust their choices to the lower information value. To understand when mis-

[1] See, for instance, Poland and Jacobson (2001) and Larson et al. (2011) for an overview of factors shaping public (dis-)trust in vaccine safety and efficacy, and their consequences for public health. Motta et al. (2018) provides evidence for widespread misinformation and overconfidence regarding medical knowledge in the general population in the U.S..

perception can alter the ranking of information experiments, we derive necessary and sufficient conditions for four broad classes of misperception. Each of these is only based on a simple comparison of outcomes that allows us to identify if and which alternative choices can mitigate the impact of misperception.

Building on these findings, we analyze how incorrect beliefs and misperception interact. This analysis can be read from either a positive or normative perspective. Since most people are behavioral in many ways (Stango and Zinman (2020)), it is important to understand how typical biases relating to information interact and how they impact welfare. For example, when observing heterogeneous choices despite identical preferences, this allows us to identify which choices could be a constrained optimum for a decision-maker facing issues with information processing and which ones are positively suboptimal. Taking a normative approach, we may ask whether a rational observer could intervene and help a biased decision-maker. Without holding any additional information, the only feasible interventions of such an intermediary are preventing the transmission of useful information or providing outright wrong information. With only one imperfection present, magnifying this imperfection cannot be beneficial. However, adding or intensifying a second mistake can have a positive welfare effect and thus beneficial interventions without superior information are feasible.

More specifically, if a decision-maker suffers from misperception, an incorrect initial belief can turn out to be welfare improving. A sufficiently biased belief pushes the decision-maker to an alternative course of action. While suboptimal in a perfect world, it might mitigate some of the welfare impact of misperception; either because the alternative is less susceptible to misperception or less sensitive to information. We show that if misperception is severe enough, such an improvement always exists. Based on the same conditions as for the utility rankings, we fully characterize when a decision-maker suffering from misperception is better off also holding an incorrect belief.

We also address the opposite question: if an objective observer notices that an agent's belief is incorrect, could this observer intervene and falsify information to the benefit of the agent. In other words, can misperception be useful when beliefs are wrong. If the observer considers the original information valuable for the agent's second-period choices, the answer is negative. Misreported information is, however, beneficial if the agent's belief is sufficiently far away from the truth such that the observer prefers the agent to take a particular action regardless of the outcome of the experiment.

We also find a somewhat surprising result regarding sophistication: a sophisticated

decision-maker may actually do worse than a naive one when both have the same form of misperception *and* the same wrong belief. This happens when a sophisticated agent mistakenly avoids some information source due to correctly adjusting the information value downward to account for misperception while at the same time undervaluing the information due to holding an incorrect belief.

We motivate and illustrate this analysis with two examples: (1) a patient consulting a doctor for a diagnostic test and (2) a company considering how to best launch a new product. (1) serves as our leading example throughout while (2) illustrates some particular implications regarding information acquisition and risk. For both examples we highlight situations in which information is best avoided - for instance by minimizing extensive market research - or even beneficially falsified - for example by deliberately providing an inaccurate medical history.

Our setting captures typical problems of information acquisition and fully characterizes the implications of potential mistakes. From a broader perspective, our results highlight the need for a comprehensive understanding of imperfections in information processing to improve decision-making. One-sided approaches that address either misperception or incorrect beliefs can have unexpected side-effects due to the intricate interaction of both biases. For instance, increasing people's confidence in the vaccine development process might be insufficient to sway sceptics and could even harden their opposition if subsequent information about vaccine safety and efficacy is open to misinterpretation. Nevertheless, as we show in this paper, interventions that jointly identify and mitigate biases can be feasible and useful.

The rest of the paper is organized as follows: the next section summarizes the relevant literature. This is followed by a description of our setting in section 3 and a characterization of the unbiased choice problem in section 4. Section 5 introduces biases in information processing. Our main results can be found in section 6, where the implications and interactions of these biases are explored. We illustrate some of the implications with an example of a company launching a new product in section 7. Finally, we conclude the paper in section 8 with a broader discussion of our four misperception classes and their implications. Furthermore, we explore the connection between misperception and ambiguity over the information content of signals. All proofs can be found in the appendix.

## 2. LITERATURE

Blackwell (1951) formalizes when an information experiment is more informative than another. Marschak and Miyasawa (1968) transfer these statistical ideas to the realm of economics. The key finding is that no rational decision-maker would choose to 'garble' their information, i.e. voluntarily introduce noise into experiments. Having more information, however, may not always be beneficial in economic settings and might sometimes even cause a disadvantage in strategic interactions. For example, Hirshleifer (1971) highlights that public information may destroy mutually beneficial insurance possibilities. Information avoidance has also been documented in bargaining (Schelling (1956), Schelling (1960), Conrads and Irlenbusch (2013), Poulsen and Roos (2010)) and holdup problems (Tirole (1986), Rogerson (1992), Gul (2001)). Strategic benefits can also arise when a behavioral agent plays intrapersonal games. Carrillo and Mariotti (2000) and Benabou and Tirole (2002) show that garbling of information can increase the current self's payoff when individuals are time-inconsistent.

Holding incorrect beliefs can equally entail benefits in some settings. Ludwig et al. (2011) show that overconfidence can improve an agent's relative and absolute performance in contests by inducing higher efforts. In moral hazard problems, an agent's overconfidence makes it easier for the principal to induce effort, which can improve the agent's welfare (De La Rosa (2011)).

There have been many studies that suggest people hold incorrect beliefs. On average, people tend to have unrealistically positive views of their traits or prospects. To mention a few, see Weinstein (1980) for health and salaries, Guthrie et al. (2001) for rates of overturned decisions on appeal by judges, and Fischoff et al. (1977) as well as Lichtenstein et al. (1982) for estimates of ones' own likelihood to answer correctly. Recent papers document overconfidence in entrepreneurs (Landier and Thesmar (2009)), in CEOs (Malmendier and Tate (2005)), and in laboratory settings (Burks et al. (2013), Charness et al. (2018), Benoit and Moore (2015)).

Perception bias has first been documented in the psychology literature, see, for example, Bruner and Potter (1964), Fischoff et al. (1977) Lichtenstein et al. (1982), and Darley and Gross (1983). The literature has explored many ways of modeling such biases, with different implications for learning. For example, Rabin and Schrag (1999) formalized them in a model of confirmation bias. They show how the tendency to misinterpret new information as supportive evidence for one's currently held hypothesis can not only lead to overconfidence in the incorrect hypothesis, but even cause someone to become fully

convinced of it. Recent evidence for such one-sided updating include Mobius et al. (2014) and Eil and Rao (2011).

Another strand of the literature analyzes how various behavioral features, which may or may not be shortcomings, can be improved upon by overconfidence, biased beliefs, or misperception. Often, these papers aim to provide a motivation why overconfidence exists in the first place. For example, Compte and Postlewaite (2004) highlights how failing to recall past failures can improve welfare for it counteracts the fear of failures. In the extreme, agents may simply derive utility from beliefs. In Brunnermeier and Parker (2005), agents prefer to hold too optimistic beliefs about the future because they derive immediate benefits from these expectations. The closest paper to ours in this regard is Steiner and Stewart (2016), who suggest that pure noise inherent in information processing creates problems akin to the winner's curse when unbiased perception strategies are employed. Optimal perception must therefore be biased, correcting for the mistake by inducing more cautious evaluations. While our focus is also on decision problems, not games, our emphasis lies on choices and welfare consequences taking perception as given, not on explaining perception (mistakes). To our knowledge, this is the first paper that, instead of focusing on a particular perception shortcoming, reduces them to a general but simple misperception matrix that can accommodate many different types of biases.

## 3. THE SETTING

We consider a two period model with two states of the world $\Omega = \{A, B\}$. In the first period, the agent chooses an action from a finite space $X_1$ and subsequently receives a signal $s \in S$ about the state. The quality of the signal depends on the action chosen. For each action $x \in X_1$, there is a distinct binary set of possible signals $S(x) = \{a(x), b(x)\} \subset S$ to which we will refer to as $a$- and $b$-signals. The probability of receiving an $a$-signal in state A is denoted as $\pi(x) \equiv \Pr(a(x)|A, x)$ and to keep the notation simple, we assume that the signal structure is symmetric between states. The probability of receiving a $b$-signal in state B thus equals the probability of receiving an $a$-signal in state A, meaning $\pi(x) = \Pr(b(x)|B, x) = \Pr(a(x)|A, x)$. This could, however, be relaxed without affecting the results. Let signals also be at least weakly informative in the sense that $a$-signals are (weakly) more likely than $b$-signals in state $A$ and vice versa. After observing the signal, the agent decides on a second action from another finite space $X_2$ and receives a payoff determined by the action profile as well as the state of the world. Let the probability that state $A$ materializes

be $\Pr(\omega = A) = p \in (0,1)$. Denote the agent's prior belief that the state is $A$ by $\mu$. This belief may or may not coincide with the correct probability $p$.

Let $u(x, y|\omega) \in \mathbb{R}$ denote the payoff if $x$ is taken at $t = 1$, $y$ at $t = 2$, and the state is $\omega$. We assume that payoffs are bounded. To avoid trivial scenarios, we implicitly assume that in each state there are unique best actions $x$ and $y$, and actions are not generally payoff equivalent. Furthermore, ties are broken deterministically. As the first action not just directly affects payoffs but also (possibly) reveals information about the realized state, we view it as a type of experiment. The agent can react to the signal and adjust the action in the second period. Notice, however, that an experiment is only useful if it is not too costly. For instance, $X_1$ may contain a choice that perfectly reveals the state but also reduces the attainable utility to an extent that it is better to choose a noisier experiment.

We call $\mathbf{x} = (x, \{x_a, x_b\})$ an *action profile* where $x \in X_1$ is the experiment and $x_a, x_b \in X_2$ represent the actions taken after an $a$- and $b$-signal. We denote the set of all such action profiles $\mathbf{X}^*$.[2] To simplify later notation, we exclude trivially suboptimal choices by restricting $\mathbf{X}^*$ to profiles where any action taken after an $a$-signal has a weakly greater payoff in state $A$ than the action taken after a $b$-signal and vice versa.[3] An action profile is said to be *signal sensitive* if $x_a \neq x_b$. It is called *simple* if the second-period choice is not conditional on the signal, $x_a = x_b$. For brevity, simple profiles are denoted by $(x, y)$, with $x \in X_1$ and $y \in X_2$. A particular profile is said to be *chosen* at some belief $\mu$ if it maximizes expected utility at that belief.

# 4. UNBIASED CHOICE PROBLEM

## 4.1. PERIOD 2 CUTOFF-STRATEGY

Reverse-engineering the agent's decision problem, we first look at optimal actions in the second period. The expected utility of an action $y \in X_2$ from the second period's perspective depends on the posterior after receiving a signal from the experiment in the first period. Furthermore, it can also depend directly on the previous action. Let $\mu(s)$ be the posterior belief after receiving signal $s$ computed according to Bayes' rule. At $t = 2$, the agent chooses an action that maximizes their expected utility given that posterior and the

[2] While technically fully contingent action profiles lie in $X_1 \times S \times X_2$, since every action results in distinct $a$- or a $b$-signals, we can reduce this space to $X_1 \times \{a, b\} \times X_2$.

[3] This does, of course, not exclude the possibility that there are $y, y' \in X_2$ where $y$ achieves a higher payoff than $y'$ in both states and both are part of some action profile in $\mathbf{X}^*$. However, they cannot be part of the *same* action profile.

previous action $x$:

$$\max_{y \in X_2} \quad E\left[u(x,y) \mid \mu(s)\right]$$

The expected payoff from any action in period 2 is monotonic in beliefs. Hence, actions in $X_2$ can be ordered according to their expected payoffs based on $\mu(s)$. This gives rise to a cutoff-type decision rule which is shown formally in Lemma 1 in the Appendix. To briefly illustrate the argument, let's looks at the expected payoffs of two actions $y, z \in X_2$ given some $x \in X_1$ and posterior $\mu(s)$:

$$E[u(x,y) \mid \mu(s)] = \mu(s) \cdot u(x,y|A) + \big(1-\mu(s)\big) \cdot u(x,y|B)$$
$$E[u(x,z) \mid \mu(s)] = \mu(s) \cdot u(x,z|A) + \big(1-\mu(s)\big) \cdot u(x,z|B)$$

If one of the two actions is strictly better in both states, i.e. $u(x,y|\omega) > u(x,z|\omega)$ for both $\omega \in \{A,B\}$, it is strictly preferred for all beliefs. If instead, $u(x,y|A) < u(x,z|A)$ and $u(x,y|B) > u(x,z|B)$, then there exists some $\mu^*(s) \in (0,1)$ such that for all $\mu(s) > \mu^*(s)$, $z$ is strictly preferred to $y$ and vice versa for $\mu(s) < \mu^*(s)$. Iterating this argument over all available actions allows us to conclude that the region in which a given action is chosen must be an interval.

**Result 1:** *For every $x \in X_1$, there exists a partition $\mathscr{P}_x$ of $[0,1]$ such that for every two consecutive elements $p_i$ and $p_{i+1}$ of the partition, there is an action $y_i \in X_2$ such that $E[u(x,y_i)|\mu] \geq E[u(x,z)|\mu]$ for all $z \in X_2$ and $p_i < \mu < p_{i+1}$.*

The second period choice depends on the posterior, which is determined by the signal. Result 1 implies that differences in the posterior are only welfare relevant if they fall into different elements of the partition. For binary signals, there are at most two choices resulting in potentially four different outcomes. Since those are pinned down for every $x \in X_1$ by $\mathscr{P}_x$, we can collapse the problem to a comparison of experiments in period 1, fixing the corresponding optimal period-2 choices.

## 4.2. CHOICE IN PERIOD 1

The optimal choice in period 1 balances the information value of an experiment with the utility derived from it directly. A very informative action leads to very different posteriors and, keeping in mind the partition structure, to different actions in period 2. When an action has little information value, the posterior is close to the prior $\mu$ and might trigger the same course of action independent of the outcome. We can write the objective function

for the utility maximization problem in period 1 as:

$$\begin{aligned} &\mu \cdot \Big[\pi(x)\,u(x, x_a^*|A) + \big(1-\pi(x)\big)\,u(x, x_b^*|A)\Big] \\ +(1-\mu) \cdot &\Big[\big(1-\pi(x)\big)\,u(x, x_a^*|B) + \pi(x)\,u(x, x_b^*|B)\Big] \end{aligned} \tag{1}$$

where $\{x_a^*, x_b^*\}$ are the optimal period-2 actions for $a$ and $b$-signals given the choice $x \in X_1$. It is a weighted average of receiving the 'correct' signal and thus choosing the correct action, and the probability of receiving the 'incorrect' signal and choosing the action that yields the lower utility in the realized state. A higher informativeness in the sense of the likelihood-ratio $\frac{\pi(x)}{1-\pi(x)}$ reduces the likelihood of such a mistake. For a high informativeness, an agent might be confident enough to choose actions that have a higher payoff variation between states. We finish this section with a key property of the unbiased agent's problem. Some of the later results arise from a violation of it.

**Result 2:** *The maximum expected utility at* $t = 1$ *is convex in* $\mu$.

**Example 1 (Diagnostic Testing):** Suppose a patient has potentially been exposed to an infectious disease. The patient (or alternatively the physician) can immediately start treatment (action $x_A$), not take any treatment and continue as usual ($x_B$), or take a test to determine whether they have been infected ($x_I$) and then either seek treatment, take another test or continue as normal. Hence, there are three actions in each period, $\{x_A, x_I, x_B\} = X_1 = X_2$. $x_I$ provides an informative signal, $\pi(x_I) = 0.75$, while the other two actions yield uninformative ones, $\pi(x_A) = \pi(x_B) = 0.5$. Suppose $u(x_A, x_A|A) = 5 = u(x_B, x_B|B)$, $u(x_I, x_A|A) = u(x_I, x_B|B) = 4$, with all other combinations having a utility of zero. $\mu$ describes the patient's probability assessment of having been infected, i.e. the true state being $A$.

Comparing the different payoffs, we can see that it is never optimal to choose a combination of $x_A$ and $x_B$. Furthermore, $x_I$ represents a pure information experiment that is only useful because it indicates the true state and thus the appropriate action at $t = 2$. If the patient takes the test in period 1, then they will choose between $x_A$ and $x_B$ in period 2 depending on whether the posterior is greater or smaller than $\frac{1}{2}$. Figure 1 illustrates the expected utility outcomes. Action profiles $(x_A, x_A)$ and $(x_B, x_B)$ are optimal for more extreme beliefs. If the patient is convinced that they have been infected, it is best to start treatment without delay even though there might be a risk of unnecessary treatment with possible adverse side effects. Equivalently, if the risk is very low, it is best to continue as usual and thereby rule out the possibility of receiving unnecessary treatment. For inter-

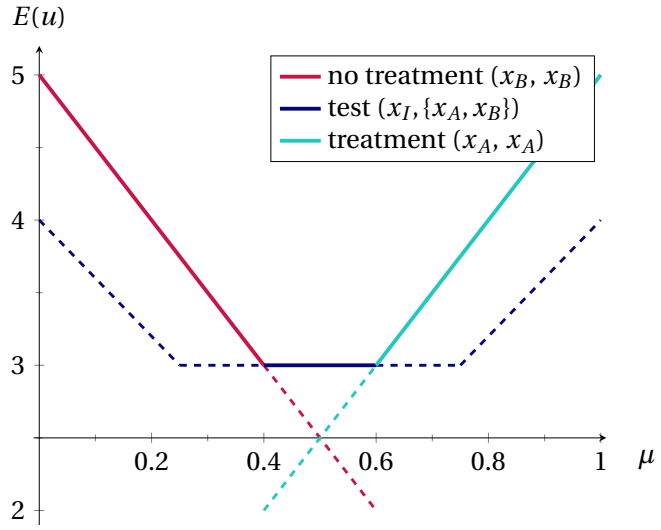


Figure 1: Example 1 expected utility outcomes for different action profiles

mediate beliefs, the informativeness of $x_I$ is valuable and the posterior is such that either $x_A$ or $x_B$ is chosen at $t = 2$ conditional on the signal. The patient can quarantine themselves and then act according to the test result. As there is delay while the patient awaits the test result, payoffs are lower in either state but there is also a lower risk of committing a mistake. As stated in Result 2, the maximum expected utility (bold line segments) is convex in $\mu$.

## 5. IMPERFECT INFORMATION PROCESSING

We now turn to our main area of investigation: scenarios in which decision-makers may fail to correctly process all available information. In particular, they might not always perceive signals accurately. We call this a perception bias or simply *misperception*. Furthermore, agents might have a prior that is deviating from the true probability distribution over states. We call this an incorrect belief or a *bias in prior*.

Misperception adds noise to the signal; a form of Blackwell garbling. It weakens the correlation between the signal and the state and thus reduces the information value of experiments. The misperception can arise from systematic errors in interpreting information or from some random mistakes. Initially, we take no stand regarding the source and exact form of the bias. Unless stated otherwise, we make the crucial assumption that agents are *naive* in the sense that they are unaware of this perception problem. They compute their posterior and choose actions as if there were no misperception issues.

Naive agents therefore arrive at an incorrect posterior following the initial signal and may choose suboptimal actions in the second period. By not recognizing the reduced informativeness of experiments, naive agents tend to overvalue information in the first period. This may result in too much information acquisition. In contrast, *sophisticated* agents are fully aware of their misperception problem. While they cannot undo their actual bias, they recognize that information is less informative and adjust their posterior beliefs, second and first period actions accordingly.

A bias in prior refers to an agent's ex-ante belief $\mu$ that differs from the true $p$. We can interpret the 'correct' prior as the probability assessment that a fully rational observer holding all previously available information would reach. An agent who did not observe all information or misinterpreted some of it in the past - possibly through misperception - could have reached a different conclusion. We do not distinguish between a sophisticated and naive agent with regards to a bias in prior. After all, an agent who is aware that their prior is biased will simply adjust it to remove the bias. A bias in prior can lead to a suboptimal choice of experiment in period 1 with potential consequences for the subsequent choices. For all but fully informative signals, the bias will also continue to distort second period beliefs and hence second period actions.

To illustrate the setting as well as the subsequent results, consider the following thought experiment in the spirit of Example 1: a patient consults a physician regarding the possibility of having contracted a disease. The doctor can order a medical test for which the outcome can be either negative or positive. On top of any inaccuracies of the test itself, suppose there is a certain chance the lab technician performing the test enters the incorrect result in the patient's file. This results in an information loss and increases the probability of false positives and negatives. For example, the technician may enter a positive result even though the test came back negative. Moreover, the doctor cannot undo the technician's mistake since they do not know whether a mistake occurred in the first place. The distortion does not have to be balanced but is independent of the state of the world; the lab technician does not have any knowledge about the truth other than through the test result. If the doctor is unaware of their technician's potential mistake, we refer to them as naive. If they take the mistakes into account, we call them sophisticated. As mentioned before, we mostly maintain the assumption of a naive doctor.[4] In the same context, an incorrect prior corresponds to the doctor putting a higher (or lower) probability on the

[4] Whiting et al. (2015) provides a review of how accurately diagnostic tests are interpreted by health professionals. They find evidence for clinicians misinterpreting (and especially overstating) the accuracy of diagnostic tests and hence updating beliefs too much based on test outcomes.

patient having a certain condition than is warranted by the information given. This can, for example, arise if a patient fails to disclose some relevant information in their medical history. This might then lead to an unnecessary test or even unnecessary treatment or failure to conduct a necessary one.[5]

Let $\tilde{s} \in S$ be the signal received by the technician conducting the experiment, and $s \in S$ the signal observed by the doctor. Then for an experiment $x \in X_1$:

$$\Pr(s, \tilde{s}|\omega, x, \mu) = \Pr(\tilde{s}|\omega, x) \cdot \Pr(s|\tilde{s}, x, \mu) \qquad \forall \omega \in \{A, B\}$$

Without any errors of the lab technician, the doctor always observes the correct signal meaning $\Pr(s|\tilde{s} = s, x, \mu) = 1$. But the more likely a mistake by the lab technician, the lower the chance the doctor receives the correct result. Any such error cannot, however, depend on $\omega$ as the true state of the world is unknown to the technician. While it is possible and even likely that the misperception depends on the prior $\mu$, for the rest of the analysis we take misperception to be constant across $\mu$. This mainly serves to keep the notation and explanations shorter. Our result go through without this assumption. We can thus define the probability of correctly transmitting a signal $s$ as a function of experiment $x \in X_1$ alone:

$$k_s(x) = \Pr(s|\tilde{s} = s, x) \quad \forall s \in S(x)$$

Subsequently we omit the argument $x$ whenever there is no confusion to which action a particular $k_s$ refers to. The distorted probabilities of an agent actually observing an $a$-signal in state $A$ and $b$-signal in state $B$ for an experiment $x \in X_1$ can be written as a linear function of $k_a$, $k_b$, and the undistorted probabilities of receiving each signal: [6]

$$\begin{aligned} \hat{\pi}_A(x) &= k_a \cdot \pi + (1 - k_b) \cdot (1 - \pi) \\ \hat{\pi}_B(x) &= k_b \cdot \pi + (1 - k_a) \cdot (1 - \pi) \end{aligned} \tag{2}$$

As the misperception is not necessarily balanced across signals, the signal probabilities might no longer be symmetric. The probability $\hat{\pi}_A$ of observing an $a$-signal in state $A$ does not necessarily equal the probability $\hat{\pi}_B$ of observing a $b$-signal in state $B$. We do, however,

[5] See Epner et al. (2013) for a discussion of diagnostic errors in relation to laboratory testing, especially through ordering of inappropriate tests, failure to order necessary tests, failure to correctly interpret results, and laboratory mistakes. Furthermore, there is evidence that misperception occurs directly at the physician, as test results are difficult to interpret and many junior doctors report insufficient training regarding their interpretation (Freedman (2015)).

[6] Note that the underlying probabilities $\pi$ and $1 - \pi$ can only enter linearly as otherwise misperception would be implicitly state dependent.

maintain the assumption that each type of signal remains at least weakly informative of the associated state.

**Assumption 1** (No signal switching)**:** *Any perception bias does not reverse the correlation between signals and states meaning that for all* $x \in X_1$ *and* $\omega, \omega' \in \{A, B\}$*:*

$$\frac{\hat{\pi}_\omega(x)}{1-\hat{\pi}_{\omega'}(x)} \geq \frac{1}{2}$$

In other words, we exclude cases where agents fundamentally misunderstand correlations in their environment. For example, suppose to the contrary that an agent misperceives every signal so that $k_a = k_b = 0$. This does not lower informativeness. Simply relabeling $a$-signals as $b$ and vice versa would achieve the same outcome as for the perfectly accurate case where $k_a = k_b = 1$. We instead focus on cases of misperception that reduce the informativeness of an experiment and cannot be overcome or improved upon by a mere re-labeling.

An experiment can be characterized by a $2 \times 2$ Markov matrix $P$, where element $p_{ij}$ refers to the probability of receiving a signal $s_j$ in state $\omega_i$ where $j = 1$ refers to an $a$-signal and $j = 2$ to a $b$-signal and equivalently for $i$. Denote the matrix associated with the unbiased information experiment arising from $x \in X_1$ by:

$$P_x = \begin{bmatrix} \pi(x) & 1-\pi(x) \\ 1-\pi(x) & \pi(x) \end{bmatrix}$$

The agent's misperception can equally be represented by a Markov matrix $M_x$:

$$M_x = \begin{bmatrix} k_a(x) & 1-k_a(x) \\ 1-k_b(x) & k_b(x) \end{bmatrix}$$

Multiplying the original experiment $P_x$ with the misperception matrix $M_x$ yields the probability structure of the distorted experiment:

$$P_x M_x = \begin{bmatrix} \hat{\pi}_A(x) & 1-\hat{\pi}_A(x) \\ 1-\hat{\pi}_B(x) & \hat{\pi}_B(x) \end{bmatrix}$$

When comparing expected utilities between information experiments, it is useful to consider three cases: the expected utility from an undistorted experiment, the expected utility from a distorted experiment where the agent is sophisticated taking into account the

distortion, and finally the case of a naive agent unaware of the distortion. Denote the expected utility of the experiment arising from $x \in X_1$ for a prior $\mu$ without distortion by:

$$\begin{aligned} V(P_x|\mu) \equiv \quad & \mu \cdot \Big[\pi(x)\,u(x, x_a^*|A) + \big(1-\pi(x)\big)\,u(x, x_b^*|A)\Big] \\ & +(1-\mu) \cdot \Big[\big(1-\pi(x)\big)\,u(x, x_a^*|B) + \pi(x)\,u(x, x_b^*|B)\Big] \end{aligned} \tag{3}$$

where $x_i^* \in X_2$ are the optimal choices for the respective signals given $x$, meaning they are maximizers of the above expression. The expected utility for the same experiment but with a misperception matrix $M_x$ is denoted as:

$$\begin{aligned} V(P_x M_x|\mu) \equiv \quad & \mu \cdot \Big[\hat{\pi}_A(x)\,u(x, x_a^{**}|A) + \big(1-\hat{\pi}_A(x)\big)\,u(x, x_b^{**}|A)\Big] \\ & +(1-\mu) \cdot \Big[\big(1-\hat{\pi}_B(x)\big)\,u(x, x_a^{**}|B) + \hat{\pi}_B(x)\,u(x, x_b^{**}|B)\Big] \end{aligned} \tag{4}$$

where again $x_i^{**}$ are the maximizers of the above expression. $V(P_x M_x|\mu)$ thus represents the expected utility from the information experiment $P_x M_x$. This implies the agent is aware of the distortions and chooses accordingly. The expected utility of a naive agent not aware of the distortions is denoted as:

$$\begin{aligned} V_n(P_x M_x|\mu) \equiv \quad & \mu \cdot \Big[\hat{\pi}_A(x)\,u(x, x_a^*|A) + \big(1-\hat{\pi}_A(x)\big)\,u(x, x_b^*|A)\Big] \\ & +(1-\mu) \cdot \Big[\big(1-\hat{\pi}_B(x)\big)\,u(x, x_a^*|B) + \hat{\pi}_B(x)\,u(x, x_b^*|B)\Big] \end{aligned} \tag{5}$$

where $x_i^*$ are identical to the ones in Equation 3 and therefore not necessarily the maximizers of the above expression. In other words, a naive agent chooses as if the experiment $x$ had the associated matrix $P_x$ even though it effectively is $P_x M_x$.

Finally, it will often be useful to consider the expected utility of a particular profile where period-2 actions are fixed and not necessarily optimally chosen among all of $X_2$. For any $\mathbf{x} = (x, \{x_a, x_b\}) \in \mathbf{X}^*$, we denote this as:

$$\begin{aligned} V(P_x M_x|\,\mathbf{x}, \mu) \equiv \quad & \mu \cdot \Big[\hat{\pi}_A(x)\,u(x, x_a|A) + \big(1-\hat{\pi}_A(x)\big)\,u(x, x_b|A)\Big] \\ & +(1-\mu) \cdot \Big[\big(1-\hat{\pi}_B(x)\big)\,u(x, x_a|B) + \hat{\pi}_B(x)\,u(x, x_b|B)\Big] \end{aligned} \tag{6}$$

where the only maximization is that both actions are optimal after the associated signal among $(x_a, x_b)$ meaning $x_a$ achieves weakly higher utility in state $A$ than $x_b$ and vice versa. As we excluded trivially suboptimal profiles, we know that for any given $\mathbf{x} \in \mathbf{X}^*$, $x_a$ does not dominate $x_b$ in both states and so each action is taken after one of the signals. This is,

for example, necessarily the case if period-2 actions are optimal given $\mu$ or $\mathbf{x}$ is chosen at $\mu$. In this case, $V(P_x|\mu) = V(P_x|\,\mathbf{x},\mu)$ and $V_n(P_xM_x|\mu) = V(P_xM_x|\,\mathbf{x},\mu)$ for any $M_x$.

# 6. BIASES AND THEIR IMPLICATIONS

## 6.1. MISPERCEPTION

We start with the observation that a perception bias always makes an agent worse off.

**Proposition 1:** *For any $\mu \in [0,1]$ and any action $x \in X_1$ with associated $P_x$ and misperception $M_x \neq I$, misperception reduces expected utility and (weakly) more so for naive than sophisticated agents:*

$$V(P_x|\mu) \geq V(P_xM_x|\mu) \geq V_n(P_xM_x|\mu)$$

*with the first inequality strict when the optimal profile $\mathbf{x} \in X^*$ conditional on $x$ is signal sensitive.*

The first inequality shows the welfare effect for sophisticated agents, which follows immediately from the well-known Blackwell (1951) result on information experiments. Misperception is a form of signal garbling and an agent cannot be better off with a garbled signal as otherwise the agent could simply take the original signal and garble it equivalently. Indeed, for signal sensitive profiles, any change in the informativeness of a signal has strictly negative welfare consequences.

The second inequality states that for the same perception error, a naive agent is (weakly) worse off than a sophisticated agent. For a given experiment, both the sophisticated and the naive agent receive the same signals with the same likelihoods. But for each signal, the sophisticated agent is aware of the lower information value, which allows them to make better decisions if a better choice is available. Whether or not this inequality is strict thus depends on the availability of alternatives. As will be shown later, for signal sensitive profiles there always exists a perception bias that leaves a naive agent strictly worse-off than a sophisticated one.

Result 3 further shows a monotonicity in the relation between welfare loss and misperception: the welfare loss for a given experiment is, at least for naive agents, strictly increasing in the perception bias. We can draw a similar conclusion for sophisticated agents but their welfare loss is capped at the point where it becomes beneficial to disregard any signals generated from the experiment.

We say that the magnitude of misperception increases if $k_a$, $k_b$, or both, decrease, which means the probability of receiving a correct signal decreases.

**Result 3:** *For any signal sensitive profile* $\mathbf{x} \in \mathbf{X}^*$ *chosen at* $\mu \in (0,1)$*, the welfare loss* $V_n(P_x|\mu) - V_n(P_x M_x|\mu)$ *of naive agents from any misperception* $M_x$ *is strictly increasing in the magnitude of misperception.*

Any perception bias is bad news for an agent who conditions their choices on signals. Misperception 'switches' signals and therefore leads to mistakes in period-2 choices. Furthermore, as misperception increases, the likelihood-ratios $\frac{\hat{\pi}_A}{1-\hat{\pi}_B}$ and $\frac{\hat{\pi}_B}{1-\hat{\pi}_A}$ decline.[7] A decision-maker should be updating less after any signal and therefore potentially choose different period-2 actions as the posterior might fall into different intervals of the partition. But only a sophisticated agent makes such an adjustment. Using the earlier example, we can illustrate the welfare consequences of misperception for a signal sensitive action profile:

**Example 1 - continued:** Recall that an agent possibly exposed to an infectious disease can take a diagnostic test (choose $x_I$), and/or decide on the appropriate treatment ($x_A$ or $x_B$). Suppose now that, without the agent being aware of this, the chance of receiving a false negative result is higher than before. This is equivalent to an agent perceiving $b(x_I)$ too often, i.e. $k_a(x_I) = 1-\delta$ while keeping $k_b(x_I) = 1$. Figure 2 plots the utility frontier for $\delta = 0.2$ which increases the likelihood of perceiving a $b$-signal in state $A$ from $1-\pi(x_I)$ to $1-\pi(x_I)+0.2\cdot\pi(x_I)$ and in state $B$ from $\pi(x_I)$ to $\pi(x_I)+0.2\cdot(1-\pi(x_I))$. The utility frontier as shown by the solid lines in Figure 2, is now strictly lower for any interval in which $x_I$ is chosen and identical otherwise. If the agent's prior is sufficiently extreme, they either take or forgo treatment and don't receive additional information. When taking a test (choosing $x_I$), however, the agent receives additional information. If the agent misjudges the accuracy by, for example, underestimating the probability of false negative results, they put too little probability on having contracted the disease after receiving a negative result. [8] This lowers their expected utility from taking the test: $V(P_I|\mu) > V(P_I M_I|\mu) \geq V_n(P_I M_I|\mu)$. Furthermore, for $\mu$ around 0.6, while it is still optimal to condition actions based on the test result when taking a test, the test itself in period 1 is not optimal. Thus $V(P_A|\mu) > V(P_I M_I|\mu) = V_n(P_I M_I|\mu)$ and similarly for $\mu$ near 0.4. A

[7] See Result A.1 in the Appendix for a formal proof.

[8] Blastland et al. (2020) provide some recent observations for such behavior, particularly when information about accuracy is missing.

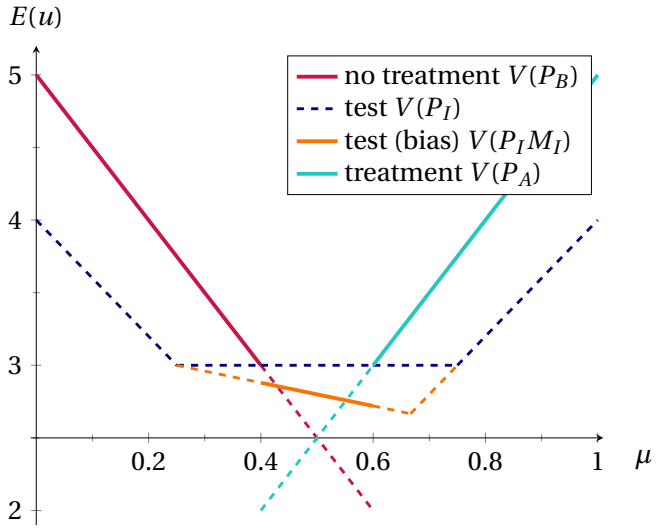


Figure 2: Non-convex utility frontier

sophisticated agent can adjust by not taking the test for a wider range of $\mu$, while a naive agent is unaware of the misperception problem and therefore doesn't adjust their choices which leads to a strictly greater welfare loss when a better alternative choice is available.

The example illustrates how misperception reduces the informativeness of experiments and thereby distorts their expected utility ranking. How severely an experiment is affected depends on how susceptible it is to misperception[9] and how valuable signals are for the action profile involving it. These differences can lead to reversals in the expected utility ranking of profiles. Whether or not this is possible thus depends on the experiments themselves as well as on how misperception differs across experiments. To systematically characterize when such reversals can occur, we analyze four instructive classes of misperception: misperception is (1) (possibly) independent across experiments, (2) the same across all experiments, (3) the same for both signal types, and (4) the same across all experiments and signal types. For those cases, we present necessary and sufficient conditions for when utility reversals can occur that rely only on a (simple) comparison of action profiles. As will be shown, these conditions nest according to the restrictions imposed on misperception and are key for characterizing the interaction between biases in prior and misperception.

[9]Section 8 briefly discusses how this can be interpreted as a form of ambiguity over the information content of signals and how our analysis fits into this interpretation.

### 6.1.1. UNRESTRICTED MISPERCEPTION

To characterize the effects of misperception in the general case without any restrictions on misperception across experiments and signal types, we introduce the following relation between profiles:

**Definition 1** (WCD)**:** *An action profile* $\mathbf{y} = (y, \{y_a, y_b\}) \in \mathbf{X}^*$ ***worst-case dominates*** $\mathbf{x} = (x, \{x_a, x_b\}) \in \mathbf{X}^*$ *at* $\mu \in (0,1)$ *if*

$$V(P_y|\,\mathbf{y}, \mu) > \min_{s\in\{a,b\}} \mu u(x, x_s)|A) + (1-\mu)\, u(x, x_s|B).$$

Worst-case dominance compares an action profile $\mathbf{y}$ to the (at most two) simple profiles that can be constructed from $\mathbf{x}$. If $\mathbf{y}$ yields higher expected utility than the worst of these simple profiles evaluated at the prior, then $\mathbf{y}$ worst-case dominates $\mathbf{x}$. In other words, we treat $\mathbf{x}$ as if it has no information value and signals only determine which action is chosen at $t = 2$. We then ask which of the choices in the contingent period-2 plan $\{x_a, x_b\}$ yields lower utility from an ex-ante point of view and compare it to the (regular) expected utility of $\mathbf{y}$. Note that $\mathbf{y}$ may or may not achieve higher expected utility than $\mathbf{x}$ when worst-case dominance holds. The usefulness of this definition lies in the fact it enables us to identify those action profiles whose ranking can be reversed through misperception, simply by looking at the worst course of action that can occur with this profile:

**Proposition 2:** *For any two action profiles* $\mathbf{x}, \mathbf{y} \in \mathbf{X}^*$*, there exist misperception matrices* $M_x$ *and* $M_y$ *such that* $V(P_y M_y|\,\mathbf{y}, \mu) > V(P_x M_x|\,\mathbf{x}, \mu)$ *if and only if* $\mathbf{y}$ *worst-case dominates* $\mathbf{x}$ *at* $\mu$*.*

Proposition 2 implies that if an action profile $\mathbf{x}$ is optimal but worst-case dominated at some belief, we can find a distortion such that an agent will be strictly better off choosing another profile $\mathbf{y}$. This is exactly the case for which all inequalities in Proposition 1 hold strictly. A sophisticated agent adjusts their choices to $\mathbf{y}$ to mitigate the downside-risk from misperception. A naive agent, instead, generally commit two mistakes: First, they fail to optimally use the information they do receive. By overstating the signal's accuracy, the choices after receiving the signal tend to be sub-optimal. Second, they fail to adjust the first period choice, i.e. acquire better information (accounting for misperception) or opt for a safer alternative less reliant on information. If a better profile is available given the misperception - which can only happen if some profile worst-case dominates the one chosen by the naive agent - the naive agent is strictly worse-off than a sophisticated one as $V(P_y M_y|\mu) > V(P_x M_x|\,\mathbf{x}, \mu) = V_n(P_x M_x|\mu)$.

Of course, to reverse the utility ranking of two profiles, a specific type of misperception that affects one experiment much more than another might be required.[10] While this might be difficult to justify in some setting, it applies to situations when a particular experiment yields much more ambiguous and harder-to-interpret results than another. Alternatively, the agent may be biased against one information source and hence more likely to misperceive its signal. We will consider the case where all signals are equally distorted later.

An extreme way to avoid the effects from misperception is to ignore any information arising from one's actions or take actions that don't yield informative signals. While Proposition 2 tells us that this would be optimal if an action is worst-case dominated and misperception is too severe, it does not tell us whether such a dominating profile exists. As it turns out, this possibility exists for any signal sensitive action profile and prior, except for the knife-edge case where the prior makes the agent just indifferent between their period-2 actions. Consequently, a naive agent can be harmed by the availability of an informative experiment.

**Result 4:** *For every signal sensitive profile* $\mathbf{x}$ *and almost every* $\mu$*, there exists a misperception matrix* $M_x$ *and a simple profile* $\mathbf{y}$ *such that* $V(P_y|\mu) > V_n(P_x M_x|\mu)$.

In terms of the previous thought experiment, Result 4 shows that exactly in those cases when a (naive) physician prefers to condition the treatment on the test result, there is a mistake the lab technician could commit such that this becomes a suboptimal course of action. If the lab technician performing the test is very prone to errors, a physician should treat the patient based on their initial assessment rather than condition treatment on the test.

### 6.1.2. Equal misperception across experiments

From the previous discussion, we can conclude that there always exists some misperception that sufficiently distorts an experiment to make a given signal-sensitive profile suboptimal. By extending the result to a more restrictive setting where all experiments are distorted equally, we show that this does not rely on overly distorting only one experiment.

In the context of our thought experiment, the lab technician might still enter incorrect results but the tendency to make such a mistake is now equalized across tests rather than

[10] More precisely, the signals $S(x)$ of experiment $x \in X_1$ might have to be misperceived with much higher probability than the signals $S(y)$ for $y \in X_1$ for a reversal in expected utility.

one test being more error prone. It is important to note that mistakes may still depend on the actual test result so that, for instance, $a$-signals can be misperceived with higher probability than $b$-signals. In a later section, we also consider the additional restriction that the test results themselves are misperceived equally.

**Definition 2** (strong WCD)**:** *An action profile* $\mathbf{y} = (y, \{y_a, y_b\}) \in \mathbf{X}^*$ *strongly worst-case dominates an action profile* $\mathbf{x} = (x, \{x_a, x_b\}) \in \mathbf{X}^*$ *at* $\mu \in (0,1)$ *if*

$$\mu u(x, x_a)|A) + (1-\mu)u(x, x_a|B) < \mu u(y, y_a)|A) + (1-\mu)u(y, y_a|B), \text{ or}$$
$$\mu u(x, x_b)|A) + (1-\mu)u(x, x_b|B) < \mu u(y, y_b)|A) + (1-\mu)u(y, y_b|B).$$

While worst-case dominance compares the expected utility of one action profile with the expected utility of the worst simple profile generated from another, strong worst-case dominance breaks both profiles down into their simple profiles and compares their expected utility individually. It can be shown that strong worst-case dominance implies worst-case dominance (Appendix, Result A.2) and is, as the name suggest, a strengthening of the previous notion. Returning to our example, strong worst-case dominance compares two contingent treatment plans by individually comparing the courses of action that occur after each signal. For example, depending on the test result, a patient might either take the test and start treatment or take the test and continue without treatment. Strong worst-case dominance compares these outcomes ignoring any information value from the test. As Proposition 3 demonstrates, if one treatment plan performs worse in expectation for one of the two possible test results, we can find a type of mistake a lab technician could commit equally across tests so that the otherwise optimal course of action becomes inferior.

**Proposition 3:** *For any two action profiles* $\mathbf{x}, \mathbf{y} \in \mathbf{X}^*$ *with* $V(P_x|\,\mathbf{x}, \mu) > V(P_y|\,\mathbf{y}, \mu)$*, there exist a misperception matrix* $M_x = M_y = M$ *such that* $V(P_y M|\,\mathbf{y}, \mu) > V(P_x M|\,\mathbf{x}, \mu)$ *if and only if* $\mathbf{y}$ *strongly worst-case dominates* $\mathbf{x}$ *at* $\mu$*.*

Result 4, which states there always is a type of misperception such that an agent would be better off from an uninformative action, goes through as is, noting that when comparing signal sensitive to a simple profiles, worst-case dominance is equivalent to strong worst-case dominance.

## 6.2. BIASED PRIOR

We now turn to the second source of suboptimal choice: an incorrect prior. As discussed previously, we see $p$ as the probability assessment that a fully rational observer taking into account all available information would arrive at. If $\mu = p$, the agent will - at least from the perspective of such an observer - choose the optimal action profile. If, however, $\mu \neq p$, then the true expected utility does not necessarily match the one evaluated at $\mu$. Equation (1) illustrates how a different $\mu$ can lead to different choices. An inaccurate $\mu$ puts too much weight on one of the states and thus favours actions appropriate for that state. As the state realizes with a different probability $p$, the choice might be suboptimal.

**Result 5:** *For any $\mu \geq p$, expected utility is (weakly) decreasing in $\mu$. Equivalently, for $\mu \leq p$ expected utility is (weakly) increasing in $\mu$.*

Small deviations from $p$ generally remain without consequences when actions are discrete as the same action profile remains optimal. In the context of the thought experiment, a physician holding a belief $\mu$ close to $p$ would still order the same test and administer the same course of treatment in response. If the difference is more pronounced, the physician might, for instance, immediately order treatment without waiting for a test result. This could negatively affect the expected outcome if the true probability is rather different from $\mu$.

## 6.3. INTERACTION OF BIASES

So far, we have seen that in isolation misperception causes agents to more frequently take the wrong course of action in response to their signals. The informativeness of experiments is reduced and thus naive agents are too confident in their observed signals. In turn, this may cause them to choose effectively inferior actions; something that can occur if and only if worst-case dominance holds (or strong worst-case dominance in the case of symmetric misperception). While not affecting the informativeness of actions, a bias in prior similarly leads to inferior choices. One might then expect that the interaction of both reduces expected utility even further, a 'double whammy'. After all, we might place little confidence in a decision-maker that holds far-out beliefs and misconstrues empirical evidence. However, it turns out that this is not always justified. To the contrary, a naive agent with both misperception and a biased prior can be better off than an agent who only suffers from one of the two. The fundamental reason behind this is that the mispercep-

tion shuffles the otherwise straight-forward ranking of experiments at different priors to a different degree.

To better illustrate these interactions, we break the analysis down into two parts by in turn holding constant one of the biases and increasing the other. For a naive agent, holding constant the bias in prior but varying the misperception fixes the choice of action profile while varying the signal strength and hence the utility ranking of choices. In contrast, holding constant the misperception and varying the bias in prior fixes the utility ranking of choices while varying the choice of action profile. Although ultimately symmetric, disentangling the effects demonstrates the influence of either type of bias.

#### 6.3.1. ADDING A BIASED PRIOR TO MISPERCEPTION

In this section, we assume that an agent's perception is biased 'from the start' and characterize the welfare effects of adding a bias in prior on top of the misperception; both when misperception is specific to the experiment as well when all experiments are affected equally. Returning to the thought experiment, we ask whether a patient could be better off being treated by a physician that has a biased initial assessment compared to one with an unbiased view, given that the lab technician commits mistakes unbeknownst to the physician. Or alternatively, a somewhat more sinister interpretation: knowing that a decision-maker suffers from misperception, could we provide them with deliberately misleading information to improve outcomes? Could a patient benefit from giving an inaccurate medical history and thus affecting the physician's initial assessment? We first present an example and then generalize this insight.

**Example 1 - continued:** If the agent is unaware of the possibility that signals might be misperceived, the utility frontier is not convex in $\mu$. This arises due to the merely perceived indifference between profiles $(x_B, x_B)$ and $(x_I, \{x_A, x_B\})$ at $\mu = 0.4$ (both achieve an expected utility of 3) and equally between $(x_A, x_A)$ and $(x_I, \{x_A, x_B\})$ at $\mu = 0.6$. At these beliefs, $(x_I, \{x_A, x_B\})$ is (strongly) worst-case dominated by either $(x_B, x_B)$ or $(x_A, x_A)$. The introduction of misperception $M_I$ thus has the potential to influence the optimality of choices. At $\mu = 0.4$, the 'true' expected utility from $x_I$ given the misperception is only 2.88. As $x_I$ yields the $b$-signal with a high probability and $B$ occurs with a relatively high chance (1-$\mu = \frac{3}{5}$) anyway, the agent would be strictly better off taking action profile $(x_B, x_B)$, avoiding the cost of the test. Similarly, at $\mu = 0.6$, the true expected utility from $x_I$ is only 2.72 as the agent perceives $b(x_I)$ signals too often. Without being aware

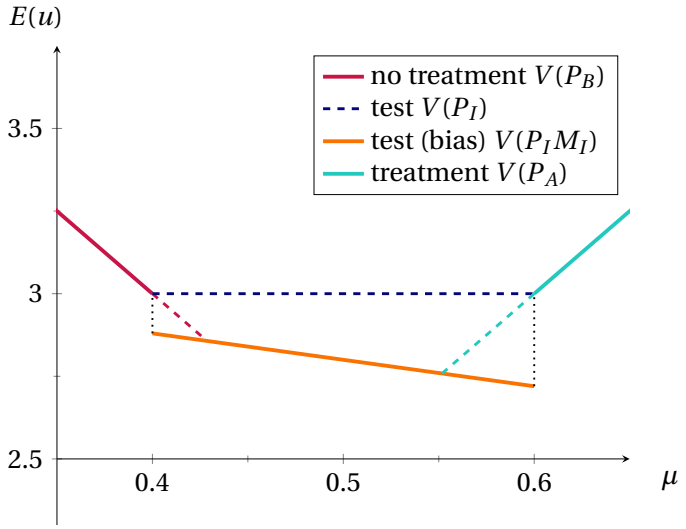


Figure 3: Welfare effects of misperception

of it, the naive agent has a higher chance of not taking any treatment ($x_B$) in a situation in which they actually contracted the disease (state $A$). Rather than being indifferent, the agent should strictly prefer action profile $(x_A, x_A)$.

Now consider the following: if the true $p$ is just below 0.6, then the actually attainable expected utility of an agent with any $\mu > 0.6$ is strictly higher compared to their expected expected utility if they holds the correct belief $\mu = p$. The choice $(x_A, x_A)$, which would be suboptimal for an agent without misperception, actually yields a strictly better outcome than the choice $x_I$. The agent with a biased prior and misperception fares better than an agent with only misperception because the suboptimal choice is not affected by misperception.

Result 6 generalizes this insight. It turns out that worst-case dominance is the critical condition for the possibility of there being a positive interaction between the two biases. Whenever a profile $\mathbf{y}$ that is optimal at some $\mu' \neq p$ worst-case dominates some profile $\mathbf{x}$ at $p$, then even if $\mathbf{x}$ is optimal at $p$ under ideal conditions, there exists some misperception such that $\mathbf{y}$ becomes the better choice. In this case, a naive agent with the incorrect prior $\mu'$ is inadvertently better off than a naive agent holding the correct prior.

**Result 6:** *For any action profile* $\mathbf{x}$ *chosen at* $\mu = p$ *and* $\mathbf{y} \neq \mathbf{x}$ *chosen at* $\mu' \neq \mu$*, there exist a misperception* $M_x$ *and a bias in prior such that a naive agent is better off with both biases than just misperception if and only if* $\mathbf{y}$ *worst-case dominates* $\mathbf{x}$ *at* $p$*.*

It is important to note, however, that a naive agent with two biases cannot be better off than a sophisticated agent who holds the correct prior. Otherwise, the sophisticated agent

would want to replicate the naive agent's choices and misperception error. Result 6 tells us by some bias in prior, a naive agent sometimes inadvertently compensates some or all of the advantage of a sophisticated agent. Interestingly though, a naive agent having two biases might be strictly better off than a sophisticated agent with the same two biases.

**Example 1 - continued:** Suppose the true prior is $p = 0.5$, yet the agent considers state $A$ more likely with prior $\mu$ just below 0.6. For a naive agent, the bias in prior is welfare neutral as the agent chooses the same action profile regardless. To see this, note that a naive agent chooses according to the utility frontier without misperception as in Figure 1 and taking a test is at the frontier for $\mu \in [0.4, 0.6]$. A sophisticated agent, however, is aware of the lower information value of $x_I$. As can be seen from Figure 3, at $\mu$ just below 0.6, the simple profile $(x_A, x_A)$ achieves higher expected utility due to the higher ex-ante probability of state $A$. The sophisticated agent thus adjust their choice at $\mu$ which leads to a welfare loss given that at the true prior $p$, $x_I$ still achieves higher expected utility despite the misperception.

The intuition behind this example is the following: Since a sophisticated agent correctly accounts for the lower-information value, they have the correct ordering of experiments at any belief. If $\mu = p$, they make the optimal choice from the perspective of a rational observer. Any deviation of $\mu$ from $p$ leads to (at least weakly) worse choices. While sophisticated agents can account for misperception, they cannot be aware of their bias in prior as this would be equivalent to not having any bias in prior in the first place. As they can compensate for one but not the other bias, they cannot benefit from any positive interaction. Holding a relatively more extreme prior, the sophisticated agent mistakenly undervalues the information generated from the test since information becomes less valuable, the more the agent is convinced of a given state. The cost from this mistake exceeds the benefit from correctly adjusting choices for the effect of misperception on the signal's informativeness.

We can repeat this analysis for the case where misperception is symmetric across experiments. Proposition 3 established that under symmetric misperception, an otherwise inferior action profile can become superior only if it strongly worst-case dominates the other. This implies that for a naive agent to benefit from an incorrect belief, the action profile chosen at that belief needs to strongly worst-case dominate the profile chosen at the correct $p$. Result 7 formalizes this:

**Result 7:** *For any signal sensitive action profile* $\mathbf{x}$ *chosen at* $\mu = p$ *and any* $\mathbf{y} \neq \mathbf{x}$ *chosen at*

*$\mu' \neq \mu$, there exists misperception $M = M_x = M_y$ and a bias in prior such that a naive agent is better off with both biases than just misperception if and only if* $\mathbf{y}$ *strongly worst-case dominates* $\mathbf{x}$ *at* $p$.

### 6.3.2. Adding Misperception to a Biased Prior

Next, we take the agent's biased prior as given and show under which conditions adding or increasing a perception bias can make them better off. In the context of our example, if a physician exhibits a biased ex-ante view, is there a way for a lab technician (who does not know the state of the world) to manipulate results such that the patient is better off? This result turns out to be more clear-cut. Adding any type of perception bias makes an agent worse off if the chosen action profile is not 'too sub-optimal' given the true $p$. A lab technician can only improve the outcome if, given $p$, a particular treatment is optimal independent of the outcome of the test.

For any signal sensitive profile $\mathbf{x} = (x, \{x_a, x_b\})$, let $\mathbf{x}_a = (x, x_a)$ and $\mathbf{x}_b = (x, x_b)$ be the respective simple profiles constructed from it.

**Proposition 4:** *Let* $\boldsymbol{x}$ *be a signal-sensitive action profile chosen at* $\mu \neq p$. *There exists misperception* $M_x$ *such that* $V(P_x M_x|\ \mathbf{x}, p) > V(P_x|\ \mathbf{x}, p)$ *if and only if* $\mathbf{x}$ *does not worst-case dominate both* $\mathbf{x}_a$ *and* $\mathbf{x}_b$ *at* $p$ *and signal strength is finite.* $V(P_x M_x|\ \mathbf{x}, p)$ *is increasing in the degree of misperception for some signal.*

We can think of Proposition 4 in terms of an observer who holds the correct prior $p$. If this observer - given their prior $p$ - prefers $\mathbf{x}$ to either of the two simple profiles that can be derived from it, then the information delivered by $\mathbf{x}$ still holds value at $p$. In that case, adding or increasing any misperception can only decrease expected utility. If instead an observer would prefer one of the simple action profiles, information generated from $\mathbf{x}$ has no value and thus conditioning actions on the outcome is not optimal. We can thus find a type of misperception that benefits an agent at $\mu \neq p$.

We can also interpret Proposition 4 in terms of how severe the bias in prior is. Recall that an action profile is signal sensitive at $\mu$ if the posteriors following different signals falls into different parts of the partition that prescribes optimal second-period choices. If we increase the prior, action $x_a$ becomes more appealing relative to $x_b$. Increasing it far enough, the agent should prefer $x_a$ regardless of the signal realization; $x_a$ is, after all, strictly better in state $A$. So if the true $p$ is sufficiently far from $\mu$, misperception that favors $a$ signals helps the agent. If the agent's prior is close to $p$, however, the agent is

always worse off due to the lower informativeness of signals. Finally, it it follows from this discussion that for every signal sensitive profile, there always exists a region (for $p$) for either case as long as signal strength is finite.[11]

This implies that for a severe enough bias, a neutral intermediary could identify the bias and manipulate signal realizations to increase welfare without holding any additional information about the state. To illustrate this in the context of our thought experiment, suppose that given the patient's history, a physician should have concluded that a certain condition is very likely and thus have started treatment immediately. Instead, the physician failed to take into account some of the information and ordered a test first. Suppose the lab technician has access to the history of the patient and reaches the correct assessment. As the physician ordered the test first, the technician can conclude that the physician's initial assessment (their prior) must be biased. Keeping in mind that the test is not perfectly accurate, by misreporting negative test results as positive (and therefore increasing misperception of the physician), the lab technician might be able to increase the patient's welfare if the differences in outcomes are severe enough. If they are not, any interference with the actual results will leave the patient strictly worse-off.

### 6.4. MARTINGALE BIAS

Until now, we imposed no restriction on how misperception can affect different signal types. In general, misperception can either be balanced, $k_a = k_b \neq 1$, or unbalanced, $k_a \neq k_b$. If one type of signal is more likely to be misperceived than another, misperception is unbalanced. This creates a drift in beliefs meaning that a rational observer, who is aware of the bias, would form an expectation over the agent's posterior that is different from the agent's prior. In contrast, a simple random error that occurs indiscriminately of the signal results in balanced misperception. In this case, an observer's expectation equals the agent's prior. We call this form of misperception a *martingale bias*. In the context of the lab technician, this describes a scenario where there is no particular relationship between the test result and the mistakes. The technician simply commits random errors, entering the wrong result in the patients' files sometimes. Analyzing this restricted class of balanced misperception, we show in contrast to Proposition 4, that an agent with a biased prior can never be better off when a martingale perception bias is added. However, analogous to Result 6, an agent with misperception might still benefit from a bias in prior.

Consider an action $x \in X_1$ with misperception of $k_a(x) = k_b(x) \equiv \kappa$. Assumption 1

[11] See Result A.3 for a formal proof.

requires that $\kappa \in [\frac{1}{2}, 1]$. The misperception matrix can be written as:

$$M_\kappa = \begin{bmatrix} \kappa & 1-\kappa \\ 1-\kappa & \kappa \end{bmatrix}$$

The effective probability of receiving an $a$-signal in state $A$ becomes $\kappa\pi(x) + (1-\kappa)\big(1-\pi(x)\big)$. As mistakes are symmetric, signal probabilities are evened out, thus weakening the signal strength. When computing expected utility, a naive agent places too much weight on the taking the right action in each state and thus considers outcomes $u(x, x_a|A)$ and $u(x, x_b|B)$ more likely than they are. The following result shows that expected utility decreases as mistakes become more likely for *any* combination of $p$ and individual prior $\mu$.

**Result 8:** *For any signal sensitive profile* $\mathbf{x}$ *and any* $\kappa \in (\frac{1}{2}, 1]$*, a decrease in* $\kappa$ *decreases the expected utility of* $\mathbf{x}$ *for any* $\mu$ *and* $p \in (0,1)$*.*

While Proposition 4 identified cases where a bias in misperception can improve an agent's welfare, Result 8 shows that a martingale bias never benefits an agent, even if they have a severe bias in prior. The martingale bias adds noise and thus reduces the signal strength equivalently for both signal types. Independent of the true state, this leads to strictly worse period-2 choices. However, as the next result shows, it is still possible that an agent suffering from random misperception benefits from an incorrect prior. As discussed before, this requires the 'reversal' of the utility ranking of two profiles and we can again characterize when this is possible. For this purpose, we introduce a suitable strengthening of worst-case dominance. Let $M_\emptyset$ be the misperception matrix that reduces the information value of any experiment to pure noise, meaning that $\kappa = \frac{1}{2}$.

**Definition 3** (NOD)**:** *An action profile* $\mathbf{y} \in \mathbf{X}^*$ ***noise dominates*** $\mathbf{x} \in \mathbf{X}^*$ *at* $\mu$ *if*

$$V(P_y|\,\mathbf{y}, \mu) > V(P_x M_\emptyset|\,\mathbf{x}, \mu)$$

It is easily verified that noise dominance implies worst-case dominance (see Result A.4 in the Appendix) while the converse is not true. Using this definition, Result 9 establishes the equivalent of Result 6 for the case where misperception is required to be symmetric across signals but not necessarily across actions.

**Result 9:** *For any signal sensitive action profile* $\mathbf{x}$ *that is chosen at* $\mu = p$*, and any action*

*profile* $\mathbf{y} \neq \mathbf{x}$ *that is chosen at* $\mu' \neq \mu$*, there exist* $\kappa$ *such that* $V(P_y|\,\mathbf{y},\mu) > V_n(P_x M_\kappa|\mu)$ *if and only if* $\mathbf{y}$ *noise dominates* $\mathbf{x}$ *at* $\mu$*.*

Just as in the case of symmetric misperception across experiments, symmetry across signals reduces the possible positive interaction of biases to a smaller set of action profiles. But the basic result remains the same for the profiles in this set. Whether misperception can reverse the utility ranking of two action profiles $\mathbf{x}$ and $\mathbf{y}$ hinges on whether $\mathbf{x}$ is still preferred if signals are pure noise. If so, there is no bias that would make an agent better off choosing $\mathbf{y}$. If not, an agent could benefit from a bias in prior that makes them choose $\mathbf{y}$ and thus avoid the noisier than expected experiment $\mathbf{x}$.

We can introduce a final restriction on the bias: misperception is required to be symmetric across both experiments and signals. Rather than committing different random errors for different experiments, we now restrict the lab technician to commit the same random error across all experiments. The technician might sometimes be distracted and enters results incorrectly independently of which test is being performed and which result is realized. We can again show that this requires a suitable strengthening of strong worst-case dominance as well as noise dominance but the basic idea of the previous result goes through.

**Definition 4** (strong NOD)**:** *An action profile* $\mathbf{y} \in \mathbf{X}^*$ ***strongly noise dominates*** *a profile* $\mathbf{x} \in \mathbf{X}^*$ *at* $\mu$ *if*

$$V(P_y M_\emptyset|\,\mathbf{y},\mu) > V(P_x M_\emptyset|\,\mathbf{x},\mu)$$

For completeness, Result A.5 shows formally that strong noise dominance implies strong worst-case dominance. Based on Result 8, we can also conclude that strong noise dominance implies noise dominance and is indeed a strengthening.

**Result 10:** *For any signal sensitive action profile* $\mathbf{x}$ *that is chosen at* $\mu$*, and any signal-sensitive action profile* $\mathbf{y} \neq \mathbf{x}$ *that is chosen at* $\mu' \neq \mu$*, there exists a misperception* $M_\kappa$ *such that* $V(P_y M_\kappa|\,\mathbf{y},\mu) > V_n(P_x M_\kappa|\mu)$ *if and only if* $\mathbf{y}$ *strongly noise dominates* $\mathbf{x}$ *at* $\mu$*.*

Similar to Result 9, Result 10 establishes that if we want to check whether an agent with prior $\mu'$ could be better off than one at the true $p$, we have to compare profiles $\mathbf{y}$ and $\mathbf{x}$ - the profiles chosen by a naive agent at these priors - at the extreme point where the two experiments yield just noise. If at this point $\mathbf{y}$ is preferred, such a reversal in the ordering of $\mathbf{x}$ and $\mathbf{y}$ is possible. But of course, this could already happen at a less extreme distortion.

In the context of our thought experiment, we can conclude that if tests are noisier than anticipated by the doctor, a patient might be better off from some treatment plan where

information has less value and resulting treatment options are less extreme. Suppose a patient is at least somewhat likely to have contracted a disease but the treatment has significant side effects. If the test for this condition is noisier than the physician expects, the patient could benefit from withholding some relevant information from their medical history so that the physician concludes that the condition is rather unlikely in the patient. The doctor might then rely on some further observations first, thus avoiding the noise of the experiment and reducing the risk of overtreatment.

## 7. EXAMPLE: PRODUCT LAUNCH

To illustrate our previous results in a different context, we turn to a typical business problem: a CEO contemplates several options how to introduce a new product to the market; from a grand product launch with an expensive marketing campaign to a soft launch in only a few selected regions. While a large-scale introduction might generate higher cashflow right away, it also entails greater risks. In contrast, a soft launch allows the company to gain more information about how the product will be received. This example highlights another aspect of flawed information acquisition: in addition to overpaying for information, the company may subscribe to a strategy that is too risky. In particular, it may be riskier than what the actual information warrants and what the decision-maker would have chosen in the absence of any information acquisition.

We model this as a two stage process. First, a risk-neutral CEO can acquire information about the state of the world through a potential soft launch. For simplicity, we assume the signal from this perfectly reveals the state. Afterwards, the CEO decides how many resources to allocate to the (actual) launch, captured by three alternatives, $\{z, r, e\}$. $z$ represents a low-key, relatively risk-free approach. $r$ is a riskier launch that employs a medium amount of resources to advertising and promotion. And $e$ is an extremely risky plan built around a large marketing campaign.

The cost of the soft launch is $1/2$. $z$'s payoffs are normalized to zero, $r$'s payoffs are 0.8 in state $A$ and $-1.2$ in state $B$, and $e$'s payoffs are 1 in state $A$ and $-4$ in state $B$. Suppose, however, that the CEO is not fully rational and instead exhibit a variety of biases when evaluating information. In particular, the CEO misperceives $b$-signals - information that indicates the product will not find much success - for $a$-signals with probability $\delta$; i.e. $k_a = 1$ and $k_b = 1 - \delta$. Furthermore, the CEO is naive about this bias.

At $t = 2$, the optimal choice is $z$ for $\mu \leq 0.6$, $r$ for $0.6 < \mu \leq 28/30$ and $z$ for $28/30 < \mu$. At

$t = 1$, the CEO acquires information if $\mu \in (0.5, 0.7]$.[12] As the signal is perfectly informative, the CEO opts for $z$ after a $b$-signal and $e$ after an $a$-signal. The true expected utility of this signal-sensitive profile is $\mu - (1-\mu) \cdot 4\delta - 1/2$ whereas the CEO believes it to be $\mu - 1/2$. Misperception causes the CEO to sometimes take the extremely risky choice when it is very costly. As a result, acquiring information is not optimal when $\delta$ is large. For any $p \in (0.6, 0.7]$, the CEO benefits from being slightly overconfident ($\mu \in (0.7, 20/30]$) when $\delta > \frac{0.7-\mu}{4(1-\delta)}$. While for $p \in (0.5, 0.6]$, the CEO is better off being underconfident ($\mu \leq 1/2$), when $\delta > \frac{\mu - 1/2}{4(1-\delta)}$. In both cases, holding relatively more extreme beliefs reduces the perceived benefit of information, alleviating at least some of the disadvantage of misperception. First, it allows the CEO to avoid the cost of acquiring information. Second, it ensures the CEO doesn't find themself in a position of having to make a choice with an objectively wrong belief *after* having received information.

The key problem is that the naive CEO is too confident after an $a$-signal. Instead of $\mu(a) = 1$, the CEO's posterior should be $\mu(a) = \frac{\mu}{\mu + (1-\mu)\cdot\delta}$.[13] From the correct posterior, we see that misperception makes a soft-launch less useful as it reduce the information value. In the extreme case where $\delta = 1$, nothing can be learned at all. For a naive CEO, this is harmful as the CEO may nevertheless become more convinced in the potential success after a slow roll-out. As a result, the company may opt for a grand secondary launch with large ad-spending despite the fact that it is unwarranted by the evidence.

Comparing the actual expected utility of each option requires exact knowledge about the degree of misperception $\delta$. In the absence of this, we can instead check whether acquiring information is (strongly) worst-case dominated, i.e. could at least in principle be suboptimal.

After a soft-launch, the CEO implements $e$ after an $a$-signal and $z$ after a $b$-signal. For $\mu \leq 0.8$, $e$ has lower ex-ante expected utility. The worst-case signal for those beliefs is thus $a$ as it results in $e$ being chosen at $t = 2$. Note that even though at $\mu$ close to 0.8, state $A$ is very likely, $a$ is still the worst-case signal as it results in the action that is most harmful to the CEO given *current* beliefs. From comparing payoffs, we can conclude that for beliefs that induce the CEO to purchase information, $\mu \in (0.5, 0.7]$, the strategies $z$ or $r$ without any information acquisition both strongly worst-case dominate acquiring information. Furthermore, for $\mu \leq 0.6$ choosing $z$ and not acquiring information is not strongly worst-case dominated by any other strategy. And similarly for $r$ and $\mu \in [0.6, 14/15]$. So even

[12] Indifference is broken in favor of the less risky option. Details can be found in the appendix.

[13] A sophisticated CEO would consequently find the less risky choice $r$ preferable for many intermediate levels of $\delta$.

without exact knowledge of $\delta$, we can conclude that with misperception, a low-risk or moderately risky launch without information acquisition are in a sense 'safer' choices. Furthermore, slight over- or underconfidence can be useful, in as far as it leads to these, more resilient choices.

In this case, a benevolent observer could limit the CEO's mistakes. For example, managers who are aware of their superiors' tendency to evaluate information too positively could manipulate their superiors' beliefs upward before any alternative is chosen. This causes the CEO to believe that the product is more likely to succeed than is warranted based on initial data alone. Consequently, the CEO will see less value from acquiring information in the first place and opt directly for some mid-size launch, option $r$. Such manipulation may occur by omission of useful information or active provision of wrong information in order to avoid a long, pointless phase of information acquisition. Interestingly, our results suggests that the general tendency of managers (Malmendier and Tate (2005)) or entrepreneurs (Landier and Thesmar (2009)) to be overconfident may naturally counteract their perception biases. By themselves, such biases are harmful. However, they can be beneficial in the sense of reducing the manager's perceived value of acquiring information.

## 8. DISCUSSION

We analyzed the effects of two fundamental mistakes in information processing and characterized their individual and joint impact on welfare. As a key observation, misperception can differentially impact courses of action and thus reverse their expected utility ranking. Consequently, a naive decision-maker benefits in some situations from incorrect prior beliefs. We characterized these interactions of mistakes for four classes of misperception: (1) unrestricted, (2) equal across experiments, (3) equal across signals, and (4) equal misperception across experiments and signals.

The analysis can be seen from a purely positive perspective, describing the impact of typical biases on choices. But as already hinted at, it also allows for a normative angle. It can give insights on how to shape information flows in order to identify and mitigate biases for more robust decision-making. Consider a rational observer trying to analyze a decision-making process. Unaware of the exact form of misperception, the observer may want to approach the problem from a worst-case perspective. How bad can a choice be if misperception is the worst-possible? Which alternative choices could possibly lead to

an improvement? Or alternatively, could the observed choice be optimal; at least under the worst possible misperception? The notions of worst-case and noise dominance yield some answers to these questions.

One way to approach this more generally is by regarding misperception as a source of ambiguity over the information content of signals. While ambiguity usually refers to a multiplicity in priors, here we are faced with a multiplicity in signal probabilities.[14] To an outside observer, there is ambiguity over which signal probabilities apply. Asking whether a given choice is optimal thus equates to asking whether the ambiguity in signals allows for a compatible expected utility ranking of actions. Formalizing this, let $\mathscr{M}$ be the (closed) set of all misperception matrices that can occur in a given setting. $\mathscr{M}$ captures the degree of ambiguity; from none at all, where $\mathscr{M}$ just contains the identity matrix, to the maximum possible, where for every experiment, $\mathscr{M}$ contains all misperception matrices possible under Assumption 1. Suppose that under ideal conditions, some action profile $\mathbf{x} \in \mathbf{X}^*$ yields a strictly higher expected utility than some $\mathbf{y} \in \mathbf{X}^*$. Choosing $\mathbf{y}$ over $\mathbf{x}$ can be optimal for some form of misperception in $\mathscr{M}$ if

$$\min_{(M_y, M_x) \in \mathscr{M}} \left[ V(P_x M_x | \, \mathbf{x}, \mu) - V(P_y M_y | \, \mathbf{y}, \mu) \right] < 0 \tag{7}$$

Without any ambiguity in signal probabilities, (7) simply compares the (maximum) expected utility of both action profiles and no utility reversal is possible. If $\mathscr{M}$ is expanded, the set of possible utility outcomes increases as well, hence increasing ambiguity. This then creates the possibility of reversals in the expected utility ranking of action profiles. The four classes of misperception discussed are somewhat extreme cases that highlight the effects of ambiguity in two domains: within and across experiments. Equalizing misperception for all experiments removes ambiguity across experiments. Any misperception has to equally apply to all of them. Equalizing it for both types of signals, in contrast, removes ambiguity within experiments; misperception cannot generate a predictable drift in posteriors and instead takes the form of a martingale bias. To see directly how $\mathscr{M}$ shapes this analysis, consider case (1) where there are no restrictions and, in particular, misperception is not necessarily correlated across experiments. Condition (7) simplifies to:

$$\min_{M_x \in \mathscr{M}} V(P_x M_x | \, \mathbf{x}, \mu) < \max_{M_y \in \mathscr{M}} V(P_y M_y | \, \mathbf{y}, \mu) = V(P_y | \, \mathbf{y}, \mu)$$

According to Proposition 2, this equates to worst-case dominance (WCD). In other words,

[14] See Gilboa and Schmeidler (1989) for their seminal paper on ambiguity aversion.

the inequality holds if $\mathbf{y}$ worst-case dominates $\mathbf{x}$. Reducing signal ambiguity tightens this condition. If misperception affects all experiments equally, it becomes:

$$\min_{M \in \mathcal{M}} \left[ V(P_x M | \mathbf{x}, \mu) - V(P_y M | \mathbf{y}, \mu) \right] < 0$$

which according to Proposition 3 is equivalent to strong WCD. Less ambiguity in signal probabilities allows for fewer forms of misperception that might affect the optimality of a course of action. The worst-case outcome improves and the set of action profiles whose ranking can be reversed by misperception becomes smaller. In turn, the conditions that characterize these sets become more restrictive.

Figure 4 visualizes the relationships between our four notions of (strong) WCD, and (strong) noise dominance (NOD). Not only do the strong versions imply their weaker counterparts, as was shown previously, but NOD and strong WCD overlap, with strong NOD being fully contained in their intersection.[15] This describes both when awareness of

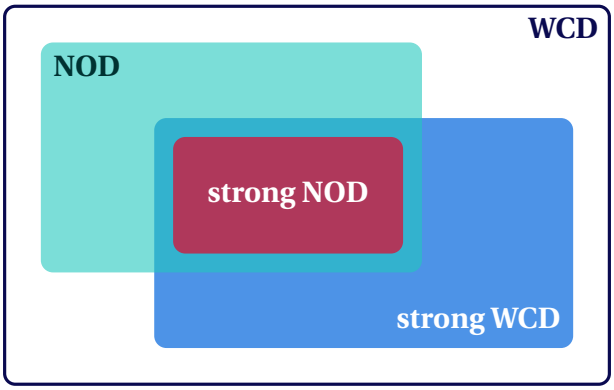


Figure 4: Four notions to compare action profiles

misperception can improve choices and when a bias in prior can mitigate mistakes from misperception for naive agents. For example, when $\mathcal{M}$ encompasses all forms of misperception, WCD is key in allowing for better choices through sophistication (Proposition 2). At the correct prior and for severe enough misperception, a sophisticated decision-maker is strictly better off than a naive one. But as Result 6 demonstrates, the impact on a naive agent could be reduced by an additional bias in prior if it leads to the choice of a profile that is less affected by misperception. This raises a more general point: observed heterogeneity in choices and beliefs might not necessarily indicate poor decision-making but can be a result of differences in misperception and approaches to mitigate its impact.

[15] See Result A.6 for the proof of the last statement.

Figure 4 also highlights that as ambiguity decreases, so do the benefits from choosing an alternative. Deviations from the unbiased, optimal choice are less likely to be the second-best choice.

Returning to our examples, if the accuracy of a test for an infectious disease is sensitive to even small errors, a patient faces significant ambiguity over how relevant the result is. If the physician is less aware of this ambiguity, for example by underestimating the probability of making a mistake in collecting the sample, then a patient has an incentive to manipulate the information they report to the physician. Just like the managers, who have an incentive to systematically manipulate the CEO's opinion, knowing about the (possibly) low information value of a soft launch. Incorrect beliefs of the physician and CEO might hence not be the problem itself but rather a fix to a more fundamental issue. For more robust tests and information experiments, ambiguity decreases and the scope for such adjustments is reduced. While errors and biases still lower the value of information, telling the full truth, taking a test, or obtaining additional information, and subsequently following the data might still be the best course of action.

In a broader context, these result highlight the need for a comprehensive understanding of imperfections in information processing to improve decision-making. One-sided approaches that address either misperception, for example by improving test standards, or correcting beliefs, for example through information campaigns, might have unexpected side-effects due to the intricate interaction of both biases. Nevertheless, as was also shown, interventions that jointly identify and mitigate biases can be feasible and sensible.

# APPENDIX A: PROOFS

**Lemma 1:** *For any action $x \in X_1$ and $y \in X_2$ with $E[u(x,y)|\mu(s)] > E[u(x,z)|\mu(s)]\ \forall z \in X_2 \setminus y$ for some $\mu(s) \in [0,1]$, there exists an interval $\mathscr{I} \subset [0,1]$ with the property that $y$ maximizes expected utility if and only if $\mu(s) \in \mathscr{I}$.*

*Proof.* If for some $(x,y)$ and some $\mu \in [0,1]$ we have $E[u(x,y)|\mu] > E[u(x,z)|\mu]\ \ \forall z \in X_2 \setminus y$ then linearity in $\mu$ implies that this is also true for at least some interval $[\mu, \overline{\mu})$ or $(\underline{\mu}, \mu]$. If this interval is $[0,1]$, the proof is complete.

If not, then there is some $\mu_j \in (0,1)$ and $y_j \in X_2$ with $y_j \neq y$ such that $E[u(x,y_j)|\mu_j] > E[u(x,y)|\mu_j]$. This implies that either $u(x,y_j|A) > u(x,y|A)$ or $u(x,y_j|B) > u(x,y|B)$. Then again by linearity for either all $\mu(s) > \mu_j$ or all $\mu(s) < \mu_j$, $y_j$ achieves a higher expected utility than $y$. Suppose $\mu_j > \mu$, then $y$ can never be optimal for any $\mu(s) \in (\mu_j, 1]$ and mutatis mutandis the argument applies to the case $\mu_j < \mu$. Iterating over any other $y' \in X_2$, the result follows. □

*Proof of Result 1.* : This follows immediately from iterating Lemma 1 over all elements of $X_2$ that are optimal for some $\mu$. As $X_2$ must have at least one such element, such a partition always exists. □

**Lemma 2:** *For any $x \in X_1$, Assumption 1 is equivalent to $k_a(x) + k_b(x) \geq 1$.*

*Proof.* For any such $x \in X_1$, we can write the likelihood-ratio for $a$-signals given some misperception $k_a = k_a(x)$ and $k_b = k_b(x)$ as

$$\frac{\hat{\pi}_A(x)}{1-\hat{\pi}_B(x)} = \frac{k_a\pi(x) + (1-k_b)(1-\pi(x))}{k_a(1-\pi(x)) + (1-k_b)\pi(x)}.$$

It is easily verified that for this ratio to be greater or equal 1, it must be that $k_a(x) + k_b(x) \geq 1$. The equivalent holds for $b$-signals. The result follows. □

*Proof of Result 2.* Let $\vec{\mu} = (\Pr(\omega = A), \Pr(\omega = B))$ be row vector of belief about the different states $\Omega$. For any $\mathbf{x} \in \mathbf{X}^*$, define $\vec{u}(\mathbf{x})$ as the state contingent vector:

$$\vec{u}(\mathbf{x}) = \begin{pmatrix} \pi(x) \cdot u(x, x_a|A) + \big(1-\pi(x)\big) \cdot u(x, x_b|A) \\ \big(1-\pi(x)\big) \cdot u(x, x_a|B) + \pi(x) \cdot u(x, x_b|B) \end{pmatrix}$$

Define the maximum expected utility as a function of $\vec{\mu}$ as $g(\vec{\mu}) \equiv \max_{x \in \mathbf{X}^*} \vec{\mu} \cdot \vec{u}(x)$. This expression coincides with the expected utility formulation of equation 1.

To show $g(\vec{\mu})$ is convex, let $\mathbf{x}'$ and $\mathbf{x}''$ be the optimal choices in $\mathbf{X}^*$ given the belief $\vec{\mu}'$ and $\vec{\mu}''$ respectively. Let $\mathbf{x}^*$ the optimal choice for the convex combination $\vec{\mu}^* = \lambda\vec{\mu}' + (1-\lambda)\vec{\mu}''$ with $\lambda \in (0,1)$. Then

$$\begin{aligned} g(\vec{\mu}^*) = g(\lambda\vec{\mu}' + (1-\lambda)\vec{\mu}'') &= (\lambda\vec{\mu}' + (1-\lambda)\vec{\mu}'') \cdot \vec{u}(\mathbf{x}^*) \\ &\leq \lambda\vec{\mu}' \cdot \vec{u}(\mathbf{x}') + (1-\lambda)\vec{\mu}'' \cdot \vec{u}(\mathbf{x}'') = \lambda g(\vec{\mu}') + (1-\lambda) g(\vec{\mu}'') \end{aligned}$$

proving convexity. □

**Result A.1:** *For any $x \in X_1$ with $\pi(x) > \frac{1}{2}$, any increase in the magnitude of the bias strictly decreases the likelihood-ratios $\frac{\hat{\pi}_A(x)}{1-\hat{\pi}_B(x)}$ and $\frac{\hat{\pi}_B(x)}{1-\hat{\pi}_A(x)}$.*

*Proof of Result A.1.* Take any $x \in X_1$ and associated misperception matrix $M_x$. We can write $P_x M_x$ as:

$$\begin{bmatrix} \pi & 1-\pi \\ 1-\pi & \pi \end{bmatrix} \begin{bmatrix} k_a & 1-k_a \\ 1-k_b & k_b \end{bmatrix} = \begin{bmatrix} \hat{\pi}_A & 1-\hat{\pi}_A \\ 1-\hat{\pi}_B & \hat{\pi}_B \end{bmatrix}$$

Now suppose we increase the magnitude of the bias increases, resulting in the misperception $M'_x$. Suppose the signal probabilities are now $\hat{\pi}'_s$ and, contrary to the statement in the result, $\frac{\hat{\pi}'_A}{1-\hat{\pi}'_B} \geq \frac{\hat{\pi}_A}{1-\hat{\pi}_B}$. This implies that:

$$(2\pi - 1)\big(k'_a - (1-k'_b)\big) \geq (2\pi-1)\big(k_a - (1-k_b)\big)$$

where $k'_a$ and $k'_b$ are the elements in $M'_x$. But since $\pi > 1/2$, this requires that either $k'_a > k_a$, or $k'_b > k_b$, or both which contradicts the increase in the magnitude of the bias. The equivalent argument applies for $\hat{\pi}'_B$ and $\hat{\pi}_B$. □

*Proof of Proposition 1.* . Let $x \in X_1$ be some experiment with associated matrix $P_x$ and misperception $M_x$. We can write the expected utility of $P_x M_x$ at $\mu$ as:

$$\begin{aligned} V_n(P_x M_x|\mu) = \quad & \mu \cdot \Big[\big(k_a\pi + (1-k_b)(1-\pi)\big)u(x,x_a|A) + \big(k_b(1-\pi) + (1-k_a)\pi\big)u(x,x_b|A)\Big] \\ + (1-\mu) \cdot & \Big[\big(k_a(1-\pi) + (1-k_b)\pi\big)u(x,x_a|B) + \big(k_b\pi + (1-k_a)(1-\pi)\big)u(x,x_b|B)\Big] \end{aligned}$$

where $x_a, x_b \in X_2$ are the maximizers of the expression at $k_a = k_b = 1$, i.e. when there is no misperception. Fixing $k_b$, an increase in misperception strictly lowers expected utility if:

$$\frac{\partial V_n(P_x M_x|\mu)}{\partial k_a} = \mu \cdot \pi\Big[u(x,x_a|A) - u(x,x_b|A)\Big] + (1-\mu)\cdot(1-\pi)\Big[u(x,x_a|B) - u(x,x_b|B)\Big] > 0 \qquad (8)$$

But this is exactly the condition required for signal sensitivity (i.e. $x_a \neq x_b$) to be optimal. If this does not hold, then $x_a = x_b$ and the effect is 0. The equivalent argument applies to $k_b$. This gives the first inequality $V(P_x|\mu) \geq V(P_x M_x|\mu)$. By definition of $V(P_x M_x|\mu)$ and $V_n(P_x M_x|\mu)$ (see Equation 4 and Equation 5), it follows immediately that $V(P_x M_x|\mu) \geq V_n(P_x M_x|\mu)$. □

*Proof of Result 3:* This follows immediately from equation (8) and the subsequent argument in the proof of Proposition 1. □

*Proof of Proposition 2.* Let $\mathbf{x} = (x, \{x_a, x_b\})$ and $\mathbf{y} = (y, \{y_a, y_b\})$ be the profiles in question, which may or may not be signal sensitive. If $V(P_y|\ \mathbf{y}, \mu) > V(P_x|\ \mathbf{x}, \mu)$, the statement is trivially true with $M_y$ and $M_x$ equal to the identity matrix. $\mathbf{y}$ necessarily worst-case dominates $\mathbf{x}$ as it achieves a higher expected utility than the optimal $\mathbf{x}$, not just the worst-case simple profile. The converse therefore also holds in this case.

Suppose now that $V(P_x|\ \mathbf{x}, \mu) \geq V(P_y|\ \mathbf{y}, \mu)$ but $\mathbf{y}$ worst-case dominates $\mathbf{x}$. We first conclude that $\mathbf{x}$ must be a signal-sensitive profile as otherwise

$$V(P_x|\ \mathbf{x}, \mu) = \min_{s \in S(x)} \big(\mu u(x, x_s|A) + (1-\mu) u(x, x_s|B)\big)$$

meaning that $\mathbf{y}$ could not worst-case dominate $\mathbf{x}$. By Definition 1, we know that $\mu u(x, x_s|A) + (1-\mu) u(x, x_s|B) < V(P_y|\ \mathbf{y}, \mu)$ for some $s \in \{a, b\}$. Suppose wlog that $s = a$. This implies that for a small enough $\epsilon > 0$, we have

$$\begin{aligned} V(P_y|\ \mathbf{y}, \mu) > &\mu\big[(1-\epsilon) u(x, x_a|A) + \epsilon u(x, x_b|A)\big] \\ &+ (1-\mu)\big[(1-\epsilon) u(x, x_a|B) + \epsilon u(x, x_b|B)\big] \end{aligned} \tag{9}$$

Now let $M_x$ be such that $k_a(x) = 1-\epsilon$ and $k_b(x) = \epsilon$ and $M_y$ be the identity matrix. Then the right-hand side of Equation 9 is exactly equal to $V(P_x M_x|\ \mathbf{x}, \mu)$ and therefore $V(P_y M_y|\ \mathbf{y}, \mu) > V(P_x M_x|\ \mathbf{x}, \mu)$ as desired.

For the converse, suppose $V(P_x|\ \mathbf{x}, \mu) \geq V(P_y|\ \mathbf{y}, \mu)$ but $V(P_y M_y|\ \mathbf{y}, \mu) > V(P_x M_x|\mathbf{x}, \mu)$ for some $M_x$ and $M_y$. From Result 3, we know that if such an $M_y$ exists, it must also be true for $M_y = I$ and so $V(P_y|\ \mathbf{y}, \mu) > V_n(P_x M_x|\ \mathbf{x}, \mu)$. Note that:

$$\begin{aligned} V(P_x M_x|\ \mathbf{x}, \mu) = \quad & \mu \cdot \Big[\big(k_a \cdot \pi + (1-k_b)(1-\pi)\big) u(x, x_a|A) + \big(k_b \cdot (1-\pi) + (1-k_a) \cdot \pi\big) u(x, x_b|A)\Big] \\ & + (1-\mu) \cdot \Big[\big(k_a \cdot (1-\pi) + (1-k_b) \cdot \pi\big) u(x, x_a|B) + \big(k_b \cdot \pi + (1-k_a)(1-\pi)\big) u(x, x_b|B)\Big] \end{aligned}$$

Wlog, assume

$$\mu \cdot \big(k_a \cdot \pi + (1-k_b)(1-\pi)\big) u(x, x_a|A) + (1-\mu) \cdot \big(k_a \cdot (1-\pi) + (1-k_b) \cdot \pi\big) u(x, x_a|B)$$
$$\geq \mu \cdot \big((1-k_a) \cdot \pi + k_b \cdot (1-\pi)\big) u(x, x_b|A) + (1-\mu) \cdot \big((1-k_a)(1-\pi) + k_b \cdot \pi\big) u(x, x_b|B)$$

It follows that

$$\frac{\partial}{\partial k_a} V(P_x M_x|\, \mathbf{x}, \mu) \geq \frac{\partial}{\partial k_b} V(P_x M_x|\, \mathbf{x}, \mu) \tag{10}$$

and because of linearity, this holds for any $k_a$ and $k_b$. Then if $V(P_y|\, \mathbf{y}, \mu) > V_n(P_x M_x|\, \mathbf{x}, \mu)$ for any $M_x$, it must be the case for $k_a = 0$ and $k_b = 1$ as by Assumption 1 and Lemma 2, $k_a + k_b \geq 1$. It follows that $V(P_y|\, \mathbf{y}, \mu) > \mu u(x, x_b|A) + (1-\mu) u(x, x_b|B)$ and hence $\mathbf{y}$ worst-case dominates $\mathbf{x}$. □

*Proof of Result 4.* Take any $\mathbf{x} = (x, \{x_a, x_b\})$ and any simple profile $\mathbf{y}$, both in $\mathbf{X}^*$. It follows from Proposition 2 that for $\mathbf{y}$ to achieve higher expected utility, it needs to worst-case dominate $\mathbf{x}$ at any such $\mu$. Worst-case dominance requires that

$$V(P_y|\, \mathbf{y}, \mu) > \min_{s \in \{a,b\}} \mu u(x_1, x_s|A) + (1-\mu) u(x_1, x_s|B).$$

Assume that the minimizer is $x_a$. Let $\mathbf{x}' = (x, x_a)$ and $\mathbf{x}'' = (x, x_b)$. It follows that $V(P_x|\, \mathbf{x}'', \mu) > V(P_x|\, \mathbf{x}', \mu)$. Set $\mathbf{y} = \mathbf{x}''$. Then $\mathbf{y}$ worst-case dominates $\mathbf{x}$ at $\mu$ and so there is an $M_x$ such that $V(P_y|\, \mathbf{y}, \mu) > V(P_x M_x|\, \mathbf{x}, \mu)$. The equivalent is true for $x_b$ being the minimizer. Because of linearity in $\mu$, there is only one $\mu \in [0,1]$ such that $V(P_x|\, \mathbf{x}'', \mu) = V(P_x|\, \mathbf{x}', \mu)$. In that case, the inequality holds weakly. Finally let $\mathbf{x}$ be such that the t=2 actions are optimal in $X_2$. Then by definition $V(P_x M_x|\, \mathbf{x}, \mu) = V_n(P_x M_x|\mu)$ and again by definition $V(P_y|\mu) \geq V(P_y|\mathbf{y}, \mu)$ and hence $V(P_y|\mu) > V_n(P_x M_x|\mu)$ as required. □

**Result A.2** (Strong worst-case dominance implies worst-case dominance)**:** *An action profile* $\mathbf{y} = (y, \{y_a, y_b\}) \in \mathbf{X}^*$ *strongly worst-case dominates* $\mathbf{x} = (x, \{x_a, x_b\}) \in \mathbf{X}^*$ *at* $\mu$ *only if* $\mathbf{y}$ *worst-case dominates* $\mathbf{x}$ *at* $\mu$*. The converse is not true.*

*Proof.* By definition $V(P_y|\, \mathbf{y}, \mu) \geq \max_{s \in \{a,b\}} \mu u(y, y_s|A) + (1-\mu) u(y, y_s|B)$. A necessary condition for $\mathbf{y}$ to strongly worst-case dominate $\mathbf{x}$ is:

$$\max_{s \in \{a,b\}} \mu u(y, y_s|A) + (1-\mu) u(y, y_s|B) > \min_{s \in \{a,b\}} \mu u(x, x_s|A) + (1-\mu) u(x, x_s|B) \tag{11}$$

Which by definition of $V(\cdot)$ implies that $V(P_y|\,\mathbf{y},\mu) > \min_{s\in\{a,b\}} \mu u(x,x_s|A)+(1-\mu)u(x,x_s|B)$. But this is the definition of worst-case dominance. To disprove the converse statement, we can see directly that the last inequality can be fulfilled without inequality (11) holding as long as the difference $u(y,y_s|\omega) - u(x,y_x|\omega)$ is sufficiently large for some state. As inequality (11) is necessary for strong worst-case dominance, the proof is complete. □

*Proof of Proposition 3.* *Sufficiency:* Suppose some action profile $\mathbf{y} = (y,\{y_a,y_b\}) \in \mathbf{X}^*$ strongly worst-case dominates some $\mathbf{x} = (y,\{x_a,x_b\}) \in \mathbf{X}^*$. Assume WLOG that

$$\mu u(y,y_a|A) + (1-\mu)u(y,y_a|B) > \mu u(x,x_a|A) + (1-\mu)u(a,x_a|B)$$

Consider the misperception matrix $M = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ where the probability of receiving an $a$-signal is 1 in both states. This implies that

$$V(P_yM|\,\mathbf{y},\mu) = \mu u(y,y_a|A) + (1-\mu)u(y,y_a|B)$$

and equivalently for $\mathbf{x}$. It follows that $V(P_yM|\,\mathbf{y},\mu) > V(P_xM|\,\mathbf{x},\mu)$ as required.

*Necessity:* Suppose $V(P_x|\,\mathbf{x},\mu) \geq V(P_y|\,\mathbf{y},\mu)$ with a distortion $M$ such that $V(P_yM|\,\mathbf{y},\mu) > V(P_xM|\,\mathbf{x},\mu)$. As misperception is equal across experiments, we can write the actual expected utility explicitly as a linear combination of the probability of receiving $a$ and $b$-signals:

$$\begin{aligned}
&V(P_xM|\,\mathbf{x},\mu) = \\
&\quad \mu\,[k_a\pi(x) + (1-k_b)(1-\pi(x))]\,u(x,x_a|A) \;+\; (1-\mu)[k_a(1-\pi(x)) + (1-k_b)\pi(x)]\,u(x,x_a|B) \\
&\quad +\mu\,[k_b(1-\pi(x)) + (1-k_a)\pi(x)]\,u(x,x_b|A) \;+\; (1-\mu)[k_b\pi(x) + (1-k_a)(1-\pi(x))]\,u(x,x_b|B)
\end{aligned}$$

$$\begin{aligned}
&V(P_yM|\,\mathbf{y},\mu) = \\
&\quad \mu\,[k_a\pi(y) + (1-k_b)(1-\pi(y))]\,u(y,y_a|A) \;+\; (1-\mu)[k_a(1-\pi(y)) + (1-k_b)\pi(y)]\,u(y,y_a|B) \\
&\quad +\mu\,[k_b(1-\pi(y)) + (1-k_a)\pi(y)]\,u(y,y_b|A) \;+\; (1-\mu)[k_b\pi(y) + (1-k_a)(1-\pi(y))]\,u(y,y_b|B)
\end{aligned}$$

Taking the derivative w.r.t. $k_a$, we get:

$$\frac{V(P_xM|\,\mathbf{x},\mu)}{\partial k_a} = \mu\pi(x)\big(u(x,x_a|A) - u(x,x_b|A)\big) + (1-\mu)(1-\pi(x))\big(u(x,x_a|B) - u(x,x_b|B)\big)$$

$$\frac{V(P_yM|\,\mathbf{y},\mu)}{\partial k_a} = \mu\pi(y)\big(u(y,y_a|A) - u(y,y_b|A)\big) + (1-\mu)(1-\pi(y))\big(u(y,y_a|B) - u(y,y_b|B)\big)$$

These do no depend on $k_b$ and, because of linearity, not on $k_a$. Suppose

$$\frac{V(P_x M|\,\mathbf{x},\mu)}{\partial k_a} - \frac{V(P_y M|\,\mathbf{y},\mu)}{\partial k_a} \geq \frac{V(P_x M|\,\mathbf{x},\mu)}{\partial k_b} - \frac{V(P_y M|\,\mathbf{y},\mu)}{\partial k_b}$$

then, if $V(P_y M|\,\mathbf{y},\mu) > V(P_x M|\,\mathbf{x},\mu)$ for some $k_a$ and $k_b$, it must also be the case for $k_a = 0$ and $k_b = 1$. But then

$$\mu u(x,x_b)|A) + (1-\mu)\,u(x,x_b|B) < \mu u(y,y_b)|A) + (1-\mu)\,u(y,y_b|B).$$

Suppose instead that

$$\frac{V(P_x M|\,\mathbf{x},\mu)}{\partial k_a} - \frac{V(P_y M|\,\mathbf{y},\mu)}{\partial k_a} < \frac{V(P_x M|\,\mathbf{x},\mu)}{\partial k_b} - \frac{V(P_y M|\,\mathbf{y},\mu)}{\partial k_b}$$

then the equivalent argument holds for $k_a = 1$ and $k_b = 0$ and so: $\mu u(x,x_a)|A) + (1-\mu)\,u(x,x_a|B) < \mu u(y,y_a)|A) + (1-\mu)\,u(y,y_a|B)$ as required.

□

*Proof of Result 5.* Take the action profile $(x,\{x_a,x_b\})$. It's actual expected utility is

$$\begin{aligned} &p\cdot[\pi(x)\,u(x,x_a|A) + (1-\pi(x))\,u(x,x_b|A)] \\ +&(1-p)\cdot[(1-\pi(x))\,u(x,x_a|B) + \pi(x)\,u(x,x_b|B)] \end{aligned}$$

If we evaluate this at some $\mu > p$, then the weight on the outcome in state $A$ (the first line of the equation) increases. This has two potential effects. First, as implied by Lemma 1, a different action might be chosen in period 2 as the potential posteriors are now different. Furthermore, a different $x$ might become optimal. Denote the potentially different choice by $(y,\{y_a,y_b\})$. This must satisfy

$$\begin{aligned} &\pi(y)\,u(y,y_a|A) + \big(1-\pi(y)\big)u(y,y_b|A) > \pi(x)\,u(x,x_a|A) + \big(1-\pi(x)\big)u(x,x_b|A) \\ &\big(1-\pi(y)\big)u(y,y_a|B) + \pi(y)\,u(y,y_b|B) < \big(1-\pi(x)\big)u(x,x_a|B) + \pi(x)\,u(x,x_b|B). \end{aligned}$$

Otherwise either $(y,\{y_a,y_b\})$ would not be optimal as it yields lower expected utility at $\mu$, or the choice at $p$ would not be optimal, as it results in lower utility in both states. It follows that if any choice of action is different when comparing optimal choices under $\mu$ and $p$, such a $\mu$ leads to a utility loss. Furthermore, fixing some $\mu > p$ and iterating this argument for $\mu' > \mu$ yields the result. The equivalent argument applies to $\mu < p$. □

*Proof of Result 6.* Let $\mathbf{x}$ and $\mathbf{y}$ be the profiles as in Proposition 2. As $\mathbf{x}$ is chosen at $p = \mu$, we know that $V(P_x|\,\mathbf{x},\mu) = V_n(P_x|\mu) > V_n(P_y|\mu) \geq V(P_y|\,\mathbf{y},\mu)$. It follows from Proposition 2 that this utility ranking can only be reversed with some misperception $M_x$ if and only if $\mathbf{y}$ worst-case dominates $\mathbf{x}$. In this case $V(P_y|\mu) > V_n(P_xM_x|\mu)$. As the agent is naive, $\mathbf{y}$ is only chosen for some $\mu' \neq p$ where $V(P_y|\mu') > V(P_x|\mu')$. But then given misperception $M_x$, the agent is better off with prior $\mu'$ than $\mu = p$. □

*Proof of Result 7.* This follows almost immediately from Proposition 3. As $\mathbf{x} \neq \mathbf{y}$ and $\mathbf{x}$ is chosen at $\mu = p$, we know that $V_n(P_x|\mu) = V(P_x|\,\mathbf{x},\mu) > V(P_y|\mu) \geq V(P_y|\,\mathbf{y},\mu)$. It follows from Proposition 3 that $V(P_xM|\,\mathbf{x},\mu) < V(P_yM|\,\mathbf{y},\mu)$ if and only if $\mathbf{y}$ strongly worst-case dominates $\mathbf{x}$. This completes the proof. □

*Proof of Proposition 4.* Consider the expected utility of $P_xM_x$ evaluated at $p$. Taking the derivative w.r.t. $k_a$ (see proof of Proposition 3), we get

$$\frac{V(P_xM|\,\mathbf{x},p)}{\partial k_a} = p\cdot\pi\Big(u(x,x_a|A) - u(x,x_b|A)\Big) + \big(1-p\big)\big(1-\pi\big)\Big(u(x,x_a|B) - u(x,x_b|B)\Big).$$

For expected utility to be increasing in misperception (decreasing in $k_s$ for some $s$), we thus need:

$$p\cdot\pi\cdot u(x,x_a|A) + \big(1-p\big)\big(1-\pi\big)u(x,x_a|B) < p\cdot\pi\cdot u(x,x_b|A) + \big(1-p\big)\big(1-\pi\big)u(x,x_b|B)$$

For an infinitely strong signal, $\pi = 1$, this cannot hold as by definition $u(x,x_a|A) > u(x,x_b|A)$. Suppose signal strength is finite instead. The condition is equivalent to:

$$\begin{aligned} V(P_x|\,\mathbf{x},p) =& p\cdot\Big[\pi\cdot u(x,x_a|A) + \big(1-\pi\big)u(x,x_b|A)\Big] + (1-p)\cdot\Big[\pi\cdot u(x,x_b|B) + \big(1-\pi\big)u(x,x_a|B)\Big] \\ <& p\cdot u(x,x_b|A) + (1-p)\cdot u(x,x_b|B) \end{aligned}$$

and thus expected utility is decreasing in $k_a$ if and only if $\mathbf{x}$ does not worst-case dominate $\mathbf{x_b}$. The equivalent argument can be made for $\mathbf{x_a}$ and $k_b$. The result follows. □

**Result A.3:** *For every signal sensitive action profile $\mathbf{x} \in X^*$ with finite signal strength, there exists $p$ and $M_x$ such that $V(P_xM_x|\,\mathbf{x},p) > V(P_x|\,\mathbf{x},p)$.*

*Proof.* As $\mathbf{x} = (x,\{x_a,x_b\}) \in X^*$ is signal sensitive, we know by definition of $X^*$ that $u(x,x_a|A) > u(x,x_b|A)$. Take $M_x$ such that $k_a = 1 = 1-k_b$ meaning that the agent always receives signal

$a$. The actual expected utility of $\mathbf{x}$, $V(P_x M_x|\,\mathbf{x}, p)$, is then strictly increasing in $p$. As the signal strength is finite, there exists a $p'$ such that for all $p > p'$, the simple profile $\mathbf{x}_a = (x, x_a)$ achieves strictly higher expected utility than $\mathbf{x}$. But this is equal to the expected utility of $\mathbf{x}$ with misperception $M_x$ and thus $V(P_x M_x|\,\mathbf{x}, p) > V(P_x|\,\mathbf{x}, p)$ as asserted. □

*Proof of Result 8.* Write the expected utility from $\mathbf{x}$ as:

$$\begin{aligned} V(P_x M_x|\,\mathbf{x}, \mu) = \quad & \mu \cdot \Big[\big(\kappa\pi + (1-\kappa)(1-\pi)\big)u(x, x_a|A) + \big((1-\kappa)\pi + \kappa(1-\pi)\big)u(x, x_b|A)\Big] \\ & + (1-\mu) \cdot \Big[\big(\kappa(1-\pi) + (1-\kappa)\pi\big)u(x, x_a|B) + \big((1-\kappa)(1-\pi) + \kappa\pi\big)u(x, x_b|B)\Big] \end{aligned}$$

It follows from Proposition 1 that any decrease in $\kappa$ reduces this expected utility which shows the result for $\mu = p$. But we can see further from the symmetry in distortions that any reduction in $\kappa$ reduces the weight on both $u(x, x_a|A)$ and $u(x, x_b|B)$ while increasing the weight on $u(x, x_b|A)$ and $u(x, x_a|B)$. This reduces expected utility for any $\mu$ and thus $V(P_x M_x|\mathbf{x}, p)$ is strictly decreasing as $\kappa$ decreases for any $p$. □

**Result A.4:** *An action profile* $\mathbf{y} \in \mathbf{X}^*$ *noise dominates a profile* $\mathbf{x}$ *at* $\mu$ *only if* $\mathbf{y}$ *worst-case dominates* $\mathbf{x}$ *at* $\mu$. *The converse is not true.*

*Proof.* By the definition of noise dominance,

$$V(P_y|\,\mathbf{y}, \mu) > \frac{\mu}{2}\big(u(x, x_a|A) + u(x, x_b|A)\big) + \frac{1-\mu}{2}\big(u(x, x_a|B) + u(x, x_b|B)\big)$$

Suppose now wlog

$$\frac{\mu}{2}u(x, x_a|A) + \frac{1-\mu}{2}u(x, x_a|B) \geq \frac{\mu}{2}u(x, x_b|A) + \frac{1-\mu}{2}u(x, x_b|B)$$

then substituting the right-hand side of this inequality back into the previous inequality, we get

$$V(P_y|\,\mathbf{y}, \mu) > \mu u(x, x_b|A) + (1-\mu)u(x, x_b|B)$$

and therefore $\mathbf{y}$ worst-case dominates $\mathbf{x}$ as asserted.

To disprove the converse, suppose $\mathbf{y}$ worst-case dominates $\mathbf{x}$ and assume wlog that

$$V(P_y|\,\mathbf{y}, \mu) > \mu u(x, x_a|A) + (1-\mu)u(x, x_a|B)$$

but then for sufficiently large $u(x, x_b|A)$ and $u(x, x_b|B)$, we can still have

$$\frac{\mu}{2}\big(u(x, x_a|A) + u(x, x_b|A)\big) + \frac{1-\mu}{2}\big(u(x, x_a|B) + u(x, x_b|B)\big) > V(P_y|\,\mathbf{y}, \mu)$$

violating noise dominance. □

*Proof of Result 9. Sufficiency:* Suppose $\mathbf{y}$ noise dominates $\mathbf{x}$ at $\mu$. As $\mathbf{x}$ is chosen at $\mu$, we know that $V(P_x M_x|\,\mathbf{x}, \mu) = V_n(P_x M_x|\mu)$. For $M_x = M_\emptyset$ and $M_y = I$, we immediately get $V(P_y M_y|\,\mathbf{y}, \mu) > V(P_x, M_x|\,\mathbf{x}, \mu) = V_n(P_x M_x|\mu)$ as required.

*Necessity:* Suppose $\mathbf{y}$ does not noise dominate $\mathbf{x}$ at $\mu$. It follows from Result 8 that for any $M_\kappa$:

$$\frac{V_n(P_x M_x|\mu)}{\partial \kappa} = \frac{V(P_x M_x|\,\mathbf{x}, \mu)}{\partial \kappa} > 0$$

where the first equality follows from $\mathbf{x}$ being chosen at $\mu$. We can then conclude that for any $\kappa \in (\frac{1}{2}, 1]$ and associated $M_\kappa$:

$$V_n(P_x M_\kappa|\mu) > V_n(P_x M_\emptyset|\mu) \geq V(P_y|\,\mathbf{y}, \mu)$$

which proves the contrapositive. □

**Result A.5:** *An action profile* $\mathbf{y} \in \mathbf{X}^*$ *strongly noise dominates* $\mathbf{x} \in \mathbf{X}^*$ *only if* $\mathbf{y}$ *strongly worst-case dominates* $\mathbf{x}$. *The converse is not true.*

*Proof.* Suppose that indeed $V(P_y M_\emptyset|\,\mathbf{y}, \mu) > V(P_x M_\emptyset|\,\mathbf{x}, \mu)$. Then

$$\begin{aligned}
&\frac{\mu}{2}\big(u(y, y_a|A) + u(y, y_b|A)\big) + \frac{1-\mu}{2}\big(u(y, y_a|B) + u(y, y_b|B)\big) \\
>&\frac{\mu}{2}\big(u(x, x_a|A) + u(x, x_b|A)\big) + \frac{1-\mu}{2}\big(u(x, x_a|B) + u(x, x_b|B)\big)
\end{aligned}$$

which can be rearranged to

$$\begin{aligned}
&\mu\big(u(y, y_a|A) - u(x, x_a|A)\big) + (1-\mu)\big(u(y, y_a|B) - u(x, x_a|B)\big) \\
>&\mu\big(u(x, x_b|A) - u(y, y_b|A)\big) + (1-\mu)\big(u(x, x_b|B) - u(y, y_b|B)\big).
\end{aligned}$$

For this inequality to hold we need that either

$$\mu u(y, y_a|A) + (1-\mu)\,u(y, y_a|B) > \mu u(x, x_a|A) + (1-\mu)\,u(x, x_a|B)$$

or

$$\mu u(y, y_b|A) + (1-\mu)u(y, y_b|B) > \mu u(x, x_b|A) + (1-\mu)u(x, x_b|B)$$

or both. It follows that $\mathbf{y}$ strongly worst-case dominates $\mathbf{x}$. As strong worst-case dominance requires only one of the inequalities to hold without imposing any restrictions on their relative magnitude, we can see directly that it is not sufficient for noise dominance and so the converse is not true. □

**Result A.6:** *An action profile* $\mathbf{y} \in \mathbf{X}^*$ *strongly noise dominates* $\mathbf{x} \in \mathbf{X}^*$ *at* $\mu$ *only if* $\mathbf{y}$ *strongly worst-case dominates* $\mathbf{x}$ *and* $\mathbf{y}$ *noise dominates* $\mathbf{x}$ *at* $\mu$. *The converse is not true.*

*Proof.* Necessity follows from Result A.5 and Result 8. To show that NOD and strong WCD together are not equal to strong NOD, consider the following numerical example: $u(y, y_a|A) = 5$, $u(y, y_b|B) = 4$, $u(y, y_a|B) = u(y, y_b|A) = 0$ and $u(x, x_a|A) = 4$, $u(x, x_b|B) = 3$, $u(x, x_b|A) = 2$, $u(x, x_a|B) = 1$. For $\mu > \frac{1}{2}$, $\mathbf{y}$ strongly worst-case dominates $\mathbf{x}$. For $\pi(y) > \frac{3}{5}$, $\mathbf{y}$ also noise dominates $\mathbf{x}$ for the same range of $\mu$. However, again for $\mu > \frac{1}{2}$, $\mathbf{x}$ strongly noise dominates $\mathbf{y}$ contradicting sufficiency of strong WCD and NOD for strong NOD. □

*Proof of Result 10. Sufficiency:* Follows directly from the definition as it is the case for $\kappa = \frac{1}{2}$ and continuity guarantees that for small enough $\epsilon > 0$, this is also the case for some $\kappa_\epsilon = \frac{1}{2} + \epsilon$.

*Necessity:* Note that for $\mathbf{x}$ to be chosen at $p$, we need $V(P_x|p) > V(P_y|\mu)$ as the agent is unaware of the perception bias and chooses according to this ordering. The effect of a decrease in $\kappa$ (increase in misperception) is

$$\begin{aligned} -\frac{\partial V_n(P_x M_\kappa|p)}{\partial \kappa} = \quad & -p\cdot\big(2\pi(x)-1\big)\big[u(x, x_a|A) - u(x, x_b|A)\big] \\ & -(1-p)\cdot\big(2\pi(x)-1\big)\big[u(x, x_b|B) - u(x, x_a|B)\big] \end{aligned}$$

This does not depend on $\kappa$ itself. Therefore, if

$$\left[\frac{\partial V_n(P_x M_\kappa|p)}{\partial \kappa}\right]_{\kappa=1} > \left[\frac{\partial V[P_y M_\kappa|\,\mathbf{y}, p)}{\partial \kappa}\right]_{\kappa=1} \tag{12}$$

then the inequality also holds for all $\kappa \in [\frac{1}{2}, 1]$.

Take any $\kappa' \in [\frac{1}{2}, 1]$ such that $V(P_y M_{\kappa'}|\,\mathbf{y}, p) > V_n(P_x M_{\kappa'}|p)$. As $V(P_x|p) = V_n(P_x|p) > V_n(P_y|p) = V(P_y|p)$, we know that the inequality in 12 is satisfied and therefore $V(P_y M_\emptyset|\mathbf{y}, p) > V_n(P_x M_\emptyset|p) = V(P_x M_x|\,\mathbf{x}, p)$. $\mathbf{y}$ strongly noise dominates $\mathbf{x}$ at $\mu$ as desired. □

*Product launch details, section 7.* The expected utility of each alternative is

$$\begin{aligned}
E[U(z)] &= 0 \\
E[U(r)] &= \mu \cdot 0.8 + (1-\mu) \cdot (-1.2) = 2\mu - 1.2 \\
E[U(e)] &= \mu + (1-\mu) \cdot (-4) = 5\mu - 4.
\end{aligned}$$

Hence, the optimal choice $x$ for a given $\mu$ is

$$x(\mu) = \begin{cases} z & \text{if} \quad \mu \leq 0.6 \\ r & \text{if} \quad 0.6 < \mu \leq 28/30 \\ e & \text{if} \quad 28/30 < \mu. \end{cases}$$

The expected utility of acquiring information is

$$E[U(\text{info})] = \mu \cdot 1 + (1-\mu) \cdot 0 - 1/2 = \mu - 1/2$$

It follows the CEO prefers acquiring information for $\mu \in (0.5, 0.7]$. The CEO's true expected utility of information given her misperception bias is

$$\mu \cdot 1 + (1-\mu) \cdot \big(-4\delta + (1-\delta) \cdot 0\big) - 1/2$$

For any initial belief in $p = \mu \in (0.6, 0.7]$, the CEO benefits from holding $\mu \in [0.7, 28/30)$ instead whenever $\delta > \frac{0.7-\mu}{4(1-\delta)}$. To see this, note that for $\mu > 0.6$, the CEO prefers $r$ over $z$. Information is worse when $\mu - 4\delta \cdot (1-\mu) - 1/2 < 2\mu - 1.2$ or $\delta > \frac{0.7-\mu}{4(1-\delta)}$. Repeating the calculation for $\mu \in (0.5, 0.6]$ results in $\mu \leq 0.5$ being optimal whenever $\delta > \frac{\mu - 1/2}{4(1-\delta)}$. □

# REFERENCES


Roland Benabou and Jean Tirole. Self-confidence and personal motivation. *Quarterly Journal of Economics*, 117(3):817–915, 2002.

Jean-Pierre Benoit and Don A. Moore. Does the better-than-average effect show that people are overconfident? Two experiments. *Journal of the European Economic Association*, 13(2):293–329, 2015.

David Blackwell. Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 93–102. University of California Press, 1951.

Michael Blastland, Alexandra LJ Freeman, Sander van der Linden, Theresa M Marteau, and David Spiegelhalter. Five rules for evidence communication. *Nature Publishing Group*, 2020.

Jerome S. Bruner and Mary C. Potter. Interference in visual recognition. *Science*, 144 (3617):424–425, 1964.

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

Stephen V. Burks, Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini. Overconfidence and social signalling. *The Review of Economic Studies*, 80(3):949–983, 2013.

Juan D. Carrillo and Thomas Mariotti. Strategic ignorance as a self-disciplining device. *The Review of Economic Studies*, 67(3):529–544, 2000.

Gary Charness, Aldo Rustichini, and Jeroen Van de Ven. Self-confidence and strategic behavior. *Experimental Economics*, 21(1):72–98, 2018.

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

Julian Conrads and Bernd Irlenbusch. Strategic ignorance in ultimatum bargaining. *Journal of Economic Behavior and Organization*, 92:104–115, 2013.

John M. Darley and Paget H. Gross. A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44(1):20–33, 1983.

Leonidas Enrique De La Rosa. Overconfidence and moral hazard. *Games and Economic Behavior*, 73:429–451, 2011.

David Eil and Justin M. Rao. The good news–bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 2 (3):114–138, 2011.

Nicholas Epley and Thomas Gilovich. The mechanics of motivated reasoning. *Journal of Economic Perspectives*, 30(3):133–40, September 2016.

Paul L Epner, Janet E Gans, and Mark L Graber. When diagnostic testing leads to harm: a new outcomes-based approach for laboratory medicine. *BMJ quality & safety*, 22(Suppl 2):ii6–ii10, 2013.

Baruch Fischoff, Paul Slovic, and Sarah Lichtenstein. Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4):552–564, 1977.

Danielle B Freedman. Towards better test utilization–strategies to improve physician ordering and their impact on patient outcomes. *EJIFCC*, 26(1):15, 2015.

Itzhak Gilboa and David Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989.

Faruk Gul. Unobservable investment and the hold-up problem. *Econometrica*, 69(2):343–376, 2001.

Chris Guthrie, Jeffrey Rachlinski, and Andrew Wistrich. Inside the judicial mind: Heuristics and biases. *Cornell Law Review*, 56:777–830, 2001.

Jack Hirshleifer. The private and social value of information and the reward to inventive activity. *The American Economic Review*, 61(4):561–574, 1971.

Augustin Landier and David Thesmar. Financial contracting with optimistic entrepreneurs. *Review of Financial Studies*, 22(1):177–150, 2009.

Heidi J Larson, Louis Z Cooper, Juhani Eskola, Samuel L Katz, and Scott Ratzan. Addressing the vaccine confidence gap. *The Lancet*, 378(9790):526 – 535, 2011.

Sarah Lichtenstein, Baruch Fischoff, and Phillips Lawrence. Calibration of probabilities: The state of the art to 1980. In Daniel Kahneman, Paul Slovic, and Amos Tversky, editors, *Judgement under uncertainty: Heuristics and biases*, pages 306–334. Cambridge University Press, Cambridge, 1982.

Sandra Ludwig, Philipp C. Wichardt, and Hanke Wickhorst. Overconfidence can improve an agent's relative and absolute performance in contests. *Economic Letters*, 110:193–196, 2011.

Ulrike Malmendier and Geoffrey Tate. CEO overconfidence and corporate investment. *Journal of Finance*, 60(6):2661–2700, 2005.

Jacob Marschak and Koichi Miyasawa. Economic comparability of information systems. *International Economic Review*, 9(2):137–174, 1968.

Markus M. Mobius, Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. Managing self-confidence. *working paper*, 2014.

Don A. Moore and Paul J. Healy. The trouble with overconfidence. *Psychological Review*, 115(2):502–517, 2008.

Matthew Motta, Timothy Callaghan, and Steven Sylvester. Knowing less but presuming more: Dunning-Kruger effects and the endorsement of anti-vaccine policy attitudes. *Social Science & Medicine*, 211:274–281, 2018.

Gregory A. Poland and Robert M. Jacobson. Understanding those who do not understand: a brief review of the anti-vaccine movement. *Vaccine*, 19(17):2440 – 2445, 2001.

Anders U. Poulsen and Michael W. M. Roos. Do people make strategic commitments? experimental evidence on strategic information avoidance. *Experimental Economics*, 13(2):206–225, 2010.

Mathew Rabin and Joel Schrag. First impressions matter: A model of confirmatory bias. *Quarterly Journal of Economics*, 114(2):37–82, 1999.

William P. Rogerson. Contractual solutions to the hold-up problem. *Review of Economic Studies*, 59(4):777–793, 1992.

Thomas C. Schelling. An essay on bargaining. *American Economic Review*, 46(3):281–306, 1956.

Thomas C. Schelling. *The Strategy of Conflict*. Harvard University Press, Cambridge, MA, 1960.

Victor Stango and Jonathan Zinman. We are all behavioral, more or less: A taxonomy of consumer decision making. *NBER Working Paper No. 28138*, 2020.

Jakub Steiner and Colin Stewart. Perceiving prospects properly. *American Economic Review*, 106(7):1601–1631, 2016.

Jean Tirole. Procurement and renegotiation. *Journal of Political Economy*, 94(2):235–259, 1986.

Neil D. Weinstein. Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39(5):806–820, 1980.

Penny F Whiting, Clare Davenport, Catherine Jameson, Margaret Burke, Jonathan AC Sterne, Chris Hyde, and Yoav Ben-Shlomo. How well do health professionals interpret diagnostic information? A systematic review. *BMJ open*, 5(7), 2015.