Braghieri, Luca

Working Paper

Biased Decoding and the Foundations of Communication

CESifo Working Paper, No. 10432

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Braghieri, Luca (2023) : Biased Decoding and the Foundations of Communication, CESifo Working Paper, No. 10432, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/279181

CESifo WORKING PAPERS

10432
2023

May 2023

Biased Decoding and the Foundations of Communication

Luca Braghieri

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10432*

Biased Decoding and the Foundations of Communication

Abstract

One-way communication between an informed sender and an uninformed receiver involves two fundamental processes: a process of encoding – whereby the sender maps states of the world or concepts into arbitrary signals – and a process of decoding – whereby the receiver makes inferences about the state of the world conditional on each signal realization. In this paper, I develop machinery to study the process of decoding for an agent who might have inaccurate beliefs about the information environment (*a biased decoder*) and show how such machinery can help shed light on foundational aspects of communication.

Luca Braghieri
Bocconi University / Milan / Italy
luca.braghieri@unibocconi.it

April 26, 2023
I thank Alex Bloedel, Sarah Eichmeyer, Mira Frick, Yannai Gonczarowski, Ryota Iijima, Massimo Marinacci, Clément Mazet-Sonilhac, Stephen Morris, and Marco Ottaviani for very helpful comments. I thank Joel Sobel for kindly giving me feedback on an early draft of the paper and for a lengthy and insightful email exchange. I thank Gleb Kozliakov for excellent research assistance and Shuige Liu for carefully proofreading the paper. All remaining mistakes are my own.

1 Introduction

Communication is fundamental to any entity that is capable of conditioning its behavior on the realization of a signal, including humans, non-human animals, and mechanical objects such as computers. Even in its simplest and most stylized form, namely one-way communication between a sender and a receiver, communication involves two distinct processes: a process of encoding, which describes the way in which the sender maps states of the world or concepts into arbitrary signals, and a process of decoding, which describes, for each possible signal realization, the receiver's inferences about the state of the world (Locke, 1689; De Saussure, 1916; Shannon, 1948; Crawford and Sobel, 1982).

The encoding process has been studied extensively. It is a cornerstone of Shannon's information theory, of Blackwell's work on the comparison of experiments, and of many economic models ranging from signaling, to cheap talk, to persuasion and information design (Shannon, 1948; Blackwell, 1951, 1953; Spence, 1973; Crawford and Sobel, 1982; Kamenica and Genzkow, 2011). In contrast, the decoding process has been studied less extensively.[1] Since the transmission of information during communication depends on both the encoding and the decoding processes, and since communication is often defined as information exchange (Merriam Webster, 2023), a deeper understanding of the decoding process might help shed light on foundational aspects of communication.

In this paper, I first develop machinery to study the process of decoding for an agent who might have inaccurate beliefs about the information environment (a *biased decoder*) and, then, I apply such machinery to the study of communication. As a result, the paper is divided into two broad parts. The overarching goal of the first part of the paper is to develop a language to describe biased decoding and its consequences for the agent's welfare. I obtain such language by taking canonical objects in decision theory – such as the value of information, the value of perfect information, and a host of measures of information transmission – and extending them to encompass the possibility of biased decoding. Thus, the milestones in the first part of the paper are: i) a notion of the value of information for a biased decoder (B-EVSI); ii) a set of propositions describing key features of B-EVSI; iii) a notion of the value of perfect information for a biased decoder (B-EVPI); iv) the characterization of a set of measures of *meaningful information transmission* that capture both the features of the signal structure (encoding) and the agent's interpretations (decoding). In the second part of the paper, I deploy the machinery developed in the first part to study communication. The milestones in the second part of the paper are a set of definitions and propositions regarding foundational concepts in communication such as successful communication, vagueness, misunderstandings, deception, lying, and thwarted deception.

A simple example might help build intuition about the importance of the decoding process for communication. Consider a driver who would like to enter a parking garage in Germany and

[1] There are some notable exceptions; for instance, Morris and Shin (1997) and Frick, Iijima, and Ishii (2022).

imagine the driver sees two signs: one that says *Einfahrt* (entrance) and one that says *Ausfahrt* (exit). If the driver is proficient in German, she will correctly interpret the signs and enter the parking garage through the entrance. If the driver does not understand German, she might park on the side of the street and look up the translation on her phone before entering the parking garage. Lastly, if the driver does not understand German but mistakenly thinks she does, the following two questions become pressing. First, how can we think about and measure the instrumental value of observing the two signs for such a driver? Second, how can we think about and measure the amount of information that such a driver extracts from the two signs? Clearly, the driver's subjective assessment of the instrumental value of the signs or of the amount of information contained in them need not provide reliable guidance. After all, the driver might deliberately enter the parking garage through the *Ausfahrt* gate thinking it is the entrance and crash against an oncoming car. Similarly, any measure of the instrumental value of the signs or of the amount of information contained in them that ignores the driver's misperceptions will be misleading. After all, the signs are perfectly informative only to someone who understands German. Therefore, in order to study the instrumental value of observing a signal and the amount of information that an agent extracts from a signal, it is necessary to consider, besides the features of the signal structure (encoding), the agent's possibly incorrect interpretations (decoding).

The first part of the paper, namely the one aimed at developing a language to describe biased decoding and its welfare consequences, is structured around two guiding questions: i) how can we think about and measure the instrumental value of information for a biased decoder? ii) How can we think about and measure the amount of information that a biased decoder extracts from a signal? Such questions, which I believe are essential to the study of communication, can be asked more generally about any agent who samples from a signal structure before facing a standard decision problem under uncertainty.[2] Therefore, for the sake of generality, the first part of the paper abstracts away from communication and considers the broader problem of decision-making under uncertainty.

The basic setup in the first part of the paper is thus that of a standard decision problem under uncertainty. After the agent makes her decision, one out of a set of mutually exclusive and exhaustive states of the world is realized. The agent's payoff depends both on the realized state of the world and on her decision. The agent observes a potentially informative signal about the state of the world and can condition her decision on the observed signal. When making her decision, the agent is a subjective expected utility maximizer; i.e. she maximizes her expected utility, with the expectation taken according to her subjective beliefs about the distribution of states of the world.

The non-standard aspect of the setup involves the agent's beliefs. First, the agent's prior beliefs about the distribution of states of the world need not equal the actual probability distribution of

[2] I use the terms signal structure, signal-generating process, Blackwell experiment, and encoding mapping interchangeably. Similarly, I use the terms decoding strategy and interpretations interchangeably.

states of the world. Second, the agent's beliefs about the features of the signal-generating process – i.e., the conditional distribution of signals given states of the world – need not equal the actual features of the signal-generating process. Third, when updating her beliefs upon observing the realization of a signal, the agent need not follow Bayes' rule. I refer to such an agent as a *biased decoder*.[3] Importantly, I assume that a biased decoder has no metacognition of the fact that her beliefs about the information environment might be incorrect.[4]

In order to begin answering the question "how can we think about and measure the instrumental value of information for a biased decoder?", I introduce the notion of *expected value of sample information for a biased decoder* (B-EVSI).[5] I define B-EVSI as the difference between the expected utility that the agent obtains if she is allowed to condition her action on the realization of the signal and the expected utility that she obtains if she is not allowed to do so. Rather than being evaluated from the perspective of the agent's subjective beliefs, however, the expectations are evaluated according to the actual distribution of states of the world and to the actual features of the signal-generating process. Therefore, in this framework, the agent takes actions that are optimal according to her subjective beliefs, but the welfare consequences of those actions are evaluated from the perspective of an external observer who has knowledge of the true probabilities.[6]

A simple decomposition brings into sharp relief the commonalities and differences between B-EVSI and the canonical expected value of sample information for an unbiased decoder (EVSI). As far as commonalities are concerned, B-EVSI nests EVSI as a special case: unsurprisingly, whenever the agent is Bayesian, has accurate prior beliefs about the distribution of states of the world, and has a correct understanding of the signal-generating process, B-EVSI reduces to EVSI. As far as differences are concerned, B-EVSI can be strictly larger than EVSI – whenever observing the realization of the signal mitigates the agent's prior misperceptions about the distribution of states of the world – or it can be strictly negative – whenever the agent is severely misled by the signal.

The observation that, in contrast to the value of information for an unbiased decoder, B-EVSI can be strictly negative prompts an exploration of the settings in which the ability to condition her action on the realization of a signal makes a biased decoder better or worse off. I begin by characterizing the set of beliefs that guarantee a non-negative B-EVSI in all decision problems that the biased decoder might encounter in the simple case in which both the state of the world

[3] An agent would also be a biased decoder if there was noise in the communication channel and the agent mistakenly thought the channel was noiseless or that it featured a different type of noise. In this paper, I abstract away from the possibility of there being noise in the communication channel.

[4] A different set of questions arises when the agent is assumed to have some metacognition of possibly having incorrect beliefs. See, for instance, Cerreia-Vioglio et al. (2022).

[5] The nomenclature Expected Value of Sample Information (EVSI) and Expected Value of Perfect Information (EVPI) is from statistical decision theory. Outside statistical decision theory, EVSI is generally referred to simply as the value of information and EVPI is generally referred to simply as the value of perfect information.

[6] The idea of "true probabilities" might make some readers uncomfortable. Happily, one can coherently replace "true probabilities" with "beliefs of a third-party Bayesian observer" virtually everywhere in the paper and interpret the results as relating to the perspective of a third-party Bayesian observer rather than to the "true probabilities."

and the signal are binary. Specifically, a necessary and sufficient condition for B-EVSI to be non-negative in all decision problems is that, for each signal, the biased decoder's posterior beliefs after observing the signal be a convex combination of the true posterior probabilities after the signal realization and the true prior. Such result relates to the discussion in experimental economics about over- and under-inference in belief updating by showing that over-inference has robustly more pernicious consequences for the agent's welfare then under-inference (Edwards, 1968; Benjamin 2019; Augenblick, Lazarus, and Thaler, 2023; Ba, Bohren, and Imas 2023).

The characterization above helps shed new light on binary signals that can be ranked according to the canonical Blackwell order. In the context of decision problems with a binary state of the world and a binary signal, not only do Blackwell-more-informative signal structures yield a higher value of information for unbiased decoders; if the biased decoder knows which signal is diagnostic of which state of the world, Blackwell-more-informative signal structures are also robustly welfare improving for biased decoders. Specifically, they induce a negative B-EVSI for a smaller set of biased beliefs and they increase the biased decoder's welfare holding her biased beliefs fixed.

In larger state and signal spaces, I characterize the set of biased beliefs that guarantee a non-negative B-EVSI in all decision problems for a biased decoder who knows the prior distribution of states of the world and is Bayesian. The key qualitative feature that emerges from the general characterization is, once again, an asymmetry in the effects of holding posterior beliefs that are too extreme versus too conservative. Specifically, whenever the agent holds subjective posterior beliefs that are more extreme than the true posterior probabilities, there always exists a decision problem in which B-EVSI is strictly negative. Conversely, whenever the agent holds subjective posterior beliefs that are more conservative than the true posterior probabilities, B-EVSI might be non-negative in all decision problems that the agent encounters.

I conclude my investigation of the question "how can we think about and measure the instrumental value of information for a biased decoder?" by introducing the notions of *expected value of perfect information for a biased decoder* (B-EVPI) and of *absolute welfare loss for a biased decoder* (B-Loss). B-EVPI is defined as the difference between the expected utility that the agent obtains if she is allowed to condition her action directly on the realized state of the world and the expected utility that she obtains if she has to rely solely on her prior information. B-Loss is defined as the difference between B-EVPI and B-EVSI and has the appealing feature of being additively separable across two components. The first component relates exclusively to the features of the signal-generating process and measures the welfare loss resulting from the signal structure providing less than full information even to an unbiased decoder. The second component relates to the agent's interpretations and measures the welfare loss resulting from the biased decoder taking suboptimal actions as a result of her biased beliefs. In environments in which the signal structure is endogenously determined by a sender via the encoding process, the additively separable nature of the measure of absolute welfare loss allows one to decouple the amount of welfare that the receiver

loses as a result of the encoding mapping providing less than full information and the amount of welfare that she loses as a result of her biased beliefs.

The notions of B-EVSI and B-EVPI allow me to address the second guiding question in the first part of the paper, namely "how can we think about and measure the amount of information that a biased decoder extracts from a signal?" Specifically, B-EVSI and B-EVPI naturally give rise to a set of measures of the amount of information generated by signal for a potentially biased decoder and of the uncertainty implicit in a biased decoder's beliefs. I refer to such measures as *meaningful information* and *meaningful uncertainty*, because they take into account the objective features of the information environment as well as the interpretations/meaning that the agent assigns to observed signals. I show that meaningful information measures extend all canonical information measures that correspond to the instrumental value of information in some decision problem that the agent might encounter (Frankel and Kamenica, 2019). For instance, they extend Shannon's notion of information transmission – mutual information – to contexts in which the receiver might misinterpret the information encoded by the sender (Shannon, 1948). Similarly, I show that meaningful uncertainty measures extend all canonical uncertainty measures that capture the extent to which, in some decision problem, holding a certain belief reduces a decision-maker's utility relative to an omniscient benchmark (Frankel and Kamenica, 2019). For instance, they extend Shannon's entropy function to contexts in which the agent has biased beliefs. Online Appendix C complements the theoretical exploration with two illustrations of how to estimate meaningful information measures on empirical data.

In the second part of the paper, I employ the machinery developed in the first part to study one-way communication between an informed sender and an uninformed receiver. I show how concepts such as *successful communication*, *vagueness*, *misunderstandings*, *deception*, *lying*, and *thwarted deception* can be analyzed by considering the relationship between the sender's encoding mapping, the receiver's decoding strategy, and the sender's beliefs about the receiver's decoding strategy.

I define successful communication as occurring whenever: i) the sender employs an encoding mapping that she correctly believes the receiver will decode accurately, and ii) there is a signal realization that induces the receiver to update her beliefs about the state of the world. As a result, the value of communication to the receiver and the amount of meaningful information transmitted is determined by the features of the sender's encoding mapping.[7] Equilibria of game-theoretic models of communication other than babbling equilibria can be thought of as studying successful communication.

Misunderstandings occur whenever the sender inadvertently employs an encoding mapping that the receiver misinterprets. Therefore, whenever communication involves a misunderstanding, the receiver is a biased decoder. The results about B-EVSI developed in the first part of the pa-

[7] In the context of communication, I refer to the expected value of sample information for a biased decoder as the value of communication.

per show that, compared to a counterfactual scenario in which the receiver interprets messages correctly, misunderstandings always lower the receiver's welfare and the amount of meaningful information transmitted. Furthermore, they show that, compared to a counterfactual scenario of no communication, the implications of misunderstandings are ambiguous. Some misunderstandings – for instance the ones that occur whenever the agent underestimates the precision of a signal and develops posterior beliefs that are more conservative than the true posterior probabilities – never harm the receiver's welfare and never lead to a negative amount of meaningful information transmitted; other misunderstandings always do.

Deception occurs whenever the sender deliberately and successfully sends a message that leads the receiver to develop beliefs about the realized state of the world that, compared to her prior beliefs, are less accurate. As in the case of misunderstandings, the receiver is a biased decoder as a result of deception. From the point of view of the receiver's welfare and of the amount of meaningful information transmitted, the effects of deception are more pernicious than the effects of misunderstandings. In particular, the results about B-EVSI developed in the first part of the paper show that, whenever the receiver is deceived, there always exists a decision problem in which the value of communication to the receiver is strictly negative and a measure of meaningful information transmission for which the amount of information transmitted is strictly negative. In the context of a binary state of the world and a binary signal, the negative effects of deception on the receiver's welfare are particularly stark: deception leads to a negative value of communication to the receiver in all decision problems and to a negative amount of meaningful information transmitted for all measures of meaningful information characterized in this paper.

Following Sobel (2020), I define lying in the context of a conventional language in which messages have a commonly understood meaning and refer to subsets of the state space. Lying occurs whenever the sender deliberately makes a statement that, in light of the realized state of the world and according to the statement's conventional meaning, is false. I show that, in the presence of a conventional language, if the sender correctly believes that the receiver interprets messages according to the linguistic convention, deception and lying are equivalent.

This paper contributes to various strands of the literature. One strand is the economic literature about biased learning and misspecified models (Berk, 1966; Nyrko, 1991; Fudenberg, Romanyuk, and Strack, 2017; Heidhues, Koszegi, and Strack 2018, 2021; He, 2022; Bushong and Gagnon-Bartsch, 2022; Bohren and Hauser, 2021; Esponda, Pouzo, and Yamamoto, 2021; Fudenberg, Lanzani, and Strack, 2021, Frick, Iijima, and Ishii, 2022). The closest paper in that literature is Morris and Shin (1997). Morris and Shin (1997) consider a Bayesian decision maker who has correct prior beliefs about the distribution of states of the world, but who might misperceive the features of the signal-generating process. In that context, the authors introduce a version of B-EVSI and characterize the settings in which B-EVSI is guaranteed to be non-negative independently of the agent's decision problem. I expand on the work of Morris and Shin (1997) in three main ways:

first, only a relatively small subset of the results developed in this paper overlaps with the ones in Morris and Shin (1997). Specifically, the decomposition of B-EVSI, the possibility that B-EVSI is strictly larger than EVSI, the conditions under which B-EVSI is non-positive in all decision problems, the implications for Blackwell informativeness, the notions of B-EVPI and B-Loss, the measures of meaningful information transmission and meaningful uncertainty, the implications for communication, and the discussion of the empirical measurement of meaningful information transmission have no counterpart in Morris and Shin (1997). Second, I consider a broader set of decoding biases than Morris and Shin (1997) and, in the case of a binary state of the world and a binary signal, I characterize the conditions for B-EVSI to be non-negative in the context of such broader set. In particular, in contrast to Morris and Shin (1997), I allow the agent's prior beliefs to be misspecified and the agent to be non-Bayesian. Third, the general characterization in Morris and Shin (1997) involves mathematical objects that are not easy to interpret; in this paper, I provide an equivalent set of necessary and sufficient conditions that, arguably, have a more straightforward interpretation.

This paper also relates the strand of the economics literature that characterizes various measures of information and uncertainty. A set of papers close to the rational inattention literature characterizes measures of uncertainty that reflect the costs of information acquisition (Sims, 2003; Hébert and Woodford, 2022; Morris and Strack, 2019; Mensch 2021; Pomatto, Strack and Tamuz, 2023). Another set of papers studies measures of information transmission that reflect the benefits of information acquisition (Frankel and Kamenica, 2019). This paper complements the work of Frankel and Kamenica (2019) by extending all the measures of information and uncertainty characterized in their paper, including mutual information and entropy, to the case of biased decoding.

This paper also contributes to the economics literature on communication (Crawford and Sobel, 1982; Milgrom and Roberts, 1986; Kamenica and Gentzkow, 2011, Kartik, 2009; Kartik, Ottaviani, and Squintani, 2007; Farrell and Rabin, 1996; Farrell, 1993; Ottaviani and Sorensen, 2006; Morris, 2001; Green and Stokey, 2007, Sobel, 2020; Abeler, Nosesnzo, and Raymond, 2019). In contrast to most models of communication in economics, the framework developed in this paper aims to provide a language to study communication out of equilibrium and in the presence of decoding biases. The closest papers in this strand of the literature are the ones by Sobel (2020,2023), who studies lying and deception in the context of one-way communication. I discuss the differences between my approach and Sobel's at the end of Section 5.

Online Appendix B describes the paper's relation and contribution to the literature outside of economics, namely in statistics, information theory, artificial intelligence, and mathematical biology. The appendix also contains a detailed discussion of the relationship between the theory of communication presented in this paper and the one in Shannon (1948).

The remainder of this paper is organized as follows: Section 2 introduces the basic framework to study the decoding process. Section 3 introduces the notions of B-EVSI, B-EVPI, and B-Loss, and

analyzes them. Section 4 introduces the notions of meaningful information transmission, meaningful uncertainty, and meaningful information loss. Section 5 discusses the implications for theories of communication. Section 6 concludes.

Most proofs can be found in the Appendix at the end of the manuscript; some of the more tedious and less insightful proofs are relegated to Online Appendix A.

2 Setup

2.1 Information Environment and Beliefs

Let $\Omega = \{\omega_1, ..., \omega_n\}$ denote a set of mutually exclusive and exhaustive states of the world. A state of the world is drawn once and for all according to a distribution $p \in \Delta(\Omega)$ with full support.[8] The agent's beliefs about the prior distribution of states of the world are denoted by $\beta \in \Delta(\Omega)$ and are assumed to have full support as well. The agent's beliefs β may be inaccurate, in the sense that they need not equal the actual distribution of states of the world p.

The agent does not observe $\omega \in \Omega$ directly, but instead observes a signal that might reveal information about ω. A signal structure $\left(\mathcal{S}, \{m|\omega\}_{\omega\in\Omega}\right)$, which I sometimes refer to as a Blackwell experiment or as an encoding mapping, consists of a finite signal space $\mathcal{S} = \{s_1, ..., s_k\}$ and a set of conditional distributions $\{m|\omega\}_{\omega\in\Omega}$, with $m|\omega \in \Delta(\mathcal{S})$ denoting the probability distribution over the space of signals conditional on the state of the world being $\omega \in \Omega$. From p and $\{m|\omega\}_{\omega\in\Omega}$, it is easy to recover the unconditional distribution of signals $m \in \Delta(\mathcal{S})$. Similarly, it is easy to recover $p|s \in \Delta(\Omega)$: the probability distribution over states of the world conditional on the signal realization being $s \in \mathcal{S}$. Denote the collection of such probability distributions, one per possible signal realization, by $\{p|s\}_{s\in\mathcal{S}}$.

The agent might incorrectly perceive the signal structure and/or might not be Bayesian. Denote the agent's posterior beliefs about the state of the world after observing signal s by $\beta|s \in \Delta(\Omega)$. Let the collection of such posterior beliefs, one per possible signal realization, be denoted by $\{\beta|s\}_{s\in\mathcal{S}}$. Let the agent's perception of the encoding mapping be denoted by $\{\mu|\omega\}_{\omega\in\Omega}$ and the agent's beliefs about the unconditional distribution of signals be denoted by $\mu \in \Delta(\mathcal{S})$. I sometimes refer to the agent's perception of the encoding mapping as the agent's decoding strategy.

Definition 1. The agent is a *biased decoder* if either $\beta \neq p$ or $\beta|s \neq p|s$ for at least one $s \in \mathcal{S}$. Otherwise, the agent is an *unbiased decoder*.

[8] From a frequentist standpoint, p can be thought of as a parameter estimated from multiple independent observations. From a Bayesian standpoint, p can be thought of as the beliefs of a third-party Bayesian observer.

2.2 Decision Problem

A decision problem $\mathcal{D}(\mathcal{A}, u)$ consists of an action set $\mathcal{A}$ and a utility function $u : \mathcal{A} \times \Omega \to \mathbb{R} \cup \{-\infty\}$.[9] I assume the existence of some action $a \in \mathcal{A}$ such that $u(a, \omega)$ is finite for every $\omega \in \Omega$. Furthermore, I adopt $-\infty \times 0 = 0$, so that actions that yield $-\infty$ in some state of the world do not affect utility if the state has or is perceived to have zero probability. I require the agent's problem, $\max_{a \in \mathcal{A}} E_{\omega \sim q}[u(a, \omega)]$, to admit a solution for all $q \in \Delta(\Omega)$.[10] Finally, I let $a_q \in \mathcal{A}$ denote an arbitrary element of $\arg\max_{a \in \mathcal{A}} E_{\omega \sim q}[u(a, \omega)]$.

3 Value of Information for a Potentially Biased Decoder

If the agent does not observe a signal about the state of the world, she chooses an action that is subjectively optimal given her prior beliefs; i.e., she chooses some action $a_\beta \in \arg\max_{a \in \mathcal{A}} E_{\omega \sim \beta}[u(a, \omega)]$. I assume that, whenever $\arg\max_{a \in \mathcal{A}} E_{\omega \sim \beta}[u(a, \omega)]$ is not unique at a certain belief $\beta \in \Delta(\Omega)$, the agent consistently picks the same action $a_\beta \in \mathcal{A}$. According to the actual probability distribution over states of the world $p \in \Delta(\Omega)$, the actual expected payoff from taking action a_β is $E_{\omega \sim p}[u(a_\beta, \omega)]$; i.e., the actual expected payoff from taking action a_β is simply the expected payoff of taking action a_β according to the actual distribution of states of the world p.

If the agent is allowed to condition her action on the signal realization, she takes an action that maximizes her subjective expected utility given the signal realization; i.e., if the signal realization is $s \in \mathcal{S}$, she takes some action $a_{\beta|s} \in \arg\max_{a \in \mathcal{A}} E_{\omega \sim \beta|s}[u(a, \omega)]$. According to the actual probability distribution over states of the world conditional on signal $s \in \mathcal{S}$, the agent's actual expected payoff of taking action $a_{\beta|s}$ is $E_{\omega \sim p|s}\left[u\left(a_{\beta|s}, \omega\right)\right]$. Taking an additional expectation over the space of signals and recalling that the unconditional distribution of signals is denoted by $m \in \Delta(\mathcal{S})$, the actual ex-ante expected payoff of observing the signal is $E_{s \sim m}\left[E_{\omega \sim p|s}\left[u\left(a_{\beta|s}, \omega\right)\right]\right]$. I refer to this object as the biased decoder's welfare. Specifically, $\mathcal{W} : \Delta(\Omega)^{2(|\mathcal{S}|+1)} \to \mathbb{R} \cup \{-\infty\}$ such that[11]

$$\mathcal{W}\left(p, \{p|s\}_{s \in \mathcal{S}}, \beta, \{\beta|s\}_{s \in \mathcal{S}}\right) = E_{s \sim m}\left[E_{\omega \sim p|s}\left[u\left(a_{\beta|s}, \omega\right)\right]\right]$$

[9] Working with the extended reals is necessary to capture some important decision problems and measures of uncertainty such as Shannon's entropy function.

[10] $E_{\omega \sim q}[\cdot]$ denotes the expectation taken with respect probability distribution $q \in \Delta(\Omega)$ over states of the world.

[11] The definition of the biased decoder's welfare involves a slight abuse of notation. In particular, the domain of the function should be $\Delta(\mathcal{S})^{|\Omega|} \times \Delta(\Omega)^{(|\mathcal{S}|+1)}$ rather than $\Delta(\Omega)^{2(|\mathcal{S}|+1)}$. This way, the function would take as input a prior distribution $p \in \Delta(\Omega)$, a signal structure $\{m|\omega\}_{\omega \in \Omega} \in \Delta(\mathcal{S})^{|\Omega|}$, and a set of posterior beliefs by the agent $\{\beta|s\}_{s \in \mathcal{S}} \in \Delta(\Omega)^{|\mathcal{S}|}$. The notation I adopted, despite being slightly less accurate, has the convenience of allowing me to write $\mathcal{W}\left(p, \{p|s\}_{s \in \mathcal{S}}, \beta, \{\beta|s\}_{s \in \mathcal{S}}\right)$ whenever I refer to a biased decoder and $\mathcal{W}\left(p, \{p|s\}_{s \in \mathcal{S}}, p, \{p|s\}_{s \in \mathcal{S}}\right)$ whenever I refer to an unbiased decoder. The same observation applies to the definition of the expected value of sample information for a potentially biased decoder. Notice also that $\mathcal{W}\left(p, \{p|s\}_{s \in \mathcal{S}}, \beta, \{\beta|s\}_{s \in \mathcal{S}}\right)$ implicitly assumes that, whenever $\arg\max_{a \in \mathcal{A}} E_{\omega \sim q}[u(a, \omega)]$ is not unique, there is a pre-specified way of selecting a_q from $\arg\max_{a \in \mathcal{A}} E_{\omega \sim q}[u(a, \omega)]$.

I am now in a position to define the value of information for a potentially biased decoder.

Definition 2. The *expected value of sample information for a potentially biased decoder (B-EVSI)* is a function $\mathcal{V}:\Delta(\Omega)^{2(|\mathcal{S}|+1)}\to\mathbb{R}\cup\{-\infty\}$ such that

$$\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)=E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{\beta|s},\omega\right)\right]\right]-E_{\omega\sim p}\left[u\left(a_{\beta},\omega\right)\right]$$

The next proposition shows that B-EVSI can be re-written in a way that brings a few important features into sharp relief. In order to state the proposition, however, it is necessary to first define the following mathematical object: $d\left(q'',q'\right)=E_{\omega\sim q''}\left[u\left(a_{q''},\omega\right)\right]-E_{\omega\sim q''}\left[u\left(a_{q'},\omega\right)\right]$. Intuitively, $d\left(q'',q'\right)$ captures the expected utility loss, evaluated according to the true probability distribution of states of the world $q''\in\Delta(\Omega)$, that the agent incurs by picking the subjectively optimal action $a_{q'}\in\mathcal{A}$ as opposed to the objectively optimal action $a_{q''}\in\mathcal{A}$. Mathematically, $d\left(q'',q'\right)$ is a Bregman divergence associated with decision problem $\mathcal{D}(\mathcal{A},u)$.[12] The following proposition provides a decomposition of B-EVSI that highlights a few important features of the value of information for a biased decoder:

Proposition 1. $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)$ *can be written as*

$$\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)=\left\{E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s},\omega\right)\right]\right]-E_{\omega\sim p}\left[u\left(a_{p},\omega\right)\right]\right\}+d\left(p,\beta\right)-E_{s\sim m}\left[d\left(p|s,\beta|s\right)\right]$$

where $d\left(p,\beta\right)\geq 0$ *with equality if* $\beta=p$, $d\left(p|s,\beta|s\right)\geq 0$ *with equality if* $\beta|s=p|s$, *and* $E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s},\omega\right)\right]\right]-E_{\omega\sim p}\left[u\left(a_{p},\omega\right)\right]\geq 0$ *with equality if* ω *and* s *are independent.*

The decomposition in Proposition 1 can be interpreted as follows: the first component in the expression above, namely $E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s},\omega\right)\right]\right]-E_{\omega\sim p}\left[u\left(a_{p},\omega\right)\right]$, is simply the value of information (EVSI) for an unbiased decoder. $d\left(p,\beta\right)$ captures the utility loss that the agent would have incurred as a result of her misperceptions about the information environment had she made a decision prior to observing the realization of the signal. $E_{s\sim m}\left[d\left(p|s,\beta|s\right)\right]$ captures the expected utility loss that the agent incurs as a result of her misperceptions and/or non-Bayesian updating after observing the signal realization. The first two components enter the expression positively and measure the maximum amount of utility that the agent could in principle gain from correctly interpreting signal $s\in\mathcal{S}$. The third component enters the expression negatively and measures the utility loss to the agent caused by taking sub-optimal actions as a result of holding incorrect posterior beliefs.

[12] Let $G\left(q''\right)=E_{\omega\sim q''}\left[u\left(a_{q''},\omega\right)\right]$, which, by the definition of $a_{q''}$, equals $\max_{a\in\mathcal{A}}E_{\omega\sim q''}\left[u\left(a,\omega\right)\right]$. It is easy to see that $G\left(q''\right)$ is a convex function. Consider $q'\in\Delta(\Omega)$. Since $G\left(q''\right)$ is convex, we can define the Bregman divergence from q'' to q'. A Bregman divergence of G is a function $\tilde{d}:\left(\Delta(\Omega)\right)^2\to\mathbb{R}$ defined as $\tilde{d}\left(q'',q'\right)=G\left(q''\right)-G\left(q'\right)-\nabla G\left(q'\right)\cdot\left(q''-q'\right)$, where $\nabla G\left(q'\right)$ is some subgradient of G at q' (Bregman, 1967). $d\left(q'',q'\right)=E_{\omega\sim q''}\left[u\left(a_{q''},\omega\right)\right]-E_{\omega\sim q''}\left[u\left(a_{q'},\omega\right)\right]$ is the Bregman divergence from q'' to q' that uses the subgradient associated with action $a_{q'}\in\mathcal{A}$. Notice that, since $G\left(q''\right)$ is a convex function, it admits a subgradient at every $q''\in\Delta(\Omega)$ (one can define a subgradient also at boundary points as shown in Frankel and Kamenica, 2019). See Dawid (2007) for additional details.

The decomposition in Proposition 1 highlights three features of B-EVSI:

Remark 1.

1. The value of information for a potentially biased decoder (B-EVSI) reduces to the canonical notion of the value of information (EVSI) whenever the agent is an unbiased decoder.

2. There exist decision problems and misperceptions such that the value of information for a biased decoder (B-EVSI) is strictly larger than the value of information for an unbiased decoder (EVSI); i.e., there exist decision problems and misperceptions such that

$$\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) > \mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, p, \{p|s\}_{s\in\mathcal{S}}\right)$$

3. There exist decision problems and misperceptions in which the value of information for a biased decoder (B-EVSI) is strictly negative.

The first statement, which follows trivially from the definition of the value of information for a potentially biased decoder, is unsurprising but worthwhile pointing out explicitly. The second result can occur, for instance, whenever the agent is Bayesian, has an incorrect but non-dogmatic prior, has an accurate understanding of the signal-generating process, and the signal-generating process is fully informative.[13] The third result can occur, for instance, when the agent has a correct prior, the signal is utterly uninformative, but the agent mistakenly thinks the signal is informative.

The following example will help build intuition.

Example 1. Consider once again the plight of the driver who would like to enter a parking garage in Germany. Let there be two gates, one of which is the entrance to the garage and one of which is the exit. Let ω denote whether the rightmost gate from the perspective of the driver is the entrance or the exit $\omega \in \{entrance, exit\}$. As discussed, the driver sees two signs: one that says *Einfahrt* and one that says *Ausfahrt.* Let s denote the sign on top of the rightmost gate from the perspective of the driver. Then, $s \in \{Einfahrt, Ausfahrt\}$. Imagine that, across all parking garages in Germany, the true conditional distribution of signals given states is $m_{entrance}(Einfhart) = 1$, $m_{entrance}(Ausfhart) = 0$, $m_{exit}(Einfhart) = 0$, and $m_{exit}(Ausfhart) = 1$. Let the driver's action be $a \in \{0, 1\}$, where $a = 1$ means entering the garage from the rightmost gate from her perspective. Let $u(1, Entrance) = 1$, $u(1, Exit) = 0$, $u(0, Entrance) = 0$, $u(0, Exit) = 1$. Lastly, let $p = 0.8$, meaning that, in 80% of German parking garages, the rightmost gate is the entrance, which is consistent with people driving on the right side of the road. Therefore, before observing the signal, an unbiased decoder takes action $a = 1$ and her expected utility is $E_{\omega\sim p}[u(a_p, \omega)] = 0.8 \times 1 + 0.2 \times 0 = 0.8$. Furthermore, for an unbiased decoder, the expected value of sample information is $E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s}, \omega\right)\right]\right] - E_{\omega\sim p}[u(a_p, \omega)] = (0.8 \times 1 + 0.2 \times 1) - (0.8 \times 1 + 0.2 \times 0) = 0.2$.

[13] See Section 3.3 for a formal definition of a fully informative encoding.

Now suppose the driver is from the U.K. and, mistakenly, thinks that, in most German parking garages, the leftmost gate is the entrance as it generally is in the U.K. Specifically, let $\beta = 0.2$. Therefore, before observing the signal, the driver enters the garage through the leftmost gate and her expected utility is $E_{\omega\sim p}\left[u\left(0,\omega\right)\right] = 0.8 \times 0 + 0.2 \times 1 = 0.2$. Let us consider the expected value of sample information for the British driver in two different cases:

- Case 1: the driver understands German. If the driver understands German, she correctly interprets the signs even though she has an incorrect prior. Therefore, the expected value of sample information for the driver is $E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{\beta|s},\omega\right)\right]\right] - E_{\omega\sim p}\left[u\left(a_{\beta},\omega\right)\right] = (0.8 \times 1 + 0.2 \times 1) - (0.8 \times 0 + 0.2 \times 1) = 0.8$. In this case, B-EVSI is strictly larger than EVSI; in other words, the biased decoder benefits even more than the unbiased decoder from the opportunity to condition her action on the signal realization. The reason is that observing the realization of the signal prevents the British driver from basing her decisions on her erroneous prior beliefs.

- Case 2: the driver mistakenly thinks that *Ausfahrt* means entrance. Then, the driver will pick action $a = 0$ whenever she sees a sign *Einfahrt* and $a = 1$ whenever she sees the sign *Ausfahrt*. In this case, therefore, the expected value of sample information for the driver is $E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{\beta|s},\omega\right)\right]\right] - E_{\omega\sim p}\left[u\left(a_{\beta},\omega\right)\right] = (0.8 \times 0 + 0.2 \times 0) - (0.8 \times 0 + 0.2 \times 1) = -0.2$. In this case, B-EVSI is strictly negative: sampling from the signal structure further misleads the British driver and leads her to take even more sub-optimal actions after observing a signal realization than before.

In light of the example above, it is natural to wonder about the circumstances in which the ability to condition her actions on the realization of a signal makes the biased decoder worse off. In other words, when does information hurt the welfare of a biased decoder? Note that, for an unbiased decoder, the issue does not arise: to the extent that observing a signal is free, an unbiased decoder is always better off taking actions after observing the realization of a signal about the state of the world than before observing it (Blackwell, 1951, 1953). In the next section, I characterize situations in which B-EVSI is non-negative in all decision problems and situations in which it is non-positive in all decision problems. I begin by analyzing the simple case of a binary state of the world and a binary signal and then extend the results to larger state and signal spaces.

3.1 Binary Environment

Let $\omega \in \{\omega_1, \omega_2\}$, $s \in \{s_1, s_2\}$. Without loss of generality, let signal s_1 be diagnostic of state of the world ω_1; i.e., let $m|\omega_1\left(s_1\right) \geq m|\omega_2\left(s_1\right)$. Therefore, $p|s_2\left(\omega_1\right) \leq p\left(\omega_1\right) \leq p|s_1\left(\omega_1\right)$.

Before proceeding, it is useful to introduce the following intuitive definition:

Definition 3. An agent is said to *interpret the signal as noise* if for every $s \in \mathcal{S}$ and for every $\omega \in \Omega$, $\beta|s\left(\omega\right) = \beta\left(\omega\right)$.

The following assumption imposes a modicum of rationality on the agent's beliefs:

Assumption 1. *If the agent does not interpret the signal as noise, there exists an* $\alpha \in (0,1)$ *such that* $\beta = \alpha\beta|s_1 + (1-\alpha)\,\beta|s_2$

For a Bayesian agent, Assumption 1 is guaranteed by the martingale property of beliefs. In fact, for a Bayesian agent, α in the expression above equals the unconditional distribution of observing signal s_1. For a biased decoder, α need not equal the unconditional distribution of observing signal s_1. Still, Assumption 1 requires that if the biased decoder believes that signals s_i is diagnostic of state of the world ω_j, she must believe that signal s_{-i} is diagnostic of state of the world ω_{-j}. For instance, if a biased decoder takes a diagnostic test for a disease and if her belief about having the disease increases after testing positive, then her belief about having the disease should decrease after testing negative. I maintain Assumption 1 throughout this section.

The following proposition characterizes the set of posterior beliefs that guarantees a non-negative and a non-positive B-EVSI in all decision problems. The first part of the proposition is similar to the leading example in Morris and Shin (1997), with the difference that Morris and Shin (1997) consider only one potential source of biased decoding, namely incorrect beliefs about the signal structure, whereas I also consider the possibility that the agent has an incorrect prior and is non-Bayesian.

Proposition 2. *In an environment with a binary state of the world and a binary signal:*

- $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq 0$ *in all decision problems* $\mathcal{D}(\mathcal{A}, u)$ *if and only if either the agent interprets the signal as noise or* $p|s_2(\omega_1) \leq \beta|s_2(\omega_1) \leq \beta|s_1(\omega_1) \leq p|s_1(\omega_1)$.
- $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \leq 0$ *in all decision problems* $\mathcal{D}(\mathcal{A}, u)$ *if and only if either the agent interprets the signal as noise or* $\beta(\omega_1) \in [p|s_2(\omega_1), p|s_1(\omega_1)]$ *and* $\beta|s_1(\omega_1) \leq \beta|s_2(\omega_1)$.

The intuition behind Proposition 2 is as follows: unsurprisingly, if the agent interprets the signal as noise, the value of B-EVSI is exactly zero in all decision problems. This is because the agent takes the same action at her prior beliefs and after each realization of the signal. If the agent does not interpret the signal as noise, three conditions have to be met in order for B-EVSI to be non-negative in all decision problems. First, the agent's prior cannot be too misspecified and, in particular, it has to fall within $[p|s_2(\omega_1), p|s_1(\omega_1)]$. Second, the agent must understand which signal is diagnostic of which state of the world. Third, the agent must hold posterior beliefs that are not too extreme; in particular, the posterior beliefs have to fall within $[p|s_2(\omega_1), p|s_1(\omega_1)]$ as well. If the agent does not interpret the signal as noise, only two conditions have to be met in order for B-EVSI to be non-positive in all decision problems. First, the agent's prior cannot be too misspecified and has to fall within $[p|s_2(\omega_1), p|s_1(\omega_1)]$. Second, the agent has to misunderstand which signal is diagnostic of which state of the world.

The first part of Proposition 2 highlights the existence of an asymmetry between holding overly extreme versus overly conservative posterior beliefs. Specifically, consider a biased decoder who knows which signal realization is diagnostic of which state of the world. If the biased decoder holds posterior beliefs that are less extreme than the true posterior probabilities, she is always better off being able to condition her action on the realization of the signal than not being able to. Conversely, if the biased decoder holds posterior beliefs that are more extreme than the true posterior probabilities, there always exist decision problems in which conditioning her action on the realization of the signal makes her strictly worse off.

The plausibility of supposing that the biased decoder knows which signal is diagnostic of which state of the world depends on the context. For instance, imagine the biased decoder takes a diagnostic test for a disease: it seems reasonable to imagine that, independently of her prior, she will interpret a positive result as diagnostic of having the disease and a negative result as diagnostic of not having the disease. Conversely, as discussed in Section 5, imagine a communication setting in which an informed sender lies to an uninformed receiver about whether she should take a left or a right turn to arrive at her destination. If the receiver believes the lie, she is effectively a biased decoder who misunderstands which signal is diagnostic of which state of the world.

The characterization in Proposition 2 also sheds a new light on binary signals that can be ranked according to the canonical Blackwell order. According to the Blackwell order, one signal structure is more informative than another if, from the ex-ante standpoint, an unbiased decoder is better off sampling from the former signal structure than from the latter independently of the decision problem she encounters (Blackwell, 1951,1953). For the sake of exposition, I begin the discussion by presenting a result describing some properties of the most informative binary signal structures according to the Blackwell order, namely signal structures that, to an unbiased decoder, perfectly reveal the state of the world. Next, I present a more general result that implies the first as a corollary.

Corollary 1. *Suppose* $m|\omega_1(s_1) = 1$ *and* $m|\omega_2(s_1) = 0$.

1. *If the agent understands which signal is diagnostic of which state of the world,*

   $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq 0$ *in all decision problems* $\mathcal{D}(\mathcal{A}, u)$ *and, holding the agent's beliefs fixed,*

   $$\mathcal{W}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = \max_{\{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}} \mathcal{W}\left(p, \{\tilde{p}|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$$

   in all decision problems $\mathcal{D}(\mathcal{A}, u)$.

2. *If the agent misunderstands which signal is diagnostic of which state of the world,*

   $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \leq 0$ *in all decision problems* $\mathcal{D}(\mathcal{A}, u)$ *and, holding the agent's*

beliefs fixed,

$$\mathcal{W}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = \min_{\{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}} \mathcal{W}\left(p, \{\tilde{p}|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$$

in all decision problems $\mathcal{D}(\mathcal{A}, u)$*,*

Corollary 1 shows that the most informative binary signal structures according to the Blackwell order are special for two reasons. First, for such signal structures, the conditions for non-negative and non-positive B-EVSI described in Proposition 2 become simpler and symmetric. If the biased decoder's posterior beliefs reflect a correct understanding of which signal realization is diagnostic of which state of the world, then, no matter the decision problem the biased decoder encounters, no matter her prior beliefs, and no matter her belief updating procedure, she will always be better off being able to condition her action on the signal realization than not to. Conversely, if the biased decoder's posterior beliefs reflect an incorrect understanding of which signal realization is diagnostic of which state of the world, then the biased decoder will always be made worse off by the possibility of conditioning her action on the signal realization. Second, taking the agent's biased beliefs as given, the most informative binary signal structures according to the Blackwell order are either welfare maximizing or welfare minimizing for the biased decoder. Specifically, if the biased decoder's beliefs reflect a correct understanding of which signal realization is diagnostic of which state of the world, then the most informative binary signal structures according to the Blackwell order are welfare maximizing for the agent. Otherwise, they are welfare minimizing. In light of the above, the most informative binary signal structures according to the Blackwell order can be said to be robustly welfare improving for biased decoders whose posterior beliefs reflect a correct understanding of which signal realization is diagnostic of which state of the world and robustly damaging otherwise.

As discussed in more detail in Section 5, the corollary above has implications for communication. Specifically, in the binary environment, the only piece of information that a sender needs to know in order to be able to robustly benefit or hurt the receiver is which signal the receiver thinks is diagnostic of which state of the world. In particular, the sender does not need to know anything about the decision problem that the receiver is facing, anything about the receiver's prior and posterior beliefs, and anything about the actual prior distribution of states of the world.

The result above is a corollary of a more general proposition providing a characterization and describing implications of the Blackwell order in the binary environment. Before stating the proposition, it is useful to introduce the following notation: let $U_{\{m|\omega\}_{\omega\in\Omega}}$ denote the set of biased beliefs $\beta \in \Delta(\Omega)$ and $\{\beta|s\}_{s\in\mathcal{S}} \in \Delta(\Omega)^{|\mathcal{S}|}$ that guarantee $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq 0$ in all decision problems when the signal structure is $\{m|\omega\}_{\omega\in\Omega}$. Furthermore, let $L_{\{m|\omega\}_{\omega\in\Omega}}$ denote the set of biased beliefs $\beta \in \Delta(\Omega)$ and $\{\beta|s\}_{s\in\mathcal{S}} \in \Delta(\Omega)^{|\mathcal{S}|}$ that guarantee $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \leq 0$ in all decision problems when the signal structure is $\{m|\omega\}_{\omega\in\Omega}$.

Proposition 3. *Let* $\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}}, \{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}$ *with* $m|\omega, \tilde{m}|\omega \in \Delta(\{s_1, s_2\})$ *be two signal structures satisfying* $m|\omega_1(s_1) \geq m|\omega_2(s_1)$ *and* $\tilde{m}|\omega_1(s_1) \geq \tilde{m}|\omega_2(s_1)$*. Let* $\{p|s\}_{s\in\mathcal{S}}$ *denote the true posterior distributions generated by* $\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}}$ *and* $\{\tilde{p}|s\}_{s\in\mathcal{S}}$ *the true posterior distributions generated by* $\{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}$*. Then, the following statements are equivalent:*

1. $\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}} \succeq \{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}$ *according to the Blackwell order.*
2. *If the agent understands which signal is diagnostic of which state of the world,*

   $U_{\{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}} \subseteq U_{\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}}}$
3. *If the agent misunderstands which signal is diagnostic of which state of the world,*

   $L_{\{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}} \subseteq L_{\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}}}$

Furthermore, if $\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}} \succeq \{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}$ *according to the Blackwell order and if*

- $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq 0$ *and* $\mathcal{V}\left(p, \{\tilde{p}|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq 0$ *in all decision problems* $\mathcal{D}(\mathcal{A}, u)$*, then* $\mathcal{W}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq \mathcal{W}\left(p, \{\tilde{p}|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$ *in all decision problems* $\mathcal{D}(\mathcal{A}, u)$*.*
- $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \leq 0$ *and* $\mathcal{V}\left(p, \{\tilde{p}|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \leq 0$ *in all decision problems* $\mathcal{D}(\mathcal{A}, u)$*, then* $\mathcal{W}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \leq \mathcal{W}\left(p, \{\tilde{p}|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$ *in all decision problems* $\mathcal{D}(\mathcal{A}, u)$*.*

Proposition 3 describes a set of properties of the Blackwell order when the set of signals under consideration is restricted to binary signals that hold fix which signal realization is diagnostic of which state of the world. Specifically, holding which signal is diagnostic of which state of the world fixed, Blackwell-more-informative binary signals: i) guarantee a non-negative value of information for biased decoders in all decision problems for a larger set of biased beliefs, ii) guarantee a non-positive value of information for biased decoders in all decision problems for a larger set of biased beliefs, iii) robustly increase the welfare of biased decoders who understand which signal is diagnostic of which state of the world; iii) robustly reduce the welfare of biased decoders who misunderstand which signal is diagnostic of which state of the world. The first two of these conditions are not only necessary, but also sufficient for a binary signal to be Blackwell-more-informative than another when which signal realization is diagnostic of which state of the world is held fixed.[14]

In the next section, I extend some of the results from this section to larger state and signal spaces.

[14]Restrictions on the set of signals are often necessary when developing results about biased decoders. For instance, it is hard to make an apple-to-apple comparison of signal structures with different numbers of possible signal realizations, because the addition of potential signal realizations introduces entirely new dimensions that the biased decoder could have misperceptions about. Similarly, for an unbiased decoder, switching which state of the world a signal realization is diagnostic of poses no problems, because the unbiased decoder's perceptions are assumed to also change in a way that is consistent with the signal structure. Conversely, for a biased decoder, switching which state of the world a signal realization is diagnostic of affects B-EVSI, unless one assumes that the biased decoder's beliefs about which signal is diagnostic of which state of the world also switch.

3.2 Arbitrary Finite Number of States of the World and Signals

Consider a setting with an arbitrary finite number of states of the world and possible signal realizations; i.e., let $\Omega = \{\omega_1, ..., \omega_n\}$ and $\mathcal{S} = \{s_1, ..., s_k\}$ for $n, k \in \mathbb{N}$, $n \geq 2$, $k \geq 2$. I characterize the set of misperceptions for which the value of information for a biased decoder is non-negative in all decision problems under four additional assumptions. First, I assume that the agent knows the prior distribution of states of the world; second I assume the agent is Bayesian; third, I assume that the actions space is finite; fourth, I assume that for every possible signal realization $s \in \mathcal{S}$, there exists a state of the world $\omega \in \Omega$ in which signal s is sent with positive probability $(m|\omega(s) > 0)$.[15]

Under such assumptions, the following theorem from Morris and Shin (1997) applies:

Theorem. ***(Theorem 3.1 in Morris and Shin, 1997)*** *$\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, p, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq 0$ in all decision problems $\mathcal{D}(\mathcal{A}, u)$ if and only if there exist $\psi : \mathcal{S} \to \mathbb{R}_0^+$ and $\phi : \mathcal{S}^2 \to \mathbb{R}_0^+$ such that, for all $\omega \in \Omega$ and $s_i \in \mathcal{S}$,*

$$p(\omega)\, m|\omega(s_i) = \psi(s_i)\, \beta|s_i(\omega) + \sum_{j\neq i} \left[\phi(s_i, s_j)\, \beta|s_i(\omega) - \phi(s_j, s_i)\, \beta|s_j(\omega)\right]$$

Morris and Shin (1997) provide intuition for $\psi : \mathcal{S} \to \mathbb{R}_0^+$ in the equation above, showing that it implies the requirement that the prior distribution of states of the world and the agent's subjective posterior distributions be obtained from the same joint distribution. However, they have a hard time giving function $\phi : \mathcal{S}^2 \to \mathbb{R}_0^+$ an interpretation.

In the next proposition, I present a set of conditions that are equivalent to the ones in Theorem 3.1 from Morris and Shin (1997) and whose interpretation is arguably more straightforward.

Proposition 4. *$\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, p, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq 0$ in all decision problems $\mathcal{D}(\mathcal{A}, u)$ if and only if the following three conditions are satisfied:*

(i) There exists $\lambda : \mathcal{S} \to \mathbb{R}_0^+$ such that $p(\omega) = \sum_i \lambda(s_i)\, \beta|s_i(\omega)$.

(ii) For every $i \in \{1, ..., k\}$, there exists $\alpha^{s_i} \in \Delta(\mathcal{S})$ with $\alpha_i^{s_i} > 0$ such that, for every $\omega \in \Omega$, $\beta|s_i(\omega) = \sum_{j\neq i} \alpha_j^{s_i} \beta|s_j(\omega) + \alpha_i^{s_i} p|s_i(\omega)$.

(iii) The $\{\alpha^{s_i}\}_{i\in\{1,...,k\}}$ satisfy

$$\alpha_i^{s_i} = \frac{m(s_i)}{\lambda(s_i) + \sum_{j\neq i} \alpha_i^{s_j} \frac{m(s_j)}{\alpha_j^{s_j}}}$$

Proposition 4 shows that, under the assumptions stated at the beginning of this section, the value of information for a biased decoder is non-negative in all decision problems if and only if three

[15] It is possible to dispense with the first and last assumptions and still provide a set of necessary and sufficient conditions for B-EVSI to be non-negative in all decision problems. Such characterization, however, like the one in Morris and Shin (1997), features mathematical objects that are not easy to interpret.

conditions are satisfied. The first condition requires that the prior distribution of states of the world be a convex combination of the agent's subjective posterior distributions. Loosely speaking, one can think of $\lambda : \mathcal{S} \to \mathbb{R}_0^+$ as capturing the agent's subjective beliefs about the unconditional probability of each signal realization, which in this paper I denoted by $\mu : \mathcal{S} \to \mathbb{R}_0^+$.[16] The second condition requires that for each possible signal realization $s_i \in \mathcal{S}$, the n-dimensional vector describing the agent's subjective posterior beliefs $\beta|s_i$ about the state of the world be a convex combination of the agent's other subjective posterior beliefs and of the actual posterior distribution of states of the world conditional on signal realization s_i. The third condition imposes restrictions on the weights in the convex combination. Specifically, it requires that the weight $\alpha_i^{s_i}$ that the agent places on the actual posterior distribution of states of the world conditional on signal realization s_i satisfy an equation that depends on three main elements: first, the objective unconditional probability $m(s_i)$ of observing signal realization s_i; second, the objective unconditional probability $m(s_j)$ of observing signal realization s_j for each $j \neq i$; lastly, $\lambda(s_i)$, which, as discussed, can loosely be interpreted as the agent's subjective belief about the unconditional probability of observing signal s_i.

The expression for $\alpha_i^{s_i}$ depends on $m(s_i)$ and $m(s_j)$ for $j \neq i$ in an arguably intuitive way. Specifically, as the objective unconditional probability $m(s_i)$ of observing signal realization s_i increases, the agent's subjective posterior beliefs $\beta|s_i$ are required to be closer to the objective posterior probabilities $p|s_i$. In other words, the more likely signal realization s_i is to occur, the more "accurate" the agent's posterior beliefs after observing signal s_i are required to be. Conversely, as the objective unconditional probability $m(s_j)$ of other states $s_j \neq s_i$ increases, the agent's subjective posterior beliefs $\beta|s_i$ and the objective posterior probabilities $p|s_i$ are allowed to be further away.

In the binary environment, the conditions (i) and (iii) from Proposition 4 are automatically satisfied and, if the agent does not interpret the signal as noise, the requirements from Proposition 4 are identical to the requirements from Proposition 2.

One of the key qualitative features that emerged in the binary environment is that posterior beliefs that are too extreme are dangerous for the agent in that they can lead to a strictly negative B-EVSI in some decision problem. An analogous result holds in the context of larger state and signal spaces. Specifically, it is easy to show that if there exists $s' \in \mathcal{S}$ such that $\beta|s'$ does not belong to the convex hull generated by $\{p|s\}_{s\in\mathcal{S}}$, then there always exists a decision problem in which B-EVSI is strictly negative. Therefore, holding extreme posterior beliefs is also dangerous in the context of an arbitrary finite number of states of the world and possible signal realizations. The result is quite general in the sense that it can be proved without imposing the four additional assumptions introduced at the beginning of this section.

[16] The precise relationship between λ and μ is as follows. If conditions (ii) and (iii) are satisfied when λ is replaced by μ, then $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, p, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq 0$ in all decision problems $\mathcal{D}(\mathcal{A}, u)$, because μ is always guaranteed to satisfy condition (i). Conversely, if $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, p, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq 0$ in all decision problems $\mathcal{D}(\mathcal{A}, u)$, λ is guaranteed to equal μ only if vectors $\{\beta|s_i - \beta|s_k\}_{i\in\{1,\ldots,k-1\}}$ are linearly independent.

The robust observation that holding extreme posterior beliefs might lead to a negative value of information relates to the discussion in experimental economics about over- and under-inference in belief updating (Edwards, 1968; Benjamin 2019; Augenblick, Lazarus, and Thaler, 2023; Ba, Bohren, and Imas 2023).[17] In particular, a recent set of experimental papers shows that, on average, people exhibit over-inference when signals are relatively uninformative and under-inference when signals are relatively informative (Augenblick, Lazarus, and Thaler, 2023; Ba, Bohren, and Imas 2023). The characterization in this paper suggests that the two types of mistakes in belief updating have different welfare consequences for the agent. Specifically, over-inference is robustly more pernicious then under-inference, especially in the binary environment commonly studied in the experimental literature.

I conclude this section by dispensing with the four additional assumptions imposed at the beginning of the section and introducing a set of easily-verifiable sufficient conditions for B-EVSI to be non-negative or non-positive in all decision problems.

Proposition 5.

- *If for every* $s \in \mathcal{S}$*, there exists* $\alpha_s \in [0,1]$ *such that, for every* $\omega \in \Omega$*,* $\beta|s(\omega) = \alpha_s \beta(\omega) + (1-\alpha_s)\, p|s(\omega)$*, then* $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq 0$ *in all decision problems* $\mathcal{D}(\mathcal{A}, u)$*.*
- *If for every* $s \in \mathcal{S}$*, there exists* $\alpha_s \in [0,1]$ *such that, for every* $\omega \in \Omega$*,* $\beta(\omega) = \alpha_s \beta|s(\omega) + (1-\alpha_s)\, p|s(\omega)$*,* $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \leq 0$ *in all decision problems* $\mathcal{D}(\mathcal{A}, u)$*.*

The first statement in the proposition reinforces the idea that under-inference is less nefarious than over-inference. In particular, if the agent under-estimates the precision of the signal, B-EVSI is non-negative in all decision problems. Conversely, if the agent over-estimates the precision of a signal $\beta|s$ is not in the convex hull generated by $\{p|s\}_{s\in\mathcal{S}}$ and, as discussed above, there exists a decision problem in which B-EVSI is strictly negative.

The second statement in the proposition shows that if the agent systematically updates her beliefs in a direction that is diametrically opposed to the direction in which an unbiased decoder updates her beliefs, the value of information for the agent is negative in all decision problem.

3.3 Value of Perfect Information and Full Value Extraction

So far, I considered B-EVSI: the expected value of sample information for a biased decoder. In this section, I introduce and characterize the concept of B-EVPI: the expected value of perfect information for a biased decoder. B-EVPI simply captures the difference between the expected welfare that the agent achieves if she is able to condition her actions directly on the state of the world and the welfare that she achieves if she has to rely solely on her prior information. Although

[17] Under-inference refers to the behavioral tendency to treat signals as if they were less diagnostic of the state of the world than they actually are. Over-inference refers to the opposite tendency.

quite straightforward, the definitions and results in this section are necessary to paint a complete picture of the value of information for a biased decoder and to lay the building blocks for the measures of meaningful information and meaningful uncertainty that I introduce in Section 4.

Letting δ_ω denote the degenerate probability distribution that puts probability mass equal to unity on state of the world $\omega \in \Omega$ and zero on states of the world $\omega' \in \Omega$ with $\omega' \neq \omega$, I define B-EVPI as follows:

Definition 4. The *expected value of perfect information for a potentially biased decoder* (B-EVPI) is

$$\overline{\mathcal{V}}(p, \beta) = E_{\omega \sim p}\left[u\left(a_{\delta_\omega}, \omega\right)\right] - E_{\omega \sim p}\left[u\left(a_\beta, \omega\right)\right]$$

A decomposition similar to the one in Proposition 1 shows that B-EVPI can be written as follows:

$$\overline{\mathcal{V}}(p, \beta) = \left\{E_{\omega \sim p}\left[u\left(a_{\delta_\omega}, \omega\right)\right] - E_{\omega \sim p}\left[u\left(a_p, \omega\right)\right]\right\} + d(p, \beta)$$

Therefore, B-EVPI can be written as the sum of two terms, the first of which captures the expected value of perfect information for an unbiased decoder (EVPI) and the second of which captures the utility loss that the agent would have incurred as a result of her misperceptions about the information environment had she made a decision prior to observing the realization of the signal.

The remainder of this section provides a characterization of B-EVPI. In order to do so, I need to introduce a few definitions:

Definition 5. The biased decoder has completely accurate posterior beliefs in realized state of the world $\omega \in \Omega$ if for every $s \in \mathcal{S}$ with $m|\omega(s) > 0$, $\beta|s(\omega) = 1$. The biased decoder has completely accurate posterior beliefs if she has completely accurate posterior beliefs in every realized state of the world $\omega \in \Omega$.

Definition 5 should be intuitive. The biased decoder has completely accurate posterior beliefs in realized state of the world $\omega \in \Omega$ if, for every signal realization $s \in \mathcal{S}$ that she might observe in state ω, she assigns probability equal to unity to the state of the world being ω. In other words, for every signal $s \in \mathcal{S}$ that the agent might observe in state $\omega \in \Omega$, the agent learns that the state of the world is ω. If the agent is able to do this after every possible realization of the state of the world, then the agent is said to have completely accurate posterior beliefs. If the agent has completely accurate posterior beliefs, I say that *full information is transmitted*.

Accurate posterior beliefs are a joint property of the encoding mapping and of the agent's decoding strategy. The next definition introduces the notions of full encodings and correct decodings:

Definition 6.

- Information is fully encoded if for every $\omega, \omega' \in \Omega$ with $\omega \neq \omega'$ and $s \in \mathcal{S}$, $m|\omega(s) > 0$ implies $m|\omega'(s) = 0$.

- Information is correctly decoded if for every $\omega \in \Omega$ and $s \in \mathcal{S}$ such that $m|\omega(s) > 0$, $\beta|s(\omega) = \frac{m|\omega(s)\beta(\omega)}{\sum_{\omega' \in \Omega} m|\omega'(s)\beta(\omega')}$.

Information is fully encoded if, to an unbiased decoder who knows the encoding mapping $\{m|\omega\}_{\omega\in\Omega}$, each signal realization is perfectly diagnostic of one state of the world. Information is correctly decoded if the biased decoder is Bayesian and holds posterior beliefs that are consistent with the encoding mapping. The two conditions above should be very familiar: in the context of signaling and cheap talk games, they are almost identical to the definition of a separating equilibrium. In fact, the first condition is identical to requiring that different types take different actions; the second conditions is tantamount to requiring that the agents have consistent beliefs, with the minor difference that here the agent is allowed to have incorrect prior beliefs.

The following proposition provides a characterization of B-EVPI:

Proposition 6. *The following three statements are equivalent:*

1. $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = \overline{\mathcal{V}}(p,\beta)$ *in all decision problems* $\mathcal{D}(\mathcal{A}, u)$.
2. *The biased decoder has completely accurate posterior beliefs.*
3. *Information is fully encoded and correctly decoded.*

Proposition 6 should, once again, be intuitive. In order for the agent to be able to learn the realized state of the world, the realized signal has to, in principle, be able to reveal the state of the world and the agent must interpret the signal correctly. If that is the case, the agent will be able to extract the full value of information in any decision problem she encounters.

3.4 Value Loss for a Potentially Biased Decoder

Proposition 6 shows that there are two and only two possible reasons why the agent might extract less than full value from a signal: first, it might be the case that the signal encodes less than full information; second, it might be the case that the agent does not decode the signal correctly. The notion of value loss for a potentially biased decoder introduced in this section allows one to study the relative contribution of each of the two factors.

Value loss for a potentially biased decoder is defined as the difference between the maximum value that the biased decoder could in principle extract from a signal and the value that she actually extracts from the signal.

Definition 7. The *expected loss in value for a biased decoder* (B-Loss) is defined as

$$\mathscr{L}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = \overline{\mathcal{V}}(p,\beta) - \mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$$

Leveraging the decomposition from Proposition 1, we have:

$$\mathscr{L}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = \left\{E_{\omega\sim p}\left[u\left(a_{\delta_\omega}, \omega\right)\right] - E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s}, \omega\right)\right]\right]\right\} + E_{s\sim m}\left[d\left(p|s, \beta|s\right)\right]$$

The above decomposition of B-Loss is appealing because it shows that the two possible sources of loss in value for a potentially biased decoder can be additively separated. Specifically, the first component, namely $E_{\omega\sim p}\left[u\left(a_{\delta_\omega}, \omega\right)\right] - E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s}, \omega\right)\right]\right]$, relies solely on the features of the signal structure and measures the loss in value that occurs as a result of the signal structure encoding less than full information. As such, the first component captures the amount of value that would be lost even to an unbiased decoder. The second component, which relies on both the features of the signal structure and the agent's beliefs, measures the amount of value that is lost due to the biased decoder having potentially incorrect beliefs. As such, the second component only affects biased decoders.

The decomposition above can be helpful when trying to determine empirically where the loss in value to a biased decoder comes from. Does it come primarily from the features of the signal structure? Or does it come primarily from the agent's biased beliefs? These questions are particularly poignant when the signal structure is endogenously determined in response to economic incentives. For instance, in the context of a cheap-talk model, one might be interested in determining the degree to which communication is impaired because the sender's strategy is only partially informative and the degree to which communication is impaired because the receiver does not interpret the sender's messages correctly. Similarly, in the context of a signaling game, one might be interesting in determining whether information flow is primarily impaired by the agents' behaviors or by observers' inabilities to decipher the agents' behaviors. A similar line of reasoning applies to information cascades, Bayesian persuasion, disclosure models, no-trade theorems, sequential voting models, etc.

Note that B-Loss is also useful for studying externalities and counterfactuals. For instance, does the presence of behavioral agents in a model exert a negative informational externality on non-behavioral agents? One way of answering that question is to compare B-Loss in the equilibrium of a model that does not feature behavioral agents to the first component of B-Loss in the equilibrium of a model that does feature behavioral agents. For instance, the addition of behavioral agents in a disclosure model can rationalize partial disclosure equilibria observed in experimental settings (Grossman, 1981; Milgrom, 1981; Milgrom and Roberts, 1986; Eyster and Rabin, 2005; Deversi, Ispano, and Schwardmann, 2021). How big is that informational externality? How much worse off are behavioral agents than rational agents in a cursed equilibrium? What would be the welfare gain if behavioral agents were debiased? All of these questions can be studied through the lens of B-Loss.

4 Measures of Information for a Potentially Biased Decoder

Broadly speaking, Section 3 aimed to answer the question: "How can we think about and measure the instrumental value of observing a piece of news for a potentially biased decoder?" This section aims to answer the following related question: "How can we think about and measure the amount of information that a potentially biased decoder extracts from observing a piece of news?" Happily, the discussion in Section 3 provides an almost immediate answer to the question above for the important class of measures of information characterized by Frankel and Kamenica (2019).

Frankel and Kamenica (2019) study both measures of the amount of information generated by a piece of news and measures of the uncertainty implicit in a given belief. The authors refer to a measure of information as valid if it corresponds to the ex-post value of sample information in some decision problem. Similarly, they refer to a measure of uncertainty as valid if it corresponds to the expected utility loss from not knowing the state of the world in some decision problem. Lastly, they refer to a measure of information and one of uncertainty as jointly valid if they arise in the context of the same decision problem. Their paper axiomatically characterizes all measures of information and uncertainty that are valid in the sense described above. In this section, I show that all their measures of information and uncertainty can be extended to the case of biased decoding.

Specifically, consider any decision problem $\mathcal{D}(\mathcal{A}, u)$ in which, for every $\omega \in \Omega$, $u(a_{\delta_\omega}, \omega) = 0$.[18] Suppose the agent is an unbiased decoder and define $G : \Delta(\Omega) \to \mathbb{R} \cup \{-\infty\}$ as $G(p) = E_{\omega \sim p}[u(a_p, \omega)]$. $G(p)$ thus defined satisfies two properties: i) it is a convex function; ii) $\forall \omega' \in \Omega$, $G(\delta_{\omega'}) = E_{\omega \sim \delta_{\omega'}}\left[u\left(a_{\delta_{\omega'}}, \omega\right)\right] = 0$. Then, $-G(p)$ is a valid measure of uncertainty according to Frankel and Kamenica (2019). Furthermore, when the signal realization is $s \in \mathcal{S}$,

$$d(p|s, p) = E_{\omega \sim p|s}\left[u\left(a_{p|s}, \omega\right)\right] - E_{\omega \sim p|s}\left[u\left(a_p, \omega\right)\right]$$

is a valid measure of information for Frankel and Kamenica (2019). Notice that, when developing their measures of information, Frankel and Kamenica take an ex-post rather than an ex-ante perspective. Specifically, when the signal realization is $s \in \mathcal{S}$, $d(p|s, p)$ captures the utility loss that the agent would have incurred had she based her action on her prior beliefs rather than on her posterior beliefs conditional on observing signal $s \in \mathcal{S}$. In other words, after the realization of signal $s \in \mathcal{S}$, the agent's prior beliefs are treated as a misperception and $d(p|s, p)$ captures the utility loss of taking an action based on that misperception rather than taking an action conditional on the realized signal.

Although Frankel and Kamenica (2019) primarily take an ex-post perspective when developing their measures of information, it is possible to think about these measures from the ex-ante perspective. In particular, given signal structure $\{m|s\}_{s \in \mathcal{S}}$, the ex-ante version of any of the measures

[18] Recall that $\delta_\omega \in \Delta(\Omega)$ is the degenerate probability distribution that puts probability mass equal to unity on state of the world $\omega \in \Omega$ and equal to zero on all other states $\omega' \in \Omega$ with $\omega' \neq \omega$.

of information in Frankel and Kamenica (2019) can be written as:

$$E_{s\sim m}\left[d\left(p|s,p\right)\right]=E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s},\omega\right)\right]-E_{\omega\sim p|s}\left[u\left(a_{p},\omega\right)\right]\right]=$$

$$=E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s},\omega\right)\right]\right]-E_{\omega\sim p}\left[u\left(a_{p},\omega\right)\right]=\mathcal{V}\left(p,\left\{p|s\right\}_{s\in\mathcal{S}},p,\left\{p|s\right\}_{s\in\mathcal{S}}\right)$$

Therefore, from an ex-ante perspective, the measures of information in Frankel and Kamenica (2019) are simply the expected value of sample information for an unbiased decoder in some decision problem $\mathcal{D}\left(\mathcal{A},u\right)$ that satisfies the property $\forall\omega\in\Omega$, $u\left(a_{\delta_{\omega}},\omega\right)=0$.

In the context of a decision problem satisfying the property above,

$$\overline{\mathcal{V}}\left(p,p\right)=E_{\omega\sim p}\left[u\left(a_{\delta_{\omega}},\omega\right)\right]-E_{\omega\sim p}\left[u\left(a_{p},\omega\right)\right]=-E_{\omega\sim p}\left[u\left(a_{p},\omega\right)\right]=-G\left(p\right)$$

Therefore, the measures of uncertainty in Frankel and Kamenica (2019) are simply the ex-ante value of perfect information for an unbiased decoder in the special class of decision problems described above. The deep connection between the measures of information and uncertainty in Frankel and Kamenica (2019) and the value of sample and perfect information for an unbiased decoder makes it possible to extend those measures of information and uncertainty to the case of biased decoding.

Before proceeding with such extension, however, I present an example of how to derive a canonical measure of uncertainty, namely Shannon's entropy function, and a canonical measure of information transmission, namely mutual information, from a particular decision problem faced by an unbiased decoder.

Example 2. Let $\Omega=\left\{\omega_1,...,\omega_n\right\}$ and consider decision problem $\mathcal{D}\left(\mathcal{A},u\right)$ with $\mathcal{A}=\Delta\left(\Omega\right)$. Denote an action $a\in\mathcal{A}$ by vector $\left(a_1,...,a_n\right)$ and, for each $i\in\left\{1,...,n\right\}$, let $u\left(a,\omega_i\right)=\ln\left(a_i\right)$. Then, $E_{\omega\sim p}\left[u\left(a,\omega\right)\right]=\sum_{i=1}^{n}p\left(\omega_i\right)\ln\left(a_i\right)$. It is easy to show that $\arg\max_{a\in\mathcal{A}}E_{\omega\sim p}\left[u\left(a,\omega\right)\right]=\left\{\left(p\left(\omega_1\right),...,p\left(\omega_n\right)\right)\right\}$. Therefore,

$$G\left(p\right)=\max_{a\in\mathcal{A}}E_{\omega\sim p}\left[u\left(a,\omega\right)\right]=E_{\omega\sim p}\left[u\left(a_p,\omega\right)\right]=\sum_{\omega\in\Omega}\left[p\left(\omega\right)\ln\left(p\left(\omega\right)\right)\right]$$

Consider $\overline{\mathcal{V}}\left(p,p\right)=-G\left(p\right)$. $\overline{\mathcal{V}}\left(p,p\right)$ is Shannon's entropy function. Notice that $\overline{\mathcal{V}}\left(p,p\right)$ is a valid measure of uncertainty according to Frankel and Kamenica (2019), because it is concave and satisfies $\overline{\mathcal{V}}\left(\delta_{\omega'},\delta_{\omega'}\right)=-E_{\omega\sim\delta_{\omega'}}\left[u\left(a_{\delta_{\omega'}},\omega\right)\right]=0$, $\forall\omega'\in\Omega$. The valid ex-post measure of information from Frankel and Kamenica (2019) that is associated to uncertainty function $\overline{\mathcal{V}}\left(p,p\right)$ is $D_{KL}\left(p|s,p\right)$, where $D_{KL}\left(p|s,p\right)=\sum_{\omega\in\Omega}p|s\left(\omega\right)\ln\left(\frac{p|s(\omega)}{p(\omega)}\right)$ is simply the Kullback-Leibler divergence from p to $p|s$ (Kullback and Leibler, 1951). The ex-ante measure of information that is associated to uncertainty function $\overline{\mathcal{V}}\left(p,p\right)$ is

$$\mathcal{V}\left(p,\left\{p|s\right\}_{s\in\mathcal{S}},p,\left\{p|s\right\}_{s\in\mathcal{S}}\right)=E_{s\sim m}\left[D_{KL}\left(p|s,p\right)\right]=$$

$$= \sum_{s\in\mathcal{S}} m(s) \sum_{\omega\in\Omega} p|s(\omega) \ln\left(\frac{p|s(\omega)}{p(\omega)}\right) = \sum_{\omega\in\Omega}\sum_{s\in\mathcal{S}} p|s(\omega)\, m(s) \ln\left(\frac{p|s(\omega)\, m(s)}{p(\omega)\, m(s)}\right) = I(s,\omega)$$

where $I(s,\omega)$ denotes the mutual information between random variables s and ω.

The example above should be familiar: Shannon's entropy function is generally presented as a measure of uncertainty and mutual information is the canonical measure of information transmission associated to Shannon's entropy.

The following definition formalizes the notion of extending the measures of information and uncertainty for an unbiased decoder developed in Frankel and Kamenica (2019) to the case of biased decoding.

Definition 8. Consider a decision problem $\mathcal{D}^*(\mathcal{A},u)$ that, in the language of Frankel and Kamenica (2019), gives rise to jointly valid measures of uncertainty $\overline{\mathcal{V}}(p,p)$ and information $d(p|s,p)$. I say that:

- $\overline{\mathcal{V}}(p,\beta)$ extends valid measure of uncertainty $\overline{\mathcal{V}}(p,p)$ to the case of biased decoding if $\overline{\mathcal{V}}(p,\beta) = \overline{\mathcal{V}}(p,p)$ when the agent is an unbiased decoder and $\overline{\mathcal{V}}(p,\beta)$ is derived from decision problem $\mathcal{D}^*(\mathcal{A},u)$ as shown in Section 3 when the agent is a biased decoder with prior beliefs $\beta\in\Delta(\Omega)$ and posterior beliefs $\{\beta|s\}_{s\in\mathcal{S}}\in\Delta(\Omega)^{|\mathcal{S}|}$.

- $d(p|s,\beta) - d(p|s,\beta|s)$ extends valid ex-post measure of information $d(p|s,p)$ to the case of biased decoding if $d(p|s,\beta) - d(p|s,\beta|s) = d(p|s,p)$ when the agent is un unbiased decoder and $d(p|s,\beta) - d(p|s,\beta|s)$ is derived from decision problem $\mathcal{D}^*(\mathcal{A},u)$ as shown in Section 3 when the agent is a biased decoder with prior beliefs $\beta\in\Delta(\Omega)$ and posterior beliefs $\{\beta|s\}_{s\in\mathcal{S}}\in\Delta(\Omega)^{|\mathcal{S}|}$.

The following proposition highlights the connection between the measures of information and uncertainty from Frankel and Kamenica (2019) and the framework developed in this paper.

Proposition 7. *If $\overline{\mathcal{V}}(p,p)$ and $d(p|s,p)$ are jointly valid measures of uncertainty and information in Frankel and Kamenica (2019), then $\overline{\mathcal{V}}(p,\beta)$ and $d(p|s,\beta) - d(p|s,\beta|s)$ extend $\overline{\mathcal{V}}(p,p)$ and $d(p|s,p)$ to the case of biased decoding. Similarly, for any decision problem $\mathcal{D}^*(\mathcal{A},u)$ satisfying the property $u(a_{\delta_\omega},\omega) = 0$ for every $\omega\in\Omega$, $\overline{\mathcal{V}}(p,\beta)$ and $d(p|s,\beta) - d(p|s,\beta|s)$ are jointly valid measures of uncertainty and information in Frankel and Kamenica (2019) whenever the agent is an unbiased decoder.*

Therefore, Proposition 7 shows: i) that all the measures of information and uncertainty from Frankel and Kamenica (2019) can be extended to the case of biased decoding; ii) that all the measures of information and uncertainty for biased decoders developed in the context of decision problems $\mathcal{D}(\mathcal{A},u)$ satisfying the property $u(a_{\delta_\omega},\omega) = 0\ \forall\omega\in\Omega$ reduce to the measures of information in Frankel and Kamenica (2019) when the agent is an unbiased decoder.

In the context of measures of information and uncertainty, the mathematical objects introduced in Section 3 become:

$$\overline{\mathcal{V}}(p,\beta) = \overline{\mathcal{V}}(p,p) + d(p,\beta) = -E_{\omega\sim p}\left[u\left(a_p,\omega\right)\right] + d(p,\beta)$$

$$\begin{aligned}\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right) &= \overline{\mathcal{V}}(p,\beta) - E_{s\sim m}\left[\overline{\mathcal{V}}\left(p|s,\beta|s\right)\right] = \\ &= E_{s\sim m}\left[\overline{\mathcal{V}}(p,p) - \overline{\mathcal{V}}\left(p|s,p|s\right)\right] + d(p,\beta) - E_{s\sim m}\left[d\left(p|s,\beta|s\right)\right] = \\ &= \left\{E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s},\omega\right)\right]\right] - E_{\omega\sim p}\left[u\left(a_p,\omega\right)\right]\right\} + d(p,\beta) - E_{s\sim m}\left[d\left(p|s,\beta|s\right)\right]\end{aligned}$$

$$\mathscr{L}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right) = E_{s\sim m}\left[\overline{\mathcal{V}}\left(p|s,\beta|s\right)\right] = -E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s},\omega\right)\right]\right] + E_{s\sim m}\left[d\left(p|s,\beta|s\right)\right]$$

The interpretation is as follows. Rather than simply being a measure of objective uncertainty, $\overline{\mathcal{V}}(p,\beta) = -E_{\omega\sim p}\left[u\left(a_p,\omega\right)\right] + d(p,\beta)$ captures the relationship between subjective and objective uncertainty. In particular, one can think of measures of uncertainty for unbiased decoders as capturing the expected amount of "learning" that the agent has to do in each realized state of the world in order to have ex-post accurate beliefs.[19] The same interpretation holds in the case of biased decoders: $\overline{\mathcal{V}}(p,\beta)$ captures the expected amount of "learning" that a biased decoder has to do in order to hold completely accurate posterior beliefs in each realized state of the world. Because of her biased beliefs, the biased decoder has to do more learning in expectation than an unbiased decoder in order to hold completely accurate posterior beliefs; i.e. $\overline{\mathcal{V}}(p,\beta) \geq \overline{\mathcal{V}}(p,p)$.

$\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)$ captures the amount of information that is transmitted to a biased decoder. I refer to $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)$ as a measure of *meaningful information transmission*, because not only does it take into account the features of the signal-generating process; it also takes into account the interpretation/meaning that an agent assigns to observed signals. $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)$ is positive if, in expectation after observing the signal, the expected amount of "learning" that a biased decoder has to do in order to have completely accurate posterior beliefs in each realized state of the world decreases; conversely, $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)$ is negative if the expected amount of "learning" that a biased decoder has to do in order to have completely accurate posterior beliefs increases. The reason why $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)$ can sometimes be negative is that the agent's biased beliefs and incorrect belief-updating procedures might sometimes lead her to develop even less accurate beliefs as a result of sampling from the signal structure. In that case, the amount of "learning" that the biased decoder has to do in order to have completely accurate posterior beliefs increases.

Lastly, $\mathscr{L}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)$ isolates the expected amount of "learning" that a biased

[19] Different measures of uncertainty quantify the amount of learning in different ways.

decoder has to do after observing the signal in order to have completely accurate posterior beliefs in each realized state of the world. If $\mathscr{L}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = 0$, I say that full meaningful information is transmitted. One way of thinking about $\mathscr{L}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$ is that it captures impairments to full meaningful information transmission. Those impairments come from two sources. The first source relates solely to the features of the signal structure: if the signal structure is less than fully informative, then some lingering uncertainty about the state of the world would remain even if the agent was an unbiased decoder. Therefore, the first source of impairment captures the amount of meaningful information that is lost as a result of the signal structure being less than fully informative. The second source relates both to the signal structure and to the biased decoder's beliefs and captures the amount of meaningful information that is lost as a result of the agent's biased beliefs. Adopting the interpretation that $\mathscr{L}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$ captures impairments to full meaningful information transmission, I refer to it as a measure of *meaningful information loss*. The additively separable nature of the expression above shows that the two sources of impairment to full meaningful information transmission can be neatly separated.

The following remark highlights a few properties of $\overline{\mathcal{V}}(p, \beta)$, $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$, and $\mathscr{L}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$.

Remark 2. In the context of decision problems under uncertainty satisfying $u\left(a_{\delta_\omega}, \omega\right) = 0$ for every $\omega \in \Omega$, $\overline{\mathcal{V}}(p, \beta)$, $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$, and $\mathscr{L}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$ have the following properties:

- As far as $\overline{\mathcal{V}}(p, \beta)$ is concerned, we have
  - $\overline{\mathcal{V}}(p, \beta) \geq \overline{\mathcal{V}}(p, p) \geq 0$.
  - $\overline{\mathcal{V}}(p, \beta) = \overline{\mathcal{V}}(p, p)$ if the biased decoder has a correct prior.
- As far as $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$ is concerned, we have
  - $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = E_{s\sim m}\left[\overline{\mathcal{V}}(p, p) - \overline{\mathcal{V}}(p|s, p|s)\right] \geq 0$ if the agent has a correct prior and information is correctly decoded.
  - $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = -E_{s\sim m}\left[d\left(p|s, \beta|s\right)\right] \leq 0$ if the agent has a correct prior and information is fully encoded.
- As far as $\mathscr{L}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$ is concerned, we have
  - $\mathscr{L}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq 0$.
  - $\mathscr{L}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = E_{s\sim m}\left[d\left(p|s, \beta|s\right)\right] \geq 0$ if and only if information is fully encoded.
  - $\mathscr{L}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = -E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s}, \omega\right)\right]\right] \geq 0$ if information is correctly decoded and either the agent has a correct prior or information is fully encoded.

– $\mathscr{L}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = 0$ if information is fully encoded and correctly decoded.

The properties above follows straightforwardly from the definitions, from the fact that Bregman divergences are non-negative, and from the fact that $G(p) = E_{\omega\sim p}[u(a_p, \omega)] \leq 0$ in the context of the decision problems considered in this section. Furthermore, in the context of measures of meaningful information arising from decision problems satisfying $\underset{a\in\mathcal{A}=\Delta(\Omega)}{\arg\max} E_{\omega\sim p}[u(a, \omega)] = \{p\}$ (e.g., the problem for which Shannon's entropy function is a valid measure of uncertainty), all the statements in Remark 2 are if and only if.

Reprising the example involving Shannon's entropy function as a measure of uncertainty and denoting such function by $\mathcal{H} : \Delta(\Omega) \to \mathbb{R}$, we have:

$$\overline{\mathcal{V}}(p, \beta) = \mathcal{H}(p) + D_{KL}(p, \beta)$$

$$d(p|s, \beta) - d(p|s, \beta|s) = D_{KL}(p|s, \beta) - D_{KL}(p|s, \beta|s)$$

$$\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = I(s, \omega) + D_{KL}(p, \beta) - E_{s\sim m}[D_{KL}(p|s, \beta|s)]$$

$$\mathscr{L}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = \{\mathcal{H}(p) - I(s, \omega)\} + E_{s\sim m}[D_{KL}(p|s, \beta|s)]$$

Other examples of measures of information and uncertainty that can be extended to the case of biased decoding involve the Brier score, the Tsallis score, quadratic loss estimation, and more. See Dawid and Musio (2014) and Frankel and Kamenica (2019) for details.

Online Appendix C presents two examples of how to estimate meaningful information measures on empirical data. Specifically, I study the transmission of meaningful information in the context of the experiments by Ambuehl and Li (2018) and by Braghieri (2021).

5 Foundations of Communication

In this part of the paper, I show how the machinery introduced in the previous sections can help shed light on foundational concepts in communication such as successful communication, vagueness, misunderstanding, deception, and lies.

Compared to the assumptions I imposed on the agent in most of the previous sections, the assumptions in this section are more stringent. Specifically, I assume that the receiver has correct beliefs about the prior distribution of states of the world and updates her beliefs according to Bayes' rule. The rationale for imposing more stringent assumptions is that this part of the paper differs from the previous ones in terms of aims. In particular, the goal of the previous sections was to develop a flexible language to describe the behavior of an agent who interacts with the information environment around her with biased beliefs and to develop measures the amount of information transmitted to such an agent. As such, a weaker set of assumptions broadened the range of settings

that such language could be applied to.[20] Conversely, the goal of this section is to define some of the key concepts in human communication, relate them to one another, and derive basic implications. As such, a tighter set of assumptions improves the exposition by shining a light only on the most relevant features.

Although the focus of this paper is on human communication, the basic building blocks of the theory apply to communication between entities ranging from plants, to non-human animals, to mechanical objects such as computers.[21] Such building blocks are: i) a sender, ii) a receiver, iii) a stochastic state of the world that is observable to the sender but not to the receiver[22], iv) a set of signals that the sender can send with the purpose of affecting the receiver's beliefs or behaviors, v) an encoding mapping that describes the distribution of signals conditional on each state of the world, vi) a decision problem for the receiver that also affects the payoff of the sender, and vii) a behavioral response by the receiver to the realized signal.[23]

Most instances of human communication rely on language: a conventional system of symbols with a complex structure and commonly understood meanings. This paper is concerned with more basic forms of human communication that either do not rely on the existence of a conventional language or that rely only on a highly stylized version of a conventional language, as described in Section 5.1.1.

5.1 Setup

The setup and notation are similar to the ones in Section 2, with the difference that now the signal structure is determined by a sender. Specifically, I assume that the sender observes the state of the world and then communicates with the receiver by sending her a one-time message $s \in \mathcal{S}$. The receiver plays the role of the potentially biased decoder from Section 2 in the sense that she does not directly observe the state of the world and she updates her beliefs, possibly erroneously, upon observing the sender's message $s \in \mathcal{S}$. In this section, I refer to $s \in \mathcal{S}$ as message or signal realization interchangeably.

For reasons discussed in the previous section, I assume a common prior and common knowledge of Bayesian rationality. Recalling that $\{\mu|\omega\}_{\omega\in\Omega} \in \Delta(\mathcal{S})^{|\Omega|}$ denotes the receiver's beliefs about the encoding mapping, the receiver's posterior beliefs are $\beta|s(\omega) = \frac{\mu|\omega(s)p(\omega)}{\sum_{\omega\in\Omega}\mu|\omega(s)p(\omega)}$ whenever $\{\mu|\omega\}_{\omega\in\Omega}$ assigns non-zero probability to message $s \in \mathcal{S}$. Whenever the receiver observes a signal realization

[20] For instance, in Section C I apply the notion of meaningful information transmission to an experimental setting in which subjects' belief updating deviates from the Bayesian benchmark.

[21] Most of the key building blocks were identified by Claude Shannon in his 1948 paper "A Mathematical Theory of Communication" (Shannon, 1948). The same elements can be found in Crawford and Sobel's 1982 paper "Strategic Information Transmission" (Crawford and Sobel, 1982).

[22] All internal states of the sender – e.g., whether the sender is hungry, whether she is thinking about poetry, whether her foot hurts, etc. – are states of the world that are observable to the sender but not to the receiver.

[23] As shown in Section 4, belief updating itself can be treated as a behavioral response in a particular class of decision problems.

$s \in \mathcal{S}$ that $\{\mu|\omega\}_{\omega\in\Omega}$ assigns zero probability to, $\beta|s(\omega)$ is assumed to be an arbitrary element of $\Delta(\Omega)$. The receiver is not assumed to know the actual encoding mapping used by the sender; therefore, $\{\mu|\omega\}_{\omega\in\Omega}$ need not equal $\{m|\omega\}_{\omega\in\Omega}$; as a consequence, $\{\beta|s\}_{s\in\mathcal{S}}$ need not equal $\{p|s\}_{s\in\mathcal{S}}$.

Since some concepts such as deception rely on a notion of intentionality on the part of the sender, I introduce notation for the sender's higher-order beliefs. Specifically, I consider the sender's second-order beliefs about the receiver's first-order beliefs about the sender's encoding mapping. Such second order beliefs are an element of $\Delta\left(\Delta(\mathcal{S})^{|\Omega|}\right)$; however, for the purpose of this paper, it is sufficient to consider degenerate second order beliefs, which I denote by $\{\hat{\mu}|\omega\}_{\omega\in\Omega} \in \Delta(\mathcal{S})^{|\Omega|}$. I also consider the sender's second order beliefs about the receiver's first order beliefs about the state of the world conditional on the signal realization. Such second order beliefs are an element of $\Delta\left(\Delta(\Omega)^{|\mathcal{S}|}\right)$, but, once again, it is sufficient for the purpose of this paper to consider degenerate second order beliefs $\left\{\hat{\beta}|s\right\}_{s\in\mathcal{S}} \in \Delta(\Omega)^{|\mathcal{S}|}$. Since I assumed common knowledge of Bayesian rationality, $\hat{\beta}|s$ is obtained from $\{\hat{\mu}|\omega\}_{\omega\in\Omega}$ using Bayes' rule whenever possible.

5.1.1 Linguistic Conventions, Canonical Message Spaces, and Conventional Languages

Some of the concepts that I analyze in this part of the paper require the sender to have access to messages that have a commonly understood conventional meaning. In order to describe the conventional meaning of messages, I introduce the notions of linguistic convention, of canonical message space, and of conventional language. The setup is highly stylized and does not aim to capture the essential features of human language; it only aims to capture the minimal set of features that are necessary to introduce concepts such as lying.

A linguistic convention is an encoding mapping in which messages have a commonly understood meaning and refer to subsets of the state space. For instance, a linguistic convention might assign signal $s_1 \in \mathcal{S}$ to the singleton containing state of the world ω_1; in that case, the interpretation of signal s_1 according to the linguistic convention would be "State ω_1 occurred." Similarly, a linguistic convention might assign signal s_2 to subset $\{\omega_1, \omega_2\} \subseteq \Omega$; in that case, the interpretation of signal s_2 according to the linguistic convention would be "Either state ω_1 or state ω_2 occurred." Formally:

Definition 9. Consider $\mathcal{T} \subseteq 2^{\Omega}\backslash\{\emptyset\}$, where 2^{Ω} is the power set of Ω. Let $|\mathcal{S}| \geq |\mathcal{T}|$, let each element $T \in \mathcal{T}$ be associated to a signal $s_T \in \mathcal{S}$. A linguistic convention ℓ is a set of conditional distributions $\{\ell|\omega\}_{\omega\in\Omega}$, with $\ell|\omega \in \Delta(\mathcal{S})$ satisfying the following property: for every $T \in \mathcal{T}$ and $\omega, \omega' \in \Omega$, if $\omega \notin T$, $\ell|\omega(s_T) = 0$ and if ω and $\omega' \in T$, $\ell|\omega(s_T) = \ell|\omega'(s_T) > 0$.

The definition above requires that, for any set $T \subseteq \mathcal{T}$, there exists a message $s_T \in \mathcal{S}$ that is sent with positive probability if and only if some state $\omega \in T$ occurs. When set T is a singleton, a receiver that interprets messages according to linguistic convention ℓ assigns $\beta|s_T(\omega) = 1$ to state of the world $\omega \in T$ and $\beta|s_T(\omega') = 0$ to all other states of the world $\omega' \notin T$. When set T is not a singleton, a receiver that interprets messages according to linguistic convention ℓ assigns

$\beta|s_T(\omega) = \frac{\beta(\omega)}{\sum_{\omega'' \in T} \beta(\omega'')}$ to each state of the world $\omega \in T$ and $\beta|s_T(\omega') = 0$ to all other states of the world $\omega' \notin T$. As in Sobel (2020), the definition of linguistic convention allows for the existence of signals that do not have commonly understood meaning.

Next, I introduce the notion of canonical message space. A canonical message space is simply a message space that allows the sender to refer to each non-empty subset of Ω.

Definition 10. A canonical message space $\mathcal{S}^*$ is a message space of cardinality $|\mathcal{S}^*| = 2^\Omega - 1$.

In one-way communication, a canonical message space $\mathcal{S}^*$ and a linguistic convention ℓ defined on it form a conventional language. A conventional language allows the sender to refer to each non-empty subset T of Ω by means of a message s_T whose commonly understood meaning is "the realized state of the world $\omega \in \Omega$ is an element of $T \subseteq 2^\Omega \setminus \{\emptyset\}$."

Definition 11. A conventional language is a duple $(\mathcal{S}^*, \ell)$, where $\mathcal{S}^*$ is the canonical message space and ℓ is a linguistic convention.

Note that, according to my definition, a conventional language is a relatively simple object that allows agents to refer to subsets of the state space, but that does not allow them to make probabilistic statements such as "State ω_1 occurred with probability $\frac{3}{4}$." One way of thinking about a conventional language, as I define it, is as a language that is able to capture the idea that certain words – e.g., the words "animal", "mammal", and "dog" – are nested.

The following example might help build intuition:

Example 3. Let $\omega \in \{\omega_1, \omega_2, \omega_3\}$. A conventional language $(\mathcal{S}^*, \ell)$ is one in which $\mathcal{S}^* = \{s_1, ..., s_7\}$ and $\{\ell|\omega\}_{\omega \in \Omega}$ is any encoding mapping of the form:

$$\begin{pmatrix} x_1 & 0 & 0 & x_4 & x_5 & 0 & x_7 \\ 0 & x_2 & 0 & x_4 & 0 & x_6 & x_7 \\ 0 & 0 & x_3 & 0 & x_5 & x_6 & x_7 \end{pmatrix}$$

where the element of the matrix in the i^{th} row and j^{th} column denotes $\ell|\omega_i(s_j)$, where $x_i > 0$ $\forall i \in \{1, ..., 7\}$, and where the x_i are such that the matrix above is row stochastic.

The following definition relates the sender's encoding and the receiver's decoding mappings to the linguistic convention.

Definition 12. Given conventional language $(\mathcal{S}^*, \ell)$:

- The sender's encoding mapping follows the linguistic convention if for every $\omega \in \Omega$ and $s \in \mathcal{S}$ such that $m|\omega(s) > 0$, $p|s(\omega) = \frac{\ell|\omega(s)p(\omega)}{\sum_{\omega' \in \Omega} \ell|\omega'(s)p(\omega')}$.
- The receiver interprets messages according to the linguistic convention if $\mu|\omega = \ell|\omega$.[24]

[24] Sobel (2020) refers to receivers who interpret messages according to the linguistic convention as credulous.

- The sender believes that the receiver interprets messages according to the linguistic convention if $\hat{\mu}|\omega = \ell|\omega$.

Notice that, in order to follow the linguistic convention, the sender's encoding mapping need not equal the linguistic convention; in other words, it need not be the case that for every $\omega \in \Omega$, $m|\omega = \ell|\omega$. For the sender's encoding mapping to follow the linguistic convention, it is sufficient that the posteriors generated by the sender's actual encoding mapping $\{m|\omega\}_{\omega\in\Omega}$ and the posteriors generated when interpreting messages according to the linguistic convention $\{\ell|\omega\}_{\omega\in\Omega}$ be equal for all messages that are sent with positive probability. For instance, in the context of the conventional language $(\mathcal{S}^*, \ell)$ from Example 3, an encoding mapping that, in states of the world ω_1 and ω_2, sends message s_4 with probability equal to unity and, in state of the world ω_3, sends message s_3 with probability equal to unity is consistent with the linguistic convention.

5.2 Persuasion vs. Cheap Talk, Local vs. Global Properties of Communication, and Ex-ante vs. Ex-post Welfare

Before embarking in the analysis of some of the foundational concepts in communication, it is useful to distinguish between two sets of assumptions that in the economics literature are generally referred to as *persuasion* and *cheap talk*, between *global* and *local* properties of communication, and between the *ex-ante* and the *ex-post* perspectives on the receiver's welfare.

Persuasion and cheap talk make different assumptions about the sender's ability to commit ex-ante to a signal structure. Specifically, *(Bayesian) persuasion* refers to environments in which a sender is able to commit to a signal structure before observing the realized state of the world and is then forced to send the receiver a message according to the signal structure she committed to. Conversely, *cheap talk* refers to environments in which the sender first observes the state of the world and then chooses a distribution of messages in that state of the world, without being able to control the messages that are sent in the other states of the world. For the purpose of this paper, the key difference between persuasion and cheap talk is that in the persuasion environment the sender has control over the entire encoding mapping, whereas in the cheap talk environment a sender who observes that the realized state of the world is $\omega \in \Omega$ only has control over the messages that are sent in state of the world $\omega \in \Omega$.

In standard persuasion and cheap talk models, the receiver is generally assumed to have correct beliefs about the encoding mapping. In the persuasion case, such assumption is motivated by the fact that the sender publicly commits to a signal structure; in the cheap talk case, such assumption is motivated by the fact that the environment is generally studied in equilibrium.

In order to analyze concepts such as misunderstandings, lies, and deception, I allow the receiver to have incorrect beliefs about the encoding mapping in both the persuasion and the cheap talk setting. In the context of persuasion, such incorrect beliefs might arise, for instance, when the

sender makes a public investment in information gathering and the receiver has incorrect beliefs about the signal structure generated by such investment, or when the sender can secretly violate the commitment assumption. In the context of cheap talk, such incorrect beliefs naturally arise when the environment is out of equilibrium.

The difference between persuasion and cheap talk is related to the difference between *global* and *local* properties of communication. Global properties of communication depend on the entire encoding mapping. For instance, the amount of meaningful information transmitted in one-way communication is a global property of communication, because, as shown in Section 4, it depends on the features of the entire encoding mapping. Conversely, local properties of communication, despite operating against a background of global properties, can be defined more narrowly as functions of the sender's behavior in a particular state of the world. For instance, whenever a sender observes that the state of the world is $\omega \in \Omega$ and sends a message that she expects the receiver to interpret as indicating that the state of the world is $\omega' \in \Omega$ with $\omega' \neq \omega$, it is not necessary to have knowledge of the entire encoding mapping to know that the sender is attempting to deceive the receiver.

The relationship between persuasion, cheap talk, local and global properties of communication is as follows. In the persuasion case, the sender chooses the entire encoding mapping; therefore, her choice determines both the local and the global properties of communication. In the cheap talk case, a sender of type $\omega \in \Omega$ only has control over the messages that are sent in state of the world $\omega \in \Omega$; therefore, her choice only determines the local properties of communication. The global properties of communication in the cheap talk setting are emergent: they are joint properties of the individual and generally uncoordinated behaviors of senders in different states of the world.

Although both persuasion and cheap talk models give rise to global and local properties of communication, it is more natural for the purpose of this paper to study global properties from the perspective of persuasion and local properties from the perspective of cheap talk. The reasoning is as follows: analyzing global properties from the perspective of cheap talk would require studying the emergent features of the individual and uncoordinated behavior of senders in different states of the world. In cheap talk models, such emergent features arise in the context of assumptions about the economic fundamentals – e.g., the decision problems faced by senders of different types – and about the state of the system – e.g. notions of equilibrium. Since the specifics of the sender's decision problem and of various assumptions about the state of the system are tangential to the core aim of this paper, I streamline the exposition by studying the global properties of communication from the perspective of persuasion. Of course, one can consider the global properties from the perspective of cheap talk as well, with only minor changes in formalism and interpretation. The reasons for studying local properties of communication from the perspective of cheap talk are twofold: first, since local properties depend only on the individual behavior of senders in particular states of the world, issues related to emergence can often be avoided. Second, some local properties – e.g., whether a message is a lie – are defined with respect to a realized state of the world; as such, they

are more naturally analyzed from the perspective of cheap talk. Of course, one could study local properties from the perspective of persuasion as well, with only minor changes in formalism and interpretation.

The last relevant distinction is the one between the receiver's expected welfare before the state of the world is realized (*ex-ante welfare*) and the receiver's welfare after the state of the world is realized (*ex-post welfare*). For global properties of communication such as meaningful information transmission, it is important to study the receiver's welfare form the ex-ante perspective as in Sections 3 and 4. For local properties of communication such as lying, I still primarily adopt an ex-ante perspective on welfare, but I sometimes complement it with the ex-post perspective. The ex-ante value of information, denoted by $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$ and referred to as B-EVSI, was described extensively in Sections 2 through 4. Under the assumption of a common prior, the ex-post value of information in state of the world $\omega \in \Omega$ can be defined as a function $\mathcal{Z} : \Delta(\mathcal{S})^{|\Omega|} \times \Delta(\Omega)^{|\mathcal{S}|} \times \Omega \to \mathbb{R} \cup \{-\infty\}$ where

$$\mathcal{Z}\left(\{m|\omega\}_{\omega\in\Omega}, \{\beta|s\}_{s\in\mathcal{S}}, \omega\right) = E_{s\sim m|\omega}\left[u\left(a_{\beta|s}, \omega\right)\right] - u\left(a_p, \omega\right)$$

In what follows, I use the terms value of information and value of communication interchangeably and, if I don't specify whether I am referring to the ex-ante or the ex-post value, it should be assumed that I am referring to the ex-ante value.

5.3 Roadmap

I organize the discussion of the foundational concepts of communication around two axes: the sender's intentions and the receiver's interpretations. Loosely speaking, a sender has honest intentions if she sends messages with the expectation that the receiver will interpret them correctly. If, conversely, she sends messages with the expectation that the receiver will interpret them incorrectly, she has dishonest intentions. The receiver's interpretations are correct whenever he decoding strategy is correct according to Definition 6; otherwise, the receiver's interpretations are incorrect.

Concepts such as successful communication, misunderstandings, deception, and thwarted deception arise from the interaction of the sender's intentions and the receiver's interpretations, as shown in the table below.

		Receiver	
		Correct interpretations	*Incorrect interpretations*
Sender	*Honest intentions*	Successful communication	Misunderstandings
	Dishonest intentions	Thwarted deception	Deception

Table 1: Sender's intentions and receiver's interpretations.

I will discuss each cell of the table in turn starting with successful communication and proceeding clockwise. The discussion of thwarted deception is relegated to Online Appendix D.

5.4 Successful Communication

According to the Merriam-Webster dictionary, communication is "a process by which information is exchanged between individuals through a common system of symbols, signs, or behavior (Merriam-Webster, 2023)." The definition of successful communication below aims to formalize the Merriam-Webster definition from the perspective of persuasion, making explicit the fact that the exchange of information is deliberate rather than accidental.

Definition 13. Communication is successful if: i) for every $\omega \in \Omega$ and $s \in \mathcal{S}$ with $m|\omega(s) > 0$, $\hat{\beta}|s = p|s = \beta|s$, and ii) the receiver does not interpret the signal as noise.

In other words, communication is successful if: a) the sender chooses a signal structure that she correctly believes the receiver will decode accurately, and b) at least one of the signal realizations induces the receiver to update her beliefs.[25] Of course, under the assumption of a common prior, when communication is successful the receiver is an unbiased decoder. Therefore, the following properties follow immediately from the results in previous sections:

Corollary 2. *If communication is successful:*

[25]The definition of successful communication has a shortcoming that can be addressed at the cost of introducing additional notation. Specifically, according to the definition above, a situation in which the receiver thinks that the sender's intention was to deceive her and that she successfully thwarted the sender's attempt at deception might be considered an instance of successful communication. Such shortcoming can be addressed by imposing further agreement conditions on higher-order beliefs.

- *The value of communication to the receiver is non-negative in all decision problems $\mathcal{D}(\mathcal{A}, u)$. Furthermore, there exists a decision problem $\mathcal{D}^*(\mathcal{A}, u)$ in which the value of communication is strictly positive.*

- *For all measures of meaningful information introduced Section 4, a non-negative amount of meaningful information is transmitted to the receiver. Furthermore, there exists one such measure for which the amount of meaningful information transmitted is strictly positive.*

Therefore, in line with the Merriam-Webster definition, successful communication can be thought of as the deliberate and successful exchange of meaningful information between two individuals by means of a commonly understood system of signs.

5.4.1 Vagueness

Somewhat ironically, the term vagueness has been used in economics to refer to two different ideas. The first is the idea that certain words or statements fail to refer to specific and well-defined categories (Lipman, 2009). For instance, there is no objective line of demarcation between individuals who are "tall" and individuals who are not. The second is the idea that certain words or statements provide less than full information (Deversi, Ispano, and Schwardmann, 2021). For instance, referring to an academic institution as a top-10 university conveys less information than referring to it as the 7^{th} highest-ranked university. In this paper, I use the word vague to refer to the latter idea.

Successful communication need not be precise, in the sense that it need not fully reveal the state of the world to the receiver. Even in the presence of a conventional language and of linguistic statements that the receiver understands perfectly well, the sender's messages can still be vague. For instance, a person might say that she was at the hospital without specifying the reason for being there.

The degree to which messages exchanged during successful communication are vague is determined by the properties of the encoding mapping.[26] Specifically, a message is defined as vague if it leaves some lingering uncertainty about the state of the world even when communication is successful; otherwise, the message is precise. Encoding mappings can differ in the degree to which they are vague. In the definition below, I equate the degree of vagueness of an encoding mapping with the Blackwell order. Such choice seems natural is in virtue of one of the equivalent ways in which the Blackwell order can be characterized, namely by the property that an encoding mapping is Blackwell more informative than another if the posteriors generated by the former mapping are a mean-preserving spread of the posteriors generated by the latter (Blackwell, 1953).

Definition 14.

[26] Of course, as discussed in the next subsection, messages can be vague even when communication is unsuccessful.

- Message $s \in \mathcal{S}$ is precise if there exists $\omega \in \Omega$ such that $m|\omega(s) > 0$ and, for every $\omega' \in \Omega$ with $\omega' \neq \omega$, $m|\omega'(s) = 0$. Otherwise, it is vague.
- Encoding mapping $\{m|\omega\}_{\omega\in\Omega}$ is vague if there exists an $\omega \in \Omega$ and $s \in \mathcal{S}$ such that $m|\omega(s) > 0$ and $s \in \mathcal{S}$ is vague. Otherwise, it is precise.
- Encoding mapping $\{m|\omega\}_{\omega\in\Omega}$ is more vague than encoding mapping $\{m'|\omega\}_{\omega\in\Omega}$ if $\{m'|\omega\}_{\omega\in\Omega}$ is more informative than $\{m|\omega\}_{\omega\in\Omega}$ according to the Blackwell order.

If communication is successful and the sender's encoding mapping is precise, then the receiver develops completely accurate posterior beliefs about the state of the world and full meaningful information is transmitted as discussed in Section 3.3. If, conversely, the sender's encoding mapping is vague, the receiver develops correct but non-degenerate posterior beliefs about the state of the world. Of course, in the context of successful communication, the expected value of communication and the amount of information transmitted to the receiver is weakly lower for more vague encoding mappings.

Despite the negative relationship between the degree of vagueness of the sender's encoding mapping and the value of information to the receiver in the context of successful communication, the sender's use of vague encodings should not necessarily be considered nefarious. In a richer model in which information is costly to acquire, transmit, or process, vague messages would naturally arise for the sake of parsimony. For instance, a sender who knows the receiver's decision problem and whose interests are aligned with those of the receiver might optimally choose to reveal the minimum amount of information that allows the receiver to pick appropriate actions.

In Definition 14, I introduced a comparative notion of vagueness for entire encoding mappings. In the context of a conventional language $(\mathcal{S}^*, \ell)$, a related notion of vagueness can be defined for individual messages as well.

Definition 15. In the context of conventional language $(\mathcal{S}^*, \ell)$, message s_T is more vague than message $s_{T'}$ if $T' \subseteq T$, for $T', T \in 2^\Omega \setminus \{\emptyset\}$.

For instance, if the set of states of the world is $\Omega = \{\omega_1, \omega_2, \omega_3\}$, message s_T indicating subset $T = \{\omega_2, \omega_3\}$ is more vague than message $s_{T'}$ indicating subset $T' = \{\omega_3\}$.

In the presence of a conventional language, the notions of comparative vagueness defined for entire encoding mappings and the one defined for individual messages are closely related, as described in the following remark.

Remark 3. For any encoding mapping that is consistent with the linguistic convention, if each message $s_{T'}$ sent with positive probability is replaced by a more vague message s_T, the resulting encoding mapping $\{m|\omega\}_{\omega\in\Omega}$ is more vague than the original one $\{m'|\omega\}_{\omega\in\Omega}$. Furthermore, if the replacement process is such that the resulting encoding mapping still follows the linguistic convention, communication remains successful, but less meaningful information is transmitted.

The claims above follow immediately from the observation that the procedure of replacing messages with more vague ones is the description of a particular type of garbling (Blackwell, 1951, 1953; Marschak and Miyasawa, 1968). Therefore, in the context of a conventional language, the degree of vagueness of individual messages and the degree of vagueness of the entire encoding mapping are tightly linked; furthermore, for successful communication, the degree of vagueness of messages is related to canonical notions of information transmission in an arguably intuitive way.

5.5 Misunderstandings

Misunderstandings occur whenever communication is inadvertently misleading. In other words, they occur whenever the sender has honest intentions, but the receiver interprets messages incorrectly. As a result of misunderstandings in communication, the receiver develops inaccurate beliefs about the state of the world.

Definition 16. Misunderstandings occur whenever: i) for every $\omega \in \Omega$ and $s \in \mathcal{S}$ with $m|\omega(s) > 0$, $\hat{\beta}|s = p|s$ and, ii) there exist $\omega \in \Omega$ and $s \in \mathcal{S}$ with $m|\omega(s) > 0$ for which $p|s \neq \beta|s$.

Whenever a misunderstanding occurs, the receiver is a biased decoder; therefore, the results from Sections 3 and 4 apply straightforwardly. In particular, the following properties hold:[27]

Corollary 3.

- *Compared to the counterfactual scenario in which the receiver interprets messages correctly,*
  - *misunderstandings lower the value of communication to the receiver in all decision problems $\mathcal{D}(\mathcal{A}, u)$.*
  - *misunderstandings lower the amount of meaningful information transmitted to the receiver for all measures of meaningful information introduced in Section 4.*
- *Compared to the counterfactual scenario of no communication,*
  - *there exist misunderstandings and decision problems $\mathcal{D}^*(\mathcal{A}, u)$ for which the value of communication is strictly negative.*
  - *there exist misunderstandings and measures of meaningful information among the ones introduced in Section 4 for which the amount of meaningful information transmitted is strictly negative.*

Compared to the case in which the receiver interprets messages correctly, misunderstandings always lower the value of communication and the amount of meaningful information transmitted

[27] For each bullet point in Corollary 3, one of the two sub-bullet points is redundant. I include the redundant sub-bullet points for the sake of exposition. I employ a similar strategy in the next subsections as well.

to the receiver. The result follows immediately from the definition of the value of information to a biased decoder and can transparently be seen from the decomposition in Proposition 1. Compared to the counterfactual scenario of no communication, the results are more ambiguous, as discussed in Section 3.2. Specifically, some misunderstandings lead to a non-negative value of information in all decision problems and a non-negative amount of meaningful information transmitted for all measures of meaningful information characterized in Section 4. Conversely, for other misunderstandings, there exist decision problems in which the value of communication is strictly negative and measures of meaningful information transmission for which the amount of meaningful information transmitted is strictly negative.

As discussed in Proposition 5, certain kinds of misunderstandings are particularly pernicious in that they lead to a negative value of communication in all decision problems and to a negative amount of meaningful information transmitted for all measures of meaningful information characterized in Section 4. Such pernicious misunderstandings are especially intuitive to grasp in the binary environment, as shown in Proposition 2. Under the assumption of a common prior, if the receiver misunderstands which signal is diagnostic of which state of the world, the value of communication to the receiver is negative in all decision problems, and so is the amount of meaningful information transmitted.

The flip-side of the statement above is that, in the binary environment, the epistemic requirements for the sender to make sure her messages do not produce misunderstandings that are harmful to the receiver are quite simple. Specifically, as shown in Proposition 3 and Corollary 1, the only piece of knowledge that the sender needs to possess in order to guarantee a positive value of communication to the receiver and a non-negative amount of meaningful information transmitted is knowledge of which message the receiver considers diagnostic of which state of the world. The epistemic requirements to avoid harmful misunderstandings do not change if the sender and the receiver are assumed to have incorrect and possibly different beliefs about the prior distribution of states of the world, and if the assumptions of Bayesian rationality and of common knowledge of Bayesian rationality are dropped.

As mentioned in the discussion of vagueness, vague messages may well be misunderstood. Since both vague and precise encodings can be misunderstood and since vague encodings contain less information according to canonical measures of information such as Shannon's entropy than precise encodings, one might be tempted to conclude that, in order to transmit as much meaningful information as possible, a sender should employ a precise encoding. Outside the binary environment, such conclusion is, in general, unwarranted. In fact a sender might increase the amount of meaningful information transmitted to the receiver by using a vague rather than a precise encoding. The intuition is that the kind of misunderstandings that occur after after messages sent from precise encodings might be more severe than the kind of misunderstandings that occur after messages sent from vague encodings. Therefore, once again, the use of vague encoding need not necessarily

be a nefarious strategy that the sender adopts to keep the receiver in the dark; in a richer model in which the probability of mistakes in interpretation is a function of the amount of information transmitted according to conventional measures, the use of vague encodings might be a cautious benevolent strategy.

5.6 Deception and Lies

Deception and lying occur whenever the sender intentionally and successfully sends a message that misleads the receiver. As conventionally defined, deception and lies are local properties of communication in that they are properties of messages sent in particular states of the world, rather than properties of an entire encoding mapping. As such, they are more amenable to be analyzed from the perspective of cheap talk than from that of persuasion.

When adopting the perspective of cheap talk and considering a sender of type $\omega \in \Omega$, it is necessary to specify: i) the behavior of types $\omega' \in \Omega$ with $\omega' \neq \omega$, ii) the receiver's beliefs about the behavior of types ω', and iii) the beliefs held by a sender of type ω about the behavior of types ω'. For the purpose of this section, I assume that the behavior of types $\omega' \neq \omega$ is fixed and fully predictable, meaning that both the sender and the receiver have degenerate and correct beliefs about the behavior of types $\omega' \neq \omega$. Formally, I assume that for every $\omega, \omega' \in \Omega$ with $\omega' \neq \omega$, $\hat{\mu}|\omega' = \mu|\omega' = m|\omega'$. Fixing the behavior of types $\omega' \neq \omega$ and assuming it is predictable allows me to isolate the effects of lying and deception.

5.6.1 Deception

The notion of deception developed in this paper builds on a canonical definition of deception in philosophy, namely "to intentionally cause to have a false belief that is known or believed to be false" (Mahon 2016, Carson, 2009).[28] Such definition is arguably too restrictive for the purpose of this paper because it appears to require limiting attention to: i) binary beliefs that only take the values of true and false, rather than probabilistic beliefs that capture degrees of uncertainty, and ii) communication environments capable of inducing only such binary beliefs. In this section, I develop a more general definition of deception that coincides with the canonical philosophical definition whenever the restrictive assumptions about beliefs being binary and about the communication environment inducing only such beliefs are met.

[28]Neither Mahon (2016) nor Carson (2009) considers the definition above to be the most satisfactory definition of deception; however, both Mahon and Carson treat it as a conventional definition of deception in philosophy and use it as a key stepping stone for their ultimate definitions. The ultimate definitions of deception in Mahon (2016) and Carson (2009) take care of a few cases not encompassed by the definition above (e.g., instances in which the sender does not know the state of the world and yet sends a signal pretending that she does). I conjecture that that most such additional cases can be addressed by expanding the current framework, but I have not pursued that line of inquiry. In this paper, I tried to capture, as a starting point, what Mahon refers to as the "traditional definition of deception" (Mahon, 2016).

The definition of deception in this paper relies on the idea that, after a particular state of the world is realized and communication occurred, the receiver's beliefs might differ in the probability they assign to the realized state of the world. A "true belief" can then be defined as a degenerate belief that puts probability mass equal to unity on the realized state of the world; similarly, a "false belief" can be defined as a degenerate belief that puts probability mass equal to zero on the realized state of the world. Non-degenerate beliefs can be compared to one another in terms of accuracy simply by looking at the probability that they assign to the realized state of the world.[29]

The canonical philosophical definition of deception consists of two separate components: one that relates solely to the sender's intentions and one that also involves the receiver's interpretations. Specifically, in order for a sender to deceive a receiver, the sender has to: i) deliberately attempt to induce the receiver to hold a belief that the sender knows or believes to be false and, ii) succeed at doing so. In line with the philosophical definition, I draw a distinction between deceptive intentions and realized deception.

Definition 17.

- Message $s \in \mathcal{S}$ is intended to be deceptive in realized state of the world $\omega \in \Omega$ if $\hat{\beta}|s(\omega) < p(\omega)$ and $m|\omega(s) = 1$.
- Message $s \in \mathcal{S}$ is deceptive in realized state of the world $\omega \in \Omega$ if it is intended to be deceptive in state $\omega \in \Omega$ and $\hat{\beta}|s = \beta|s$.

The formalization of intended deception above tracks two core features of the canonical philosophical definition of intended deception. First, it captures the idea of a false belief, albeit in an environment in which beliefs are probabilistic rather than binary. Specifically, by sending message $s \in \mathcal{S}$ with deceptive intentions, the sender expects her message to induce the receiver to reduce the probability that she assigns to the realized state of the world and, thus, to hold less accurate beliefs. Second, the formalization of intended deception captures intentionality. In particular, in each state of the world $\omega \in \Omega$, the sender always has access to a contingent strategy that is not deceptive.[30] Specifically, if, in state of the world $\omega \in \Omega$, there exists an $s \in \mathcal{S}$ such that $\hat{\beta}|s(\omega) < p(\omega)$, then, by common knowledge of Bayesian rationality, there must exist an $s' \in \mathcal{S}$ with $s' \neq s$ such that $\hat{\beta}|s'(\omega) > p(\omega)$.[31] But then, in state of the world $\omega \in \Omega$, the sender has at least two options: she can send a message that she thinks will induce the receiver to hold more accurate beliefs about the realized state of the world or she can send a message that she thinks will induce the receiver to hold less accurate beliefs about it. The sender's choice to send the latter message rather than the former

[29] A similar taxonomy can be found in Sobel (2020, 2023). The way in which Sobel thinks about the accuracy of beliefs, however, is different from simply comparing the probability that different beliefs assign to the realized state of the world.

[30] According to Sobel (2020), this is an important desideratum of a definition of deception.

[31] Common knowledge of Bayesian rationality is stronger an assumption than needed for the definition of deception. In fact, it is sufficient that the sender believes that the receiver satisfy an assumption similar to Assumption 1

reveals her deceptive intentions. Therefore, according to the formalization of intended deception above, any attempt at deception is an intentional choice made by the sender for the purpose of causing the receiver to hold less accurate beliefs.

Of course, intended deception need not necessarily translate into successful deception. Successful deception not only requires the sender to attempt to induce the receiver to hold less accurate beliefs; the sender has to actually succeed in inducing such beliefs. For that reason, the formalization of deception in Definition 17 requires the sender's second order beliefs to be correct; in that case, the attempt at deception will be successful because the receiver will interpret messages precisely as the sender expects her to. As shown in the next subsection, the receiver might thwart the sender's attempt at deception by interpreting messages in a way that is not in line with the sender's second-order beliefs. Furthermore, the receiver can always avoid deception by interpreting the sender's messages as noise.[32]

As I alluded to at the beginning of this section, the formalization of deception in Definition 17 coincides with the canonical philosophical definition when the receiver's beliefs after communication can only take the values of true and false and when the communication environment is such as to be able to induce only such binary beliefs. Specifically, suppose the realized state of the world is $\omega \in \Omega$ and suppose that, for every $s \in \mathcal{S}$, $\hat{\beta}|s(\omega) = \beta|s(\omega) \in \{0,1\}$. Suppose the sender sends a message $s \in \mathcal{S}$ that she thinks will induce the receiver to hold belief $\hat{\beta}|s(\omega) = 0$ and that, indeed, induces the receiver to hold belief $\beta|s(\omega) = 0$. Then, by sending message s, the sender intentionally caused the receiver to hold a false belief – namely the belief that the state of the world is not ω – that the sender knows to be false. Therefore, in this setting, the canonical philosophical definition of deception and the one from Definition 17 coincide.

Since deception is both intentional and avoidable, the sender must expect to benefit from deception whenever she sends a message with deceptive intentions. Therefore, deception, when successful, should benefit the sender. How does it affect the welfare of the receiver? The following corollary and proposition provide the answer.

Corollary 4. *Compared to the counterfactual scenario in which the receiver interprets messages correctly,*

- *receiving a deceptive message in state of the world $\omega \in \Omega$ lowers the value of communication to the receiver in all decision problems $\mathcal{D}(\mathcal{A}, u)$.*
- *receiving a deceptive message in state of the world $\omega \in \Omega$ lowers the amount of meaningful information that is transmitted to the receiver for all measures of meaningful information introduced in Section 4.*

[32] When the receiver interprets the sender's messages as noise, deception cannot occur because it cannot be the case that $\hat{\beta}|s = \beta|s < p(\omega)$. This is another important desideratum of a definition of deception according to Sobel (2020).

Under the additional assumptions imposed to prove Proposition 4, the following result also holds:

Proposition 8. *Compared to the counterfactual scenario of no communication, there always exists a decision problem $\mathcal{D}^*(\mathcal{A}, u)$ for which receiving a deceptive message in state of the world $\omega \in \Omega$ leads to a strictly negative value of communication.*

The first two statements in Corollary 4 are similar to the first two statements of the corollary describing the consequences of misunderstandings. In fact, the misspecifications caused by deception are a strict subset of those generated by misunderstandings. Therefore, compared to the counterfactual of no deception, deception lowers the receiver's welfare in all decision problems and the amount of meaningful information that is transmitted to the receiver for all measures of meaningful information introduced in Section 4.

The statement in Proposition 8 is stronger than the corresponding statement for misunderstandings. As discussed in the context of Corollary 3, certain types of misunderstandings yield a non-negative value of communication in all decision problems $\mathcal{D}(\mathcal{A}, u)$. Such situations do not arise in the context of deception. In particular, whenever a sender of type $\omega \in \Omega$ sends a deceptive message to the receiver, there always exists a decision problem $\mathcal{D}^*(\mathcal{A}, u)$ for which the value of communication to the receiver is strictly negative.

Notice that, compared to the counterfactual scenario of no communication, deception need not necessarily reduce the receiver's welfare in all decision problems or lead to a negative amount of meaningful information transmitted. The intuition is that the sender and the receiver's incentives might be sufficiently misaligned for the sender to benefit from deceiving the receiver, and yet sufficiently aligned for the receiver to obtain a higher payoff after deceptive communication than in the counterfactual scenario of no communication.

The effects of deception on the receiver's welfare become robustly negative in the case of a binary state of the world and a binary signal, because, in the binary environment, the intuition from the previous paragraph does not apply. The reason is as follows: in the binary environment, the interests of a sender of type $\omega \in \{\omega_1, \omega_2\}$ and of the receiver can either be aligned, in the sense that both the sender and the receiver are better off when the receiver takes the action associated to the signal that is diagnostic of state of the world ω, or misaligned, in the sense that the receiver is better off when the she takes the action associated to the signal that is diagnostic of state of the world ω and the sender is worse off. Only in the latter case does a sender of type ω engage in deception; therefore, it is natural to expect that, from the ex-post perspective, the receiver's welfare should be lower in all decision problems. Since deception in one state of the world affects the encoding mapping in such a way as to mislead the agent in the other state of the world as well, the receiver's welfare will also be lower from the ex-ante perspective.

Corollary 5. *If message $s \in \{s_1, s_2\}$ is deceptive in state of the world $\omega \in \{\omega_1, \omega_2\}$, then:*

- *The value of communication to the receiver is non-positive in all decision problems $\mathcal{D}(\mathcal{A}, u)$ from both the ex-ante and the ex-post perspective.*

- *The amount of meaningful information that is transmitted to the receiver is non-positive for all measures of meaningful information introduced in Section 4.*

The result above is a corollary of Proposition 7 in Sobel (2020) and of Proposition 2 in this paper.

5.6.2 Lies

In line with Sobel (2020), I define a lie as a statement that the sender makes knowing that the statement is false. Therefore, in order to to be able to lie, the sender has to rely on the existence of a linguistic convention as discussed in Section 5.1.1. Specifically, in the context of a conventional language $(\mathcal{S}^*, \ell)$ in which each message $s_T \in \mathcal{S}^*$ has the commonly understood meaning "the realized state of the world $\omega \in \Omega$ is an element of $T \in 2^\Omega \setminus \{\emptyset\}$," message $s_T \in \mathcal{S}^*$ sent by a sender of type $\omega \in \Omega$ is a lie if $\omega \notin T$. In other words, a message $s_T \in \mathcal{S}^*$ is defined to be a lie in state of the world $\omega \in \Omega$ if the conventional meaning of the message rules out the occurrence of state of the world $\omega \in \Omega$, and yet the sender sends message s in state of the world ω.

Definition 18. Given conventional language $(\mathcal{S}^*, \ell)$, message $s \in \mathcal{S}^*$ is a lie in realized state of the world $\omega \in \Omega$ if $\ell|\omega(s) = 0$ and $m|\omega(s) = 1$.

A conventional language generates a set of clear expectations about the meaning of messages that helps the sender and the receiver achieve successful communication. However, such clear expectations can be abused by the sender for the purpose of misleading the receiver. In particular, a conventional language allows the receiver to lie by making false statements about the realized state of the world.

Lies need not be deceptive if the receiver anticipates them and, as a result, interprets messages in a way that is different from the one prescribed by the linguistic convention (Kartik, Ottaviani, and Squintani, 2007; Sobel, 2020). For instance, messages sent by notorious liars such as the boy who cried wolf from the eponymous fable are often ignored. Similarly, inflated claims by sellers about the quality of their products are generally discounted by expert buyers (Sobel, 2020).

Whenever the receiver interprets messages according to a linguistic convention, however, lies are deceptive. Similarly, in the context of a conventional language, all deceptive messages are lies.

Proposition 9. *Given conventional language $(\mathcal{S}^*, \ell)$, if the sender believes the receiver interprets messages according to the linguistic convention and if the sender is correct, message $s \in \mathcal{S}^*$ is a lie in state of the world $\omega \in \Omega$ if and only if it is deceptive in state of the world $\omega \in \Omega$.*

Since, in the presence of a conventional language, all lies are deceptive, Corollary 4 and Corollary 5 also describe the consequences of lying for the welfare of the receiver and for the amount of meaningful information transmitted.[33]

5.6.3 Relation to Sobel (2020, 2023)

In two recent papers, Sobel provides a definition of lying, a host of definitions of deception, and studies the effects of lying and deception on the receiver's welfare (Sobel, 2020, 2023). My analysis and Sobel's differ along two main dimensions: first, the perspective from which we study the consequences of lying and deception. Specifically, Sobel evaluates the welfare consequences of lying and deception from the interim perspective (i.e., after the state of the world is realized). In order to discuss the implications of lying and deception for meaningful information transmission, I primarily adopt an ex-ante perspective, though I do present some results from the interim perspective as well. Second, Sobel's definitions of deception and my definition are different and have different implications.

Sobel (2020, 2023) defines a sender's message as deceptive if the beliefs it induces are in some sense less accurate than the beliefs the sender could have induced by sending an alternative message. Different definitions of deception in Sobel (2020, 2023) formalize the notion of a "less accurate belief" in different ways. I define a message as deceptive if the beliefs it induces place a lower probability on the realized state of the world than do the sender's prior beliefs. Therefore, the main differences between Sobel's definitions and my definition relate to: i) the counterfactual scenario against which a message is benchmarked in order to be considered deceptive, and ii) the way in which we formalize the notion of a "less accurate belief".

One of the definitions in Sobel (2023), namely that of KL deception, defines a belief to be less accurate than another if it places a lower probability on the realized state of the world than the other. In that case, Sobel's definition of a less accurate belief coincides the one implicit in my definition of deception, and the main difference between our definitions reduces to whether the appropriate counterfactual against which to benchmark a deceptive messages is the receiver's prior or a belief that the sender could induce by sending an alternative message.

The difference in definitions affects whether vague messages are considered deceptive and whether deception is consistent with equilibrium. According to Sobel, if the sender has access to a precise message in a certain state of the world but chooses to send a vague message instead, then the sender's message is deceptive. Therefore, in the context of a conventional language, all messages that are vague according to Definition 15 are deceptive according to Sobel's definitions whenever the receiver's interpretations follow the linguistic convention. Considering vague mes-

[33] Proposition 5 does not apply directly to the case of lying, because the assumption of the existence of a conventional language implies that, in the case in which the state of the world is binary, the space of signals has three rather than two elements. It is easy to show, however, that the third signal, which if interpreted according to the linguistic convention reveals no information about the state of the world, does not affect the conclusions of the proposition.

sages deceptive seems appropriate whenever there is a presumption that the sender should disclose full information to the receiver; for instance, in the context of fiduciary duties. However, such presumption is absent in many contexts and, therefore, it is not obvious that partial information disclosure should necessarily be regarded as deceptive. For instance, suppose a celebrity answers "no comment" to a tabloid journalist inquiring about whether she is currently in a romantic relationship. I surmise that the celebrity's refusal to disclose information to the tabloid journalist would generally not be considered deceptive, even though the information withheld might be valuable to the journalist.

Separating the concepts of vagueness and deception seems to also be desirable from a conceptual standpoint because vagueness and deception impair meaningful information transmission in qualitatively different ways. Specifically, vagueness, as I define it, is a property of the encoding mapping; therefore, it affects the amount of meaningful information transmitted to biased and unbiased decoders alike. Deception, as I define it, relies on the receiver incorrectly decoding the sender's messages; therefore, it affects only the amount of meaningful information transmitted to a biased decoder.

The distinction above relates to whether deception can occur in equilibrium. According to Sobel's definition, deception is consistent with equilibrium; therefore, the receiver can be deceived even when she has full knowledge of the sender's encoding mapping. Conversely, according to my definition, deception is an out-of-equilibrium phenomenon that can only occur whenever the receiver has incorrect beliefs about the sender's encoding mapping.[34] Therefore, according to my definition, the receiver always benefits from learning the sender's actual encoding mapping, because such knowledge helps her interpret the sender's messages more accurately, prevents her from being deceived, and, as a result, improves her welfare.

6 Conclusion

Information not only needs to be encoded in a signal structure; it also needs to be decoded. In this paper, I considered the plight of an agent who does not necessarily decode signals correctly. I developed a set of results about the value of information for such an agent and introduced a set of measures of the amount of information that is transmitted to such an agent. I showed how the language developed to study biased decoding can help shed light on foundational concepts in communication such as successful communication, vagueness, misunderstandings, deception, and lies.

[34]This can be seen from the fact that, in Definition 17, the requirements that $m|\omega(s) = 1$, that $\hat{\beta}|s(\omega) < p(\omega)$, and that $\hat{\beta}|s(\omega) = \beta|s(\omega)$ imply $\beta|s(\omega) < p(\omega) < p|s(\omega)$. Therefore, the receiver cannot have equilibrium beliefs.

References

Johannes Abeler, Daniele Nosenzo, and Collin Raymond. Preferences for truth-telling. *Econometrica*, 87, 2019.

János Aczél and Zoltán Daróczy. *On Measures of Information and Their Characterization*. Academic Press Inc., 1975.

Sandro Ambuehl and Shengwu Li. Belief updating and the demand for information. *Games and Economic Behavior*, 109, 5 2018.

Kenneth J. Arrow. The theory of discrimination. In *Discrimination in Labor Markets*. Princeton University Press, 1974.

Ned Augenblick, Eben Lazarus, , and Michael Thaler. Overinference from weak signals and underinference from strong signals. *Working Paper*, 2023.

Cuimin Ba, J. Aislinn Bohren, and Alex Imas. Over- and underreaction to information. *Working Paper*, 2023.

Abhijit Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3), 1992.

Daniel J. Benjamin. Errors in probabilistic reasoning and judgment biases. In *Handbook of Behavioral Economics*, volume 2, pages 69–186. 2019.

Robert H. Berk. Limiting behavior of posterior distributions when the model is incorrect. *Annals of Mathematical Statistics*, 1966.

Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5), 1992.

David Blackwell. Comparison of experiments. In *Berkeley Symposium on Mathematical Statistics and Probability*, 1951.

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24(2), 1953.

Andreas Blume, Ernest K Lai, and Wooyoung Lim. Strategic information transmission: A survey of experiments and theoretical foundations. In *Handbook of Experimental Game Theory*. 2020.

Aislinn J. Bohren and Daniel N. Hauser. Learning with heterogeneous misspecified models. *Econometrica*, 89(6), 2021.

Aislinn J. Bohren, Kareem Haggag, Alex Imas, and Devin G. Pope. Inaccurate statistical discrimination. *Working Paper*, 2021.

Pedro Bordalo, John Conlon, Nicola Gennaioli, Spencer Y. Kwon, and Andrei Shleifer. Memory and probability. *The Quarterly Journal of Economics*, 138(1), 2023.

Luca Braghieri. Political correctness, social image, and information transmission. *Working Paper*, 2021.

Lev M. Bregman. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3), 1967.

Benjamin Bushong and Tristan Gagnon-Bartsch. Learning with misattribution of reference dependence. *Journal of Economic Theory*, 203, 7 2022.

Thomas L. Carson. Lying, deception, and related concepts. In *The Philosophy of Deception*. Oxford University Press, 2009.

Simone Cerreia-Vioglio, Lars Peter Hansen, Fabio Maccheroni, and Massimo Marinacci. Making decisions under model misspecification. *Working Paper*, 2022.

Marco Cipriani and Antonio Guarino. Herd behavior in a laboratory financial market. *The American Economic Review*, 95(5), 2005.

Bernat Corominas-Murtra, Jordi Fortuny, and Ricard V. Solé. Towards a mathematical theory of meaningful communication. *Scientific Reports*, 4, 4 2014.

Vincent P Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, 50, 1982.

A. Philip Dawid. The geometry of proper scoring rules. *Annals of the Institute of Statistical Mathematics*, 59, 3 2007.

A. Philip Dawid and Monica Musio. Theory and applications of proper scoring rules. *Metron*, 72, 1 2014.

A. Philip Dawid and Paola Sebastiani. Coherent dispersion criteria for optimal experimental design. *The Annals of Statistics*, 27, 1999.

Philip Dawid. Coherent measures of discrepancy, uncertainty and dependence, with applications to bayesian predictive experimental design. *Working Paper*, 1998.

Ferdinand de Saussure. *Course in General Linguistics*. 1916.

Marvin Deversi, Alessandro Ispano, and Peter Schwardmann. Spin doctors: An experiment on vague disclosure. *European Economic Review*, 139, 2021.

Ward Edwards. Conservatism in human information processing. In *Formal Representation of Human Judgment*. Wiley, 1968.

Ignacio Esponda, Demian Pouzo, and Yuichi Yamamoto. Asymptotic behavior of bayesian learners with misspecified models. *Journal of Economic Theory*, 195, 7 2021.

Erik Eyster and Matthew Rabin. Cursed equilibrium. *Econometrica*, 73(5), 2005.

Joseph Farrell. Meaning and credibility in cheap-talk games. *Games and Economic Behavior*, 5, 1993.

Joseph Farrell and Matthew Rabin. Cheap talk. *Journal of Economic Perspectives*, 10, 1996.

Alexander Frankel and Emir Kamenica. Quantifying information and uncertainty. *American Economic Review*, 109, 2019.

Mira Frick, Ryota Iijima, and Yuhta Ishii. Welfare comparisons for biased learning. *Working Paper*, 2022.

Drew Fudenberg, Gleb Romanyuk, and Philipp Strack. Active learning with a misspecified prior. *Theoretical Economics*, 12, 9 2017.

Drew Fudenberg, Giacomo Lanzani, and Philipp Strack. Limit points of endogenous misspecified learning. *Econometrica*, 89, 2021.

Uri Gneezy. Deception: the role of consequences. *The American Economic Review*, 95(1), 2005.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102, 3 2007.

Jerry R. Green and Nancy L. Stokey. A two-person game of information transmission. *Journal of Economic Theory*, 135, 2007.

Sanford J. Grossman. The informational role of warranties and private disclosure about product quality. *The Journal of Law and Economics*, 24(3), 1981.

Kevin He. Mislearning from censored data: The gambler's fallacy and other correlational mistakes in optimal-stopping problems. *Theoretical Economics*, 17:1269–1312, 2022.

Benjamin Hébert and Michael Woodford. Rational inattention when decisions take time. *Working Paper*, 2022.

Paul Heidhues, Botond Koszegi, and Philipp Strack. Unrealistic expectations and misguided learning. *Econometrica*, 86, 2018.

Paul Heidhues, Botond Koszegi, and Philipp Strack. Convergence in models of misspecified learning. *Theoretical Economics*, 16, 2021.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101, 10 2011.

Navin Kartik. Strategic communication with lying costs. *The Review of Economic Studies*, 76: 1359–1395, 2009.

Navin Kartik, Marco Ottaviani, and Francesco Squintani. Credulity, lies, and costly talk. *Journal of Economic Theory*, 134, 5 2007.

Goedker Katrin, Peiran Jiao, and Paul Smeets. Investor memory. *Working Paper*, 2022.

Solomon Kullback and Richard Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 1951.

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

David K. Lewis. *Convention: a Philosophical Study.* Wiley-Blackwell, 1969.

Barton L. Lipman. Why is language vague? *Working Paper*, 2009.

John Locke. *An Essay Concerning Humane Understanding.* 1689.

Ryan Lowe, Jakob Foerster, Y-Lan Boureau, Joelle Pineau, and Yann Dauphin. On the pitfalls of measuring emergent communication. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems*, 2019.

James E. Mahon. The definition of lying and deception. *Stanford Encyclopedia of Philosophy*, 2016.

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

David C. Krakauer Martin A. Nowak. The evolution of language. In *Proceedings of the National Academy of Science*, volume 96, 1999.

Yusufcan Matsalioglu, A. Yesim Orhun, and Collin Raymond. Intrinsic information preferences and skewness. *Working Paper*, 2017.

Jeffrey Mensch. Rational inattention and the monotone likelihood ratio property. *Journal of Economic Theory*, 196, 2021.

Merriam-Webster-Dictionary. Communication, 2023. URL https://www.merriam-webster.com/dictionary/communication#citations.

Paul Milgrom. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 12(2), 1981.

Paul Milgrom and John Roberts. Relying on the information of interested parties. *The RAND Journal of Economics*, 17, 1986.

Stephen Morris. Political correctness. *Journal of Political Economy*, 109, 2001.

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

Stephen Morris and Philipp Strack. The wald problem and the relation of sequential sampling and ex-ante information costs. *Working Paper*, 2019.

Yaw Nyarko. Learning in misspecified models and the possibility of cycles. *Journal of Economic Theory*, 55, 1991.

Marco Ottaviani and Peter Norman Sørensen. Reputational cheap talk. *The RAND Journal of Economics*, 37, 2006.

Luciano Pomatto, Philipp Strack, and Omer Tamuz. The cost of information. *American Economic Review*, Forthcoming, 2023.

Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27, 1948.

Claude E. Shannon. The bandwagon. *IRE Transactions on Information Theory*, 2, 4 1956.

Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.

Christopher A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50, 4 2003.

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

Joel Sobel. On the relationship between damage and deception. *Working Paper*, 2023.

Michael Spence. Job market signaling. *The Quarterly Journal of Economics*, 87(3), 1973.

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

Appendix: Proofs

Proof of Proposition 1

$$\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)=E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{\beta|s},\omega\right)\right]\right]-E_{\omega\sim p}\left[u\left(a_{\beta},\omega\right)\right]=$$
$$=E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{\beta|s},\omega\right)\right]\right]+E_{\omega\sim p}\left[u\left(a_{p},\omega\right)\right]-E_{\omega\sim p}\left[u\left(a_{p},\omega\right)\right]+$$
$$+E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s},\omega\right)\right]\right]-E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s},\omega\right)\right]\right]-E_{\omega\sim p}\left[u\left(a_{\beta},\omega\right)\right]=$$
$$=\left\{E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s},\omega\right)\right]\right]-E_{\omega\sim p}\left[u\left(a_{p},\omega\right)\right]\right\}+d\left(p,\beta\right)-E_{s\sim m}\left[d\left(p|s,\beta|s\right)\right]$$

Proof of Proposition 3

I will first show that statements 1. and 2. are equivalent, starting with 1. implying 2. Suppose $\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}}\succeq\{m'|\omega\}_{\omega\in\{\omega_1,\omega_2\}}$ according to the Blackwell order. Consider $p|s_1(\omega_1)$, $p|s_2(\omega_1)$, $\tilde{p}|s_1(\omega_1)$, $\tilde{p}|s_2(\omega_1)$. By one of the equivalent characterizations of the Blackwell order, we know that $p|s(\omega_1)$ is a mean-preserving-spread of $\tilde{p}|s(\omega_1)$ (Blackwell, 1951,1953). Furthermore, we assumed that $m|\omega_1(s_1)\geq m|\omega_2(s_1)$ and $\tilde{m}|\omega_1(s_1)\geq\tilde{m}|\omega_2(s_1)$. Together, the two imply that $p|s_2(\omega_1)\leq\tilde{p}|s_2(\omega_1)\leq p(\omega_1)\leq\tilde{p}|s_1(\omega_1)\leq p|s_1(\omega_1)$. Therefore, $[\tilde{p}|s_2(\omega_1),\tilde{p}|s_1(\omega_1)]\subseteq[p|s_2(\omega_1),p|s_1(\omega_1)]$. By Proposition 2 and Assumption 1, we know that $U_{\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}}}$ can be written as

$$U_{\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}}}=\left\{\beta,\{\beta|s\}_{s\in\mathcal{S}}\in\Delta(\Omega)^{(|\mathcal{S}|+1)}:p|s_2(\omega_1)\leq\beta|s_2(\omega_1)\leq\beta(\omega_1)\leq\beta|s_1(\omega_1)\leq p|s_1(\omega_1)\right\}$$

and $U_{\{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}}$ can be written as

$$U_{\{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}}=\left\{\beta,\{\beta|s\}_{s\in\mathcal{S}}\in\Delta(\Omega)^{(|\mathcal{S}|+1)}:\tilde{p}|s_2(\omega_1)\leq\beta|s_2(\omega_1)\leq\beta(\omega_1)\leq\beta|s_1(\omega_1)\leq\tilde{p}|s_1(\omega_1)\right\}$$

But then, it is clear that $U_{\{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}}\subseteq U_{\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}}}$. Therefore, 1. implies 2.

Let's now show that 2. implies 1. Suppose $U_{\{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}}\subseteq U_{\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}}}$. Then, it must be the case that $p|s_2(\omega_1)\leq\tilde{p}|s_2(\omega_1)\leq\tilde{p}|s_1(\omega_1)\leq p|s_1(\omega_1)$. But then, $p|s(\omega_1)$ is a mean-preserving-spread of $\tilde{p}|s(\omega_1)$ and $\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}}\succeq\{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}$ according to the Blackwell order. Therefore, 1. and 2. are equivalent.

Showing that 1. and 3. are equivalent follows an almost identical line of reasoning.

Let's now assume that $\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}}\succeq\{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}$ according to the Blackwell order, that $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq0$ and $\mathcal{V}\left(p,\{\tilde{p}|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq0$ in all decision problems $\mathcal{D}(\mathcal{A},u)$, and show that $\mathcal{W}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq\mathcal{W}\left(p,\{\tilde{p}|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)$ in all decision problems $\mathcal{D}(\mathcal{A},u)$.

Since $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq0$ and $\mathcal{V}\left(p,\{\tilde{p}|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq0$ in all decision problems $\mathcal{D}(\mathcal{A},u)$, we know $p|s_2(\omega_1)\leq\beta|s_2(\omega_1)\leq\beta|s_1(\omega_1)\leq p|s_1(\omega_1)$ and $\tilde{p}|s_2(\omega_1)\leq\beta|s_2(\omega_1)\leq\beta|s_1(\omega_1)\leq\tilde{p}|s_1(\omega_1)$. I claim that it has to be the case that $u\left(a_{\beta|s_1},\omega_1\right)\geq u\left(a_{\beta|s_2},\omega_1\right)$ and

$u\left(a_{\beta|s_2},\omega_2\right) \geq u\left(a_{\beta|s_1},\omega_2\right)$. Since $a_{\beta|s_1}$ and $a_{\beta|s_2}$ are subjectively optimal, we have

$$E_{\omega\sim\beta|s_1}\left[u\left(a_{\beta|s_1},\omega\right)\right] \geq E_{\omega\sim\beta|s_1}\left[u\left(a_{\beta|s_2},\omega\right)\right] \text{ and } E_{\omega\sim\beta|s_2}\left[u\left(a_{\beta|s_2},\omega\right)\right] \geq E_{\omega\sim\beta|s_2}\left[u\left(a_{\beta|s_1},\omega\right)\right]$$

Writing out the two inequalities explicitly, we have

$$\beta|s_1\left(\omega_1\right)u\left(a_{\beta|s_1},\omega_1\right)+\beta|s_1\left(\omega_2\right)u\left(a_{\beta|s_1},\omega_2\right) \geq \beta|s_1\left(\omega_1\right)u\left(a_{\beta|s_2},\omega_1\right)+\beta|s_1\left(\omega_2\right)u\left(a_{\beta|s_2},\omega_2\right)$$

$$\beta|s_2\left(\omega_1\right)u\left(a_{\beta|s_2},\omega_1\right)+\beta|s_2\left(\omega_2\right)u\left(a_{\beta|s_2},\omega_2\right) \geq \beta|s_2\left(\omega_1\right)u\left(a_{\beta|s_1},\omega_1\right)+\beta|s_2\left(\omega_2\right)u\left(a_{\beta|s_1},\omega_2\right)$$

Rearranging the two inequalities, we obtain

$$\beta|s_1\left(\omega_1\right)\left[u\left(a_{\beta|s_1},\omega_1\right)-u\left(a_{\beta|s_2},\omega_1\right)\right]+\left[1-\beta|s_1\left(\omega_1\right)\right]\left[u\left(a_{\beta|s_1},\omega_2\right)-u\left(a_{\beta|s_2},\omega_2\right)\right] \geq 0$$

$$\beta|s_2\left(\omega_1\right)\left[u\left(a_{\beta|s_1},\omega_1\right)-u\left(a_{\beta|s_2},\omega_1\right)\right]+\left[1-\beta|s_2\left(\omega_1\right)\right]\left[u\left(a_{\beta|s_1},\omega_2\right)-u\left(a_{\beta|s_2},\omega_2\right)\right] \leq 0$$

Therefore, $\left[u\left(a_{\beta|s_1},\omega_1\right)-u\left(a_{\beta|s_2},\omega_1\right)\right]$ and $\left[u\left(a_{\beta|s_1},\omega_2\right)-u\left(a_{\beta|s_2},\omega_2\right)\right]$ have to either both be zero or have different sign. Furthermore, since $\beta|s_1\left(\omega_1\right) \geq \beta|s_2\left(\omega_1\right)$, since the first inequality is non-negative, and the second is non-positive, it must be the case that $u\left(a_{\beta|s_1},\omega_1\right) \geq u\left(a_{\beta|s_2},\omega_1\right)$ and $u\left(a_{\beta|s_1},\omega_2\right) \leq u\left(a_{\beta|s_2},\omega_2\right)$ as initially claimed.

Consider the space in which the graph of $E_{\omega\sim q}\left[u\left(a,\omega\right)\right]$ lives. That space is $[0,1]\times\mathbb{R}$. In that space, the line segments joining point $\left(0,u\left(a_{\beta|s_1},\omega_2\right)\right)$ to $\left(1,u\left(a_{\beta|s_1},\omega_1\right)\right)$ and point $\left(0,u\left(a_{\beta|s_2},\omega_2\right)\right)$ to $\left(1,u\left(a_{\beta|s_2},\omega_1\right)\right)$ must cross exactly once in light of the arguments from the previous paragraph. Furthermore, we know $E_{\omega\sim\beta|s_1}\left[u\left(a_{\beta|s_1},\omega\right)\right] \geq E_{\omega\sim\beta|s_1}\left[u\left(a_{\beta|s_2},\omega\right)\right]$; similarly, we know that $E_{\omega\sim\beta|s_2}\left[u\left(a_{\beta|s_2},\omega\right)\right] \geq E_{\omega\sim\beta|s_2}\left[u\left(a_{\beta|s_1},\omega\right)\right]$. Taken together, the observations in this paragraph and the previous one imply that $\forall q\in\Delta\left(\Omega\right)$ with $q\left(\omega_1\right) \geq \beta|s_1\left(\omega_1\right)$, $E_{\omega\sim q}\left[u\left(a_{\beta|s_1},\omega\right)\right] \geq E_{\omega\sim q}\left[u\left(a_{\beta|s_2},\omega\right)\right]$ and that $\forall q\in\Delta\left(\Omega\right)$ with $q\left(\omega_1\right) \leq \beta|s_2\left(\omega_1\right)$, $E_{\omega\sim q}\left[u\left(a_{\beta|s_2},\omega\right)\right] \geq E_{\omega\sim q}\left[u\left(a_{\beta|s_1},\omega\right)\right]$.

Consider auxiliary decision problem $\mathcal{D}^*\left(\mathcal{A}^*,u^*\right)$, where $\mathcal{A}^*=\left\{a_{\beta|s_1},a_{\beta|s_2}\right\}$ and $u^*\left(a_{\beta|s_i},\omega_j\right)=u\left(a_{\beta|s_i},\omega_j\right)$ for $i,j\in\{1,2\}$. By Blackwell's characterization, we know that, if $\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}} \succeq \{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}$ according to the Blackwell order then

$$E_{s\sim m}\left[E_{\omega\sim p|s}\left[u^*\left(a_{p|s},\omega\right)\right]\right] \geq E_{s\sim\tilde{m}}\left[E_{\omega\sim\tilde{p}|s}\left[u^*\left(a_{\tilde{p}|s},\omega\right)\right]\right]$$

By the argument in the paragraph above, we know that, whenever $q\left(\omega_1\right) \geq \beta|s_1\left(\omega_1\right)$, the optimal action in decision problem $\mathcal{D}^*\left(\mathcal{A}^*,u^*\right)$ is $a_{\beta|s_1}$ and whenever $q\left(\omega_1\right) \leq \beta|s_2\left(\omega_1\right)$, the optimal action in decision problem $\mathcal{D}^*\left(\mathcal{A}^*,u^*\right)$ is $a_{\beta|s_2}$. But then, the biased decoder's behavior in decision problem $\mathcal{D}^*\left(\mathcal{A}^*,u^*\right)$ is optimal with respect to the true probabilities $q\left(\omega_1\right)$. Therefore, $E_{s\sim m}\left[E_{\omega\sim p|s}\left[u^*\left(a_{\beta|s},\omega\right)\right]\right]=E_{s\sim m}\left[E_{\omega\sim p|s}\left[u^*\left(a_{p|s},\omega\right)\right]\right]$ and $E_{s\sim\tilde{m}}\left[E_{\omega\sim\tilde{p}|s}\left[u^*\left(a_{\beta|s},\omega\right)\right]\right]=$

$E_{s\sim\tilde{m}}\left[E_{\omega\sim\tilde{p}|s}\left[u^*\left(a_{\tilde{p}|s},\omega\right)\right]\right]$. But then,

$$\mathcal{W}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)=$$

$$=E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{\beta|s},\omega\right)\right]\right]=E_{s\sim m}\left[E_{\omega\sim p|s}\left[u^*\left(a_{\beta|s},\omega\right)\right]\right]=E_{s\sim m}\left[E_{\omega\sim p|s}\left[u^*\left(a_{p|s},\omega\right)\right]\right]\geq$$

$$\geq E_{s\sim\tilde{m}}\left[E_{\omega\sim\tilde{p}|s}\left[u^*\left(a_{\tilde{p}|s},\omega\right)\right]\right]=E_{s\sim\tilde{m}}\left[E_{\omega\sim\tilde{p}|s}\left[u^*\left(a_{\beta|s},\omega\right)\right]\right]=E_{s\sim\tilde{m}}\left[E_{\omega\sim\tilde{p}|s}\left[u\left(a_{\beta|s},\omega\right)\right]\right]=$$

$$=\mathcal{W}\left(p,\{\tilde{p}|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)$$

Therefore, if $\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}}\succeq\{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}$ according to the Blackwell order and if $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq 0$ and $\mathcal{V}\left(p,\{\tilde{p}|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq 0$ in all decision problems $\mathcal{D}\left(\mathcal{A},u\right)$, then $\mathcal{W}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq\mathcal{W}\left(p,\{\tilde{p}|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)$ in all decision problems $\mathcal{D}\left(\mathcal{A},u\right)$.

Showing that if $\{m|\omega\}_{\omega\in\{\omega_1,\omega_2\}}\succeq\{\tilde{m}|\omega\}_{\omega\in\{\omega_1,\omega_2\}}$ according to the Blackwell order and if $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)\leq 0$ and $\mathcal{V}\left(p,\{\tilde{p}|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)\leq 0$ in all decision problems $\mathcal{D}\left(\mathcal{A},u\right)$, then $\mathcal{W}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)\leq\mathcal{W}\left(p,\{\tilde{p}|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)$ in all decision problems $\mathcal{D}\left(\mathcal{A},u\right)$ follows a similar line of reasoning. □

Proof of Proposition 4

I will first prove the "only if" direction, namely if $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},p,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq 0$ in all decision problems $\mathcal{D}\left(\mathcal{A},u\right)$, then conditions (i), (ii), and (iii) from the proposition are satisfied.

According to Theorem 3.1 in Morris and Shin (1997), the statement $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},p,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq 0$ in all decision problems $\mathcal{D}\left(\mathcal{A},u\right)$ is equivalent to the statement there exist $\psi:\mathcal{S}\to\mathbb{R}_0^+$ and $\phi:\mathcal{S}^2\to\mathbb{R}_0^+$ such that, $\forall\omega\in\Omega$ and $s_i\in\mathcal{S}$,

$$p\left(\omega\right)m|\omega\left(s_i\right)=\psi\left(s_i\right)\beta|s_i\left(\omega\right)+\sum_{j\neq i}\left[\phi\left(s_i,s_j\right)\beta|s_i\left(\omega\right)-\phi\left(s_j,s_i\right)\beta|s_j\left(\omega\right)\right]\tag{1}$$

Therefore, it is sufficient to show that when the latter statement holds, (i), (ii), and (iii) above also hold.

I will prove this in six steps.

Step 1: I claim that $\lambda:\mathcal{S}\to\mathbb{R}_0^+$ defined as $\lambda\left(s\right)=\psi\left(s\right)\ \forall s\in\mathcal{S}$ satisfies condition (i). Summing Equation 1 over signals, we obtain

$$\sum_i p\left(\omega\right)m|\omega\left(s_i\right)=\sum_i\left\{\psi\left(s_i\right)\beta|s_i\left(\omega\right)+\sum_{j\neq i}\left[\phi\left(s_i,s_j\right)\beta|s_i\left(\omega\right)-\phi\left(s_j,s_i\right)\beta|s_j\left(\omega\right)\right]\right\}$$

$$p\left(\omega\right)=\sum_i\psi\left(s_i\right)\beta|s_i\left(\omega\right)+\sum_i\sum_{j\neq i}\left[\phi\left(s_i,s_j\right)\beta|s_i\left(\omega\right)-\phi\left(s_j,s_i\right)\beta|s_j\left(\omega\right)\right]$$

$$p(\omega) = \sum_i \psi(s_i)\, \beta|s_i(\omega)$$

where all of the expressions containing $\phi(\cdot,\cdot)$ cancel out. Therefore, $\lambda : \mathcal{S} \to \mathbb{R}_0^+$ defined as $\lambda(s) = \psi(s)\ \forall s \in \mathcal{S}$ satisfies condition (i). For the remaining five steps, I replace $\psi(\cdot)$ in Equation 1 with $\lambda(\cdot)$.

Step 2: I claim that $\forall i \in \{1,...,k\}$, $m(s_i) - \lambda(s_i) = \sum_{i \neq j} [\phi(s_i, s_j) - \phi(s_j, s_i)]$. Consider once again Equation 1 and sum across states of the world

$$\sum_{\omega \in \Omega} p|s_i(\omega)\, m(s_i) = \sum_{\omega \in \Omega} \left\{ \lambda(s_i)\, \beta|s_i(\omega) + \sum_{j \neq i} [\phi(s_i, s_j)\, \beta|s_i(\omega) - \phi(s_j, s_i)\, \beta|s_j(\omega)] \right\}$$

$$m(s_i) = \sum_{\omega \in \Omega} \lambda(s_i)\, \beta|s_i(\omega) + \sum_{j \neq i} \left[\sum_{\omega \in \Omega} \phi(s_i, s_j)\, \beta|s_i(\omega) - \sum_{\omega \in \Omega} \phi(s_j, s_i)\, \beta|s_j(\omega) \right]$$

$$m(s_i) - \lambda(s_i) = \sum_{j \neq i} [\phi(s_i, s_j) - \phi(s_j, s_i)]$$

Step 3: I claim that $\forall s_i \in \mathcal{S}$, $\lambda(s_i) + \sum_{j \neq i} \phi(s_i, s_j) \neq 0$. Suppose, aiming towards contradiction, that $\lambda(s_i) + \sum_{j \neq i} \phi(s_i, s_j) = 0$. In that case, the expression from Step 2 would imply

$$m(s_i) + \sum_{j \neq i} \phi(s_j, s_i) = 0$$

but that's a contradiction, because we assumed $m(s_i) > 0\ \forall s_i \in \mathcal{S}$ and because $\phi(\cdot,\cdot) \geq 0$.

Step 4: I claim that, for arbitrary signal $s_i \in \mathcal{S}$, the vector $\left(\alpha_1^{s_i}, ..., \alpha_k^{s_i}\right) \in \mathbb{R}^k$ defined as $\alpha_i^{s_i} = \frac{m(s_i)}{\lambda(s_i) + \sum_{h \neq i} \phi(s_i, s_h)}$ and, $\forall j \neq i$, $\alpha_j^{s_i} = \frac{\phi(s_j, s_i)}{\lambda(s_i) + \sum_{h \neq i} \phi(s_i, s_h)}$ belongs to $\Delta(\mathcal{S})$. We already know that the denominator is non-zero; therefore, the expressions in the previous sentence are well-defined. We also know that $\alpha_i^{s_i} > 0$ and $\forall j \neq i$, $\alpha_j^{s_i} = \frac{\phi(s_j, s_i)}{\lambda(s_i) + \sum_{h \neq i} \phi(s_i, s_h)} \geq 0$ by the properties of $m(\cdot)$, $\lambda(\cdot)$, and $\phi(\cdot,\cdot)$. Therefore, in order to show that $\left(\alpha_1^{s_i}, ..., \alpha_k^{s_i}\right) \in \Delta(\mathcal{S})$, it is sufficient to show that $\alpha_i^{s_i} + \sum_{j \neq i} \alpha_j^{s_i} = 1$

$$\frac{m(s_i)}{\lambda(s_i) + \sum_{h \neq i} \phi(s_i, s_h)} + \sum_{j \neq i} \frac{\phi(s_j, s_i)}{\lambda(s_i) + \sum_{h \neq i} \phi(s_i, s_h)} = 1$$

The equation above can be re-written as

$$m(s_i) + \sum_{j \neq i} \phi(s_j, s_i) = \lambda(s_i) + \sum_{h \neq i} \phi(s_i, s_h)$$

which we know to be true by Step 2.

Step 5: I claim that, for each signal $s_i \in \mathcal{S}$, Equation 1 implies that there exists $\alpha^{s_i} \in \Delta(\mathcal{S})$

such that $\forall\omega\in\Omega\ \beta|s_i(\omega)=\sum_{j\neq i}\alpha_j^{s_i}\beta|s_j(\omega)+\alpha_i^{s_i}p|s_i(\omega)$. Specifically, one can rewrite Equation 1 as follows:

$$\beta|s_i(\omega)=\frac{m(s_i)}{\lambda(s_i)+\sum_{h\neq i}\phi(s_i,s_h)}p|s_i(\omega)+\frac{\sum_{j\neq i}\phi(s_j,s_i)}{\lambda(s_i)+\sum_{h\neq i}\phi(s_i,s_h)}\beta|s_j(\omega)=$$

$$=\alpha_i^{s_i}p|s_i(\omega)+\sum_{j\neq i}\alpha_j^{s_i}\beta|s_j(\omega)$$

where Step 4 already showed that $\left(\alpha_1^{s_i},...,\alpha_k^{s_i}\right)$ defined as above belong to $\Delta(\mathcal{S})$.

Step 6: I claim that the $\{\alpha^{s_j}\}_{j\in\{1,...,k\}}$ defined in the previous step satisfy

$$\frac{m(s_i)}{\alpha_i^{s_i}}=\lambda(s_i)+\sum_{j\neq i}\alpha_i^{s_j}\frac{m(s_j)}{\alpha_j^{s_j}}$$

First of all, notice that the expression is well defined because $\alpha_h^{s_h}>0\ \forall h\in\{1,...,k\}$. Subbing in $\alpha_i^{s_i}=\frac{m(s_i)}{\lambda(s_i)+\sum_{h\neq i}\phi(s_i,s_h)}$, $\alpha_j^{s_j}=\frac{m(s_j)}{\lambda(s_j)+\sum_{h\neq j}\phi(s_j,s_h)}$and, $\forall j\neq i$, $\alpha_j^{s_i}=\frac{\phi(s_j,s_i)}{\lambda(s_i)+\sum_{h\neq i}\phi(s_i,s_h)}$ we obtain

$$\frac{m(s_i)}{\frac{m(s_i)}{\lambda(s_i)+\sum_{h\neq i}\phi(s_i,s_h)}}=\lambda(s_i)+\sum_{j\neq i}\frac{\phi(s_i,s_j)}{\lambda(s_j)+\sum_{h\neq j}\phi(s_j,s_h)}\frac{m(s_j)}{\frac{m(s_j)}{\lambda(s_j)+\sum_{h\neq j}\phi(s_j,s_h)}}$$

$$\lambda(s_i)+\sum_{j\neq i}\phi(s_i,s_j)=\lambda(s_i)+\sum_{j\neq i}\phi(s_i,s_j)$$

which is true.

Therefore, if $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},p,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq 0$ in all decision problems $\mathcal{D}(\mathcal{A},u)$, then conditions (i), (ii), and (iii) from the statement of the proposition are satisfied.

I will now prove the "if" direction, namely if conditions (i), (ii), and (iii) from the statement of the proposition are satisfied, then $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},p,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq 0$ in all decision problems $\mathcal{D}(\mathcal{A},u)$.

In light of Theorem 3.1 in Morris and Shin (1997), it is sufficient to show that (i), (ii), and (iii) imply the existence of $\psi:\mathcal{S}\to\mathbb{R}_0^+$ and $\phi:\mathcal{S}^2\to\mathbb{R}_0^+$ such that, $\forall\omega\in\Omega$ and $s\in\mathcal{S}$, Equation 1 is satisfied.

In particular, we can manipulate condition (iii) as follows

$$\alpha_i^{s_i}=\frac{m(s_i)}{\lambda(s_i)+\sum_{j\neq i}\alpha_i^{s_j}\frac{m(s_j)}{\alpha_j^{s_j}}}$$

$$\frac{m(s_i)}{\alpha_i^{s_i}}\beta|s_i(\omega)-p|s_i(\omega)\,m(s_i)=\left[\lambda(s_i)+\sum_{j\neq i}\alpha_i^{s_j}\frac{m(s_j)}{\alpha_j^{s_j}}\right]\beta|s_i(\omega)-p|s_i(\omega)\,m(s_i)$$

$$\frac{m(s_i)}{\alpha_i^{s_i}}\left[\beta|s_i(\omega)-\alpha_i^{s_i}p|s_i(\omega)\right]=\left[\lambda(s_i)+\sum_{j\neq i}\alpha_i^{s_j}\frac{m(s_j)}{\alpha_j^{s_j}}\right]\beta|s_i(\omega)-p|s_i(\omega)\,m(s_i)$$

$$\frac{m(s_i)}{\alpha_i^{s_i}}\sum_{j\neq i}\alpha_j^{s_i}\beta|s_j(\omega)=\lambda(s_i)\,\beta|s_i(\omega)+\sum_{j\neq i}\alpha_i^{s_j}\frac{m(s_j)}{\alpha_j^{s_j}}\beta|s_i(\omega)-p|s_i(\omega)\,m(s_i)$$

$$p|s_i(\omega)\,m(s_i)=\lambda(s_i)\,\beta|s_i(\omega)+\sum_{j\neq i}\alpha_i^{s_j}\frac{m(s_j)}{\alpha_j^{s_j}}\beta|s_i(\omega)-\frac{m(s_i)}{\alpha_i^{s_i}}\sum_{j\neq i}\alpha_j^{s_i}\beta|s_j(\omega)$$

$$m|\omega(s_i)\,p(\omega)=\lambda(s_i)\,\beta|s_i(\omega)+\sum_{j\neq i}\left[\alpha_i^{s_j}\frac{m(s_j)}{\alpha_j^{s_j}}\beta|s_i(\omega)-\alpha_j^{s_i}\frac{m(s_i)}{\alpha_i^{s_i}}\beta|s_j(\omega)\right]$$

where the third step follows from the fact that $\forall\omega\in\Omega\ \beta|s_i(\omega)=\sum_{j\neq i}\alpha_j^{s_i}\beta|s_j(\omega)+\alpha_i^{s_i}p|s_i(\omega)$.

Letting $\psi:\mathcal{S}\to\mathbb{R}_0^+$ be defined as $\psi(s_i)=\lambda(s_i)\geq 0$ and $\phi:\mathcal{S}^2\to\mathbb{R}_0^+$ be defined as $\phi(s_j,s_i)=\alpha_j^{s_i}\frac{m(s_i)}{\alpha_i^{s_i}}$, we obtain Equation 1.

Therefore, if conditions (i), (ii), and (iii) from the statement of the proposition are satisfied, then $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},p,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq 0$ in all decision problems $\mathcal{D}(\mathcal{A},u)$. □

Proof of Proposition 5

In order to prove the first statement in the proposition, consider arbitrary decision problem $\mathcal{D}(\mathcal{A},u)$. As shown in the proof of Proposition 2, $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)$ can be rewritten as $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)=E_{s\sim m}\left[d(p|s,\beta)-d(p|s,\beta|s)\right]$. Clearly, if $\forall s\in\mathcal{S}\ d(p|s,\beta)-d(p|s,\beta|s)\geq 0$, then $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq 0$. Therefore, in order to show that $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq 0$ in arbitrary decision problem $\mathcal{D}(\mathcal{A},u)$, it is sufficient to show that $d(p|s,\beta)-d(p|s,\beta|s)\geq 0\ \forall s\in\mathcal{S}$.

Fix arbitrary $s\in\mathcal{S}$. Since we assumed that, for every $s\in\mathcal{S}$, there exists $\alpha_s\in[0,1]$ such that, for every $\omega\in\Omega$, $\beta|s(\omega)=\alpha_s\beta(\omega)+(1-\alpha_s)\,p|s(\omega)$, we know $\beta|s$, β, and $p|s$ are on the same line when thought of as vectors in $\Delta(\Omega)$. Consider, $G:\Delta(\Omega)\to\mathbb{R}\cup\{-\infty\}$ with $G(q)=E_{\omega\sim q}\left[u(a_q,\omega)\right]$. It is easy to show that $G(q)$ is a convex function (Dawid and Musio, 2014). Therefore, it is still a convex function if we limit the domain to $Q:\{q\in\Delta(\Omega)\,|q=\chi\beta+(1-\chi)\,p|s\ \textit{for some}\ \chi\in[0,1]\}$. Since $G(q)$ on the restricted domain Q is still convex, we know $d(p|s,q)$ increases weakly the further q is from $p|s$ in the restricted domain; i.e., we know that, in the restricted domain, $|p|s-q|\geq|p|s-q'|$ implies $d(p|s,q)\geq d(p|s,q')$. Since $\beta|s=\alpha_s\beta+(1-\alpha_s)\,p|s$ for some $\alpha_s\in[0,1]$, we know that, in the restricted domain, $|p|s-\beta|\geq|p|s-\beta|s|$ and, therefore, $d(p|s,\beta)\geq d(p|s,\beta|s)$. Since $s\in\mathcal{S}$ and $\mathcal{D}(\mathcal{A},u)$ were chosen arbitrarily, the argument above implies that $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)\geq 0$ in all decision problems $\mathcal{D}(\mathcal{A},u)$.

The second statement follows, mutatis mutandi, an analogous line of reasoning. □

Proof of Proposition 7

Suppose $\overline{\mathcal{V}}(p,p)$ and $d(p|s,p)$ are jointly valid measures of uncertainty and information in Frankel and Kamenica (2019). Then, there exists a decision problem $\mathcal{D}^*(\mathcal{A},u)$ satisfying the property $u(a_{\delta_\omega},\omega)=0\ \forall\omega\in\Omega$ in which, if the agent is an unbiased decoder, the measure of uncertainty is $\overline{\mathcal{V}}(p,p)$ and the ex-post measure of information is $d(p|s,p)$. As shown in Section 3, for a biased decoder with prior beliefs $\beta\in\Delta(\Omega)$ and posterior beliefs $\{\beta|s\}_{s\in\mathcal{S}}\in\Delta(\Omega)^{|\mathcal{S}|}$, one can construct $\overline{\mathcal{V}}(p,\beta)$ and $d(p|s,\beta)-d(p|s,\beta|s)$ from decision problem $\mathcal{D}^*(\mathcal{A},u)$. Therefore, $\overline{\mathcal{V}}(p,\beta)$ and $d(p|s,\beta)-d(p|s,\beta|s)$ respectively extend jointly valid measures of uncertainty $\overline{\mathcal{V}}(p,p)$ and $d(p|s,p)$ to the case of biased decoding.

Now consider $\mathcal{D}^*(\mathcal{A},u)$ satisfying the property $u(a_{\delta_\omega},\omega)=0\,\forall\omega\in\Omega$ and construct $\overline{\mathcal{V}}(p,\beta)$ and $d(p|s,\beta)-d(p|s,\beta|s)$. In the case of unbiased decoding, $\overline{\mathcal{V}}(p,\beta)$ and $d(p|s,\beta)-d(p|s,\beta|s)$ reduce to $\overline{\mathcal{V}}(p,p)$ and $d(p|s,p)$. But then, $\overline{\mathcal{V}}(p,p)$ and $d(p|s,p)$ are jointly valid measures of uncertainty and information in Frankel and Kamenica (2019). □

Proof of Proposition 8

Suppose the deceptive message that the receiver receives in state of the world $\omega\in\Omega$ is message $s_d\in\mathcal{S}$. If message s_d is deceptive in state of the world $\omega\in\Omega$, then $\beta|s_d(\omega)<p(\omega)$ and $m|\omega(s_d)=1$. Suppose, aiming towards contradiction, that there does not exist a decision problem $\mathcal{D}^*(\mathcal{A},u)$ for which receiving deceptive message $s_d\in\mathcal{S}$ in state of the world $\omega\in\Omega$ leads to a strictly negative value of communication. In other words, suppose that receiving deceptive message $s_d\in\mathcal{S}$ in state of the world $\omega\in\Omega$ leads to a non-negative value of communication in all decision problems $\mathcal{D}(\mathcal{A},u)$. As shown in Proposition 4, this implies that, $\forall i\in\{1,...,k\}$, $\exists\alpha^{s_i}\in\Delta(\mathcal{S})$ such that $\beta|s_i(\omega)=\sum_{j\neq i}\alpha_j^{s_i}\beta|s_j(\omega)+\alpha_i^{s_i}p|s_i(\omega)$ with $\alpha_i^{s_i}>0$. Since the agent is Bayesian and knows the prior distribution of states of the world, the existence of $\beta|s_d(\omega)<p(\omega)$ implies the existence of $s'\in\mathcal{S}$ such that $\beta|s'(\omega)>p(\omega)$. Consider $s_h\in\mathcal{S}$ such that $\beta|s_h(\omega)=\max\{\beta|s'(\omega)\}_{s'\in\mathcal{S}}$. Clearly $s_h\neq s_d$, because $\beta|s_h(\omega)>\beta|s_d(\omega)$. By Proposition 4, $\exists\alpha^{s_h}\in\Delta(\mathcal{S})$ such $\beta|s_h(\omega)=\sum_{j\neq h}\alpha_j^{s_h}\beta|s_j(\omega)+\alpha_h^{s_h}p|s_h(\omega)$ with $\alpha_h^{s_h}>0$. But since $m|\omega(s_d)=1$, $m|\omega(s_h)=0$, which implies $p|s_h(\omega)=0$. The fact that $\beta|s_h(\omega)=\max\{\beta|s'(\omega)\}_{s'\in\mathcal{S}}$ and the fact that $\beta|s_h(\omega)=\sum_{j\neq h}\alpha_j^{s_h}\beta|s_j(\omega)+\alpha_h^{s_h}\times 0$ with $\alpha_h^{s_h}>0$ yield the desired contradiction. □

Proof of Proposition 9

Let's first show that if message $s\in\mathcal{S}^*$ is a lie in state of the world $\omega\in\Omega$, then it is deceptive. Since $s\in\mathcal{S}^*$ is a lie in state of the world $\omega\in\Omega$, we know that $m|\omega(s)=1$ and $\ell|\omega(s)=0$. Since the receiver interprets messages according to the linguistic convention,

$$\beta|s(\omega)=\frac{\ell|\omega(s)\,p(\omega)}{\sum_{\omega'\in\Omega}\ell|\omega'(s)\,p(\omega')}=0$$

where we know the denominator of the expression is non-negative, because the existence of a conventional language $(\mathcal{S}^*, \ell)$ implies that $\forall s \in \mathcal{S}$, $\exists \omega' \in \Omega$ s.t. $\ell|\omega'(s) > 0$ and because we assumed p has full support. The full support of p also guarantees $0 = \beta|s(\omega) < p(\omega)$. Therefore, if message $s \in \mathcal{S}^*$ is a lie, then it is deceptive.

Let's now show that if message $s \in \mathcal{S}^*$ is deceptive in state of the world $\omega \in \Omega$, then it is a lie in state of the world $\omega \in \Omega$. If message $s \in \mathcal{S}^*$ is deceptive in state of the world $\omega \in \Omega$, then $\beta|s(\omega) < p(\omega)$ and and $m|\omega(s) = 1$. In the context of a conventional language $(\mathcal{S}^*, \ell)$, it is easy to see that $\forall s \in \mathcal{S}^*$ such that $\ell|\omega(s) > 0$, $\beta|s(\omega) \geq p(\omega)$. But then, $\beta|s(\omega) < p(\omega)$ must imply $\ell|\omega(s) = 0$. Therefore, if message $s \in \mathcal{S}^*$ is deceptive in state of the world $\omega \in \Omega$, then it is a lie in state of the world $\omega \in \Omega$. $\square$

Online Appendix: Not for Publication

Biased Decoding and the Foundations of Communication

Luca Braghieri

A Omitted Proofs

Proof of Proposition 2

I will divide this proof into two lemmata.

Lemma. $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq 0$ *in all decision problems* $\mathcal{D}(\mathcal{A}, u)$ *if and only if either the agent interprets the signal as noise or* $p|s_2(\omega_1) \leq \beta|s_2(\omega_1) \leq \beta|s_1(\omega_1) \leq p|s_1(\omega_1)$.

If the agent interprets the signal as noise, then $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = 0$ in all decision problems $\mathcal{D}(\mathcal{A}, u)$, because the agent takes the same action at β, $\beta|s_1$, and $\beta|s_2$.

Suppose the agent does not interpret the signal as noise. I will first show that if it is not the case that $p|s_2(\omega_1) \leq \beta|s_2(\omega_1) \leq \beta|s_1(\omega_1) \leq p|s_1(\omega_1)$, then there exists a decision problem $\mathcal{D}(\mathcal{A}, u)$ in which $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) < 0$.

There are two cases two consider: 1) $\beta|s_1(\omega_1), \beta|s_2(\omega_1) \in [p|s_2(\omega_1), p|s_1(\omega_1)]$ but $\beta|s_2(\omega_1) > \beta|s_1(\omega_1)$; 2) either $\beta|s_1(\omega_1) \notin [p|s_2(\omega_1), p|s_1(\omega_1)]$ or $\beta|s_2(\omega_1) \notin [p|s_2(\omega_1), p|s_1(\omega_1)]$.

Let's first consider case 1). Let $\mathcal{A} = [0,1]$ and define $u(a, \omega_1) = \ln(a)$ and $u(a, \omega_2) = \ln(1-a)$. Therefore, for $E_{\omega\sim q}[u(a,\omega)] = q(\omega_1)\ln(a) + (1-q(\omega_1))\ln(1-a)$. We know that $\arg\max_{a\in[0,1]} E_{\omega\sim q}[u(a,\omega)] = \arg\max_{a\in[0,1]}[q(\omega_1)\ln(a) + (1-q(\omega_1))\ln(1-a)] = \{q(\omega_1)\}$ (Aczél and Daróczy, 1975). Therefore, $\max_{a\in[0,1]} E_{\omega\sim q}[u(a,\omega)] = E_{\omega\sim q}[u(a_q,\omega)] = \sum_{i=1}^{2}[q(\omega_i)\ln(q(\omega_i))]$. Consider $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$:

$$\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{\beta|s},\omega\right)\right]\right] - E_{\omega\sim p}\left[u\left(a_\beta,\omega\right)\right] =$$

$$= E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{\beta|s},\omega\right)\right] - E_{\omega\sim p|s}\left[u\left(a_\beta,\omega\right)\right]\right] =$$

$$= E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{\beta|s},\omega\right)\right] - E_{\omega\sim p|s}\left[u\left(a_{p|s},\omega\right)\right] + E_{\omega\sim p|s}\left[u\left(a_{p|s},\omega\right)\right] - E_{\omega\sim p|s}\left[u\left(a_\beta,\omega\right)\right]\right] =$$

$$= E_{s\sim m}\left[d\left(p|s,\beta\right) - d\left(p|s,\beta|s\right)\right]$$

In order to show that $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) < 0$, it is sufficient to show that $\forall s_i \in \mathcal{S}$, $d\left(p|s,\beta\right) - d\left(p|s,\beta|s\right) < 0$. This is true for the following reason: first, since $\beta|s_2(\omega_1) > \beta|s_1(\omega_1)$, we know by assumption 1 that $\beta|s_2(\omega_1) > \beta(\omega_1) > \beta|s_1(\omega_1)$. Since we are in case 1), we have $p|s_2(\omega_1) \leq \beta|s_1(\omega_1) < \beta(\omega_1) < \beta|s_2(\omega_1) \leq p|s_1(\omega_1)$; therefore, $\forall s \in \mathcal{S}$ $|\beta|s(\omega_1) - p|s(\omega_1)| > |\beta(\omega_1) - p|s(\omega_1)|$. Second, $E_{\omega\sim q}[u(a_q,\omega)]$ is strictly convex, which implies that, in each direction, $d\left(p|s,q\right)$ strictly increases the further $q(\omega_1)$ is from $p|s(\omega_1)$. Therefore, $\forall s \in \mathcal{S}, d\left(p|s,\beta\right) - d\left(p|s,\beta|s\right) < 0$ and $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) < 0$.

For case 2), let $\mathcal{A} = [0,1]\cup\{a^*\}$. Define $u(a,\omega_1) = \ln(a)$ and $u(a,\omega_2) = \ln(1-a)$ for $a \in [0,1]$ and

$$u(a^*,\omega_1) = \frac{1 - \min\{\beta(\omega_1), p|s_2(\omega_1)\}}{\max\{\beta(\omega_1), p|s_1(\omega_1)\} - \min\{\beta(\omega_1), p|s_2(\omega_1)\}} \times$$

$$\left\{\sum_{i=1}^{2}\left[\max\left\{\beta\left(\omega_i\right),p|s_1\left(\omega_i\right)\right\}\ln\left(\max\left\{\beta\left(\omega_i\right),p|s_1\left(\omega_i\right)\right\}\right)\right]-\sum_{i=1}^{2}\left[\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\ln\left(\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\right)\right]\right\}+$$

$$+\sum_{i=1}^{2}\left[\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\ln\left(\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\right)\right]$$

$$u\left(a^*,\omega_2\right)=-\frac{\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\}}{\max\left\{\beta\left(\omega_1\right),p|s_1\left(\omega_1\right)\right\}-\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\}}\times$$

$$\left\{\sum_{i=1}^{2}\left[\max\left\{\beta\left(\omega_i\right),p|s_1\left(\omega_i\right)\right\}\ln\left(\max\left\{\beta\left(\omega_i\right),p|s_1\left(\omega_i\right)\right\}\right)\right]-\sum_{i=1}^{2}\left[\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\ln\left(\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\right)\right]\right\}+$$

$$+\sum_{i=1}^{2}\left[\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\ln\left(\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\right)\right]$$

Therefore, for $a\in[0,1]$, $E_{\omega\sim q}\left[u\left(a,\omega\right)\right]=q\left(\omega_1\right)\ln\left(a\right)+\left(1-q\left(\omega_1\right)\right)\ln\left(1-a\right)$. Conversely, for $a=a^*$,

$$E_{\omega\sim q}\left[u\left(a^*,\omega\right)\right]=\frac{q\left(\omega_1\right)-\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\}}{\max\left\{\beta\left(\omega_1\right),p|s_1\left(\omega_1\right)\right\}-\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\}}\times$$

$$\left\{\sum_{i=1}^{2}\left[\max\left\{\beta\left(\omega_i\right),p|s_1\left(\omega_i\right)\right\}\ln\left(\max\left\{\beta\left(\omega_i\right),p|s_1\left(\omega_i\right)\right\}\right)\right]-\sum_{i=1}^{2}\left[\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\ln\left(\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\right)\right]\right\}+$$

$$+\sum_{i=1}^{2}\left[\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\ln\left(\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\right)\right]$$

We know that $\arg\max_{a\in[0,1]}E_{\omega\sim q}\left[u\left(a,\omega\right)\right]=\arg\max_{a\in[0,1]}\left[q\left(\omega_1\right)\ln\left(a\right)+\left(1-q\left(\omega_1\right)\right)\ln\left(1-a\right)\right]=\left\{q\left(\omega_1\right)\right\}$ (Aczél and Daróczy, 1975). Therefore, $\max_{a\in[0,1]}E_{\omega\sim q}\left[u\left(a,\omega\right)\right]=\sum_{i=1}^{2}\left[q\left(\omega_i\right)\ln\left(q\left(\omega_i\right)\right)\right]$.

Notice that, at q such that $q\left(\omega_1\right)=\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\}$,

$$\max_{a\in[0,1]}E_{\omega\sim q}\left[u\left(a,\omega\right)\right]=\sum_{i=1}^{2}\left[\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\ln\left(\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\right)\right]$$

and

$$E_{\omega\sim q}\left[u\left(a^*,\omega\right)\right]=\sum_{i=1}^{2}\left[\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\ln\left(\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\right)\right]$$

Similarly, at q such that $q\left(\omega_1\right)=\max\left\{\beta\left(\omega_1\right),p|s_1\left(\omega_1\right)\right\}$,

$$\max_{a\in[0,1]}E_{\omega\sim q}\left[u\left(a,\omega\right)\right]=\sum_{i=1}^{2}\left[\max\left\{\beta\left(\omega_i\right),p|s_1\left(\omega_i\right)\right\}\ln\left(\max\left\{\beta\left(\omega_i\right),p|s_1\left(\omega_i\right)\right\}\right)\right]$$

and

$$E_{\omega\sim q}\left[u\left(a^*,\omega\right)\right]=\sum_{i=1}^{2}\left[\max\left\{\beta\left(\omega_i\right),p|s_1\left(\omega_i\right)\right\}\ln\left(\max\left\{\beta\left(\omega_i\right),p|s_1\left(\omega_i\right)\right\}\right)\right]$$

Since $\max_{a\in[0,1]}E_{\omega\sim q}\left[u\left(a,\omega\right)\right]$ is a strictly convex function and since we know that $E_{\omega\sim q}\left[u\left(a^*,\omega\right)\right]$ is

an affine function that intersects $\max_{a\in[0,1]} E_{\omega\sim q}\left[u\left(a,\omega\right)\right]$ at

$$\left(\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\},\sum_{i=1}^{2}\left[\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\ln\left(\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\right)\right]\right)$$

and at

$$\left(\max\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\},\sum_{i=1}^{2}\left[\max\left\{\beta\left(\omega_i\right),p|s_1\left(\omega_i\right)\right\}\ln\left(\max\left\{\beta\left(\omega_i\right),p|s_1\left(\omega_i\right)\right\}\right)\right]\right)$$

it must be the case that $E_{\omega\sim q}\left[u\left(a^*,\omega\right)\right]\geq\max_{a\in[0,1]} E_{\omega\sim q}\left[u\left(a,\omega\right)\right]$ if and only if

$$q\left(\omega_1\right)\in\left[\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\},\max\left\{\beta\left(\omega_1\right),p|s_1\left(\omega_1\right)\right\}\right]$$

But then,

$$\arg\max_{a\in[0,1]\cup\{a^*\}} E_{\omega\sim q}\left[u\left(a,\omega\right)\right]=\begin{cases}\{q\left(\omega_1\right)\} & for\ q\ s.t.\ q\left(\omega_1\right)\notin\left[\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\},\max\left\{\beta\left(\omega_1\right),p|s_1\left(\omega_1\right)\right\}\right]\\ \{a^*,\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\}\} & for\ q\ s.t.\ q\left(\omega_1\right)=\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\}\\ \{a^*,\max\left\{\beta\left(\omega_1\right),p|s_1\left(\omega_1\right)\right\}\} & for\ q\ s.t.\ q\left(\omega_1\right)=\max\left\{\beta\left(\omega_1\right),p|s_1\left(\omega_1\right)\right\}\\ \{a^*\} & for\ q\ s.t.\ q\left(\omega_1\right)\in\left(\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\},\max\left\{\beta\left(\omega_1\right),p|s_1\left(\omega_1\right)\right\}\right)\end{cases}$$

Furthermore,

$$\max_{a\in[0,1]\cup a^*} E_{\omega\sim q}\left[u\left(a,\omega\right)\right]=$$

$$\begin{cases}\sum_{i=1}^{2}\left[p\left(\omega_i\right)\ln\left(p\left(\omega_i\right)\right)\right] & for\ q\ s.t.\ q\left(\omega_1\right)\notin\left[\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\},\max\left\{\beta\left(\omega_1\right),p|s_1\left(\omega_1\right)\right\}\right]\\ \frac{q-\min\{\beta(\omega_1),p|s_2(\omega_1)\}}{\max\{\beta(\omega_1),p|s_1(\omega_1)\}-\min\{\beta(\omega_1),p|s_2(\omega_1)\}}\times & \\ \left\{\sum_{i=1}^{2}\left[\max\left\{\beta\left(\omega_i\right),p|s_1\left(\omega_i\right)\right\}\ln\left(\max\left\{\beta\left(\omega_i\right),p|s_1\left(\omega_i\right)\right\}\right)\right]-\right. & \\ \left.\sum_{i=1}^{2}\left[\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\ln\left(\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\right)\right]\right\}+ & for\ q\ s.t.\ q\left(\omega_1\right)\in\left[\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\},\max\left\{\beta\left(\omega_1\right),p|s_1\left(\omega_1\right)\right\}\right]\\ +\sum_{i=1}^{2}\left[\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\ln\left(\min\left\{\beta\left(\omega_i\right),p|s_2\left(\omega_i\right)\right\}\right)\right] & \end{cases}$$

Of course, $E_{\omega\sim q}\left[u\left(a_q,\omega\right)\right]=\max_{a\in[0,1]\cup a^*} E_{\omega\sim q}\left[u\left(a,\omega\right)\right]$ is convex. Furthermore, it is strictly convex for q such that $q\left(\omega_1\right)\notin\left[\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\},\max\left\{\beta\left(\omega_1\right),p|s_1\left(\omega_1\right)\right\}\right]$ and linear for q such that $q\left(\omega_1\right)\in\left[\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\},\max\left\{\beta\left(\omega_1\right),p|s_1\left(\omega_1\right)\right\}\right]$. Since $E_{\omega\sim q}\left[u\left(a_q,\omega\right)\right]$ is linear for q such that $q\left(\omega_1\right)\in\left[\min\left\{\beta\left(\omega_1\right),p|s_2\left(\omega_1\right)\right\},\max\left\{\beta\left(\omega_1\right),p|s_1\left(\omega_1\right)\right\}\right]$ the expected value of sample information to an unbiased decoder is zero. Furthermore, by construction, the agent takes the same action a^* at both β and p, which implies $d\left(p,\beta\right)=0$. Therefore, B-EVSI reduces to

$$\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)=-E_{s\sim m}\left[d\left(p|s,\beta|s\right)\right]\leq 0$$

Therefore, $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)<0$ if $\exists s_i\in\mathcal{S}$ such that $d\left(p|s,\beta|s\right)>0$.

Since we are in case 2), we know that either $\beta|s_1\left(\omega_1\right)\notin\left[p|s_2\left(\omega_1\right),p|s_1\left(\omega_1\right)\right]$ or $\beta|s_2\left(\omega_1\right)\notin\left[p|s_2\left(\omega_1\right),p|s_1\left(\omega_1\right)\right]$. We can consider the following mutually exclusive and exhaustive subcases.

Leveraging the fact that the agent does not interpret the signal as noise and Assumption 1, we see that, in each subcase, $\exists s_i \in \mathcal{S}$ s.t. $\beta|s_i(\omega_1) \notin [\min\{\beta(\omega_1), p|s_2(\omega_1)\}, \max\{\beta(\omega_1), p|s_1(\omega_1)\}]$ as pointed out below.

1. $\beta|s_1(\omega_1) < \beta|s_2(\omega_1) < p|s_2(\omega_1) < p|s_1(\omega_1)$. In this subcase, $\beta|s_1(\omega_1) < \min\{\beta(\omega_1), p|s_2(\omega_1)\}$.
2. $\beta|s_2(\omega_1) < \beta|s_1(\omega_1) < p|s_2(\omega_1) < p|s_1(\omega_1)$. In this subcase, $\beta|s_2(\omega_1) < \min\{\beta(\omega_1), p|s_2(\omega_1)\}$.
3. $\beta|s_1(\omega_1) < p|s_2(\omega_1) \leq \beta|s_2(\omega_1) \leq p|s_1(\omega_1)$. In this subcase, $\beta|s_1(\omega_1) < \min\{\beta(\omega_1), p|s_2(\omega_1)\}$.
4. $\beta|s_1(\omega_1) < p|s_2(\omega_1) < p|s_1(\omega_1) < \beta|s_2(\omega_1)$. In this subcase, $\beta|s_1(\omega_1) < \min\{\beta(\omega_1), p|s_2(\omega_1)\}$ and $\beta|s_2(\omega_1) > \max\{\beta(\omega_1), p|s_1(\omega_1)\}$.
5. $\beta|s_2(\omega_1) < p|s_2(\omega_1) \leq \beta|s_1(\omega_1) \leq p|s_1(\omega_1)$. In this subcase, $\beta|s_2(\omega_1) < \min\{\beta(\omega_1), p|s_2(\omega_1)\}$.
6. $p|s_2(\omega_1) \leq \beta|s_1(\omega_1) \leq p|s_1(\omega_1) < \beta|s_2(\omega_1)$. In this subcase, $\beta|s_2(\omega_1) > \max\{\beta(\omega_1), p|s_1(\omega_1)\}$.
7. $\beta|s_2(\omega_1) < p|s_2(\omega_1) < p|s_1(\omega_1) < \beta|s_1(\omega_1)$. In this subcase, $\beta|s_2(\omega_1) < \min\{\beta(\omega_1), p|s_2(\omega_1)\}$ and $\beta|s_1(\omega_1) > \max\{\beta(\omega_1), p|s_1(\omega_1)\}$.
8. $p|s_2(\omega_1) \leq \beta|s_2(\omega_1) \leq p|s_1(\omega_1) < \beta|s_1(\omega_1)$. In this subcase, $\beta|s_1(\omega_1) > \max\{\beta(\omega_1), p|s_1(\omega_1)\}$.
9. $p|s_2(\omega_1) < p|s_1(\omega_1) < \beta|s_1(\omega_1) < \beta|s_2(\omega_1)$. In this subcase, $\beta|s_2(\omega_1) > \max\{\beta(\omega_1), p|s_1(\omega_1)\}$.
10. $p|s_2(\omega_1) < p|s_1(\omega_1) < \beta|s_2(\omega_1) < \beta|s_1(\omega_1)$. In this subcase, $\beta|s_1(\omega_1) > \max\{\beta(\omega_1), p|s_1(\omega_1)\}$.

Since, in each of the subcases above $\exists s_i \in \mathcal{S}$ s.t.

$$\beta|s_i(\omega_1) \notin [\min\{\beta(\omega_1), p|s_2(\omega_1)\}, \max\{\beta(\omega_1), p|s_1(\omega_1)\}]$$

then, in each subcase, $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) < 0$. This is because, for

$$\beta|s_i(\omega_1) \notin [\min\{\beta(\omega_1), p|s_2(\omega_1)\}, \max\{\beta(\omega_1), p|s_1(\omega_1)\}]$$

$\arg\max_{a\in[0,1]\cup a^*} E_{\omega\sim\beta|s_i}[u(a,\omega)] = \left\{a_{\beta|s_i}\right\}$ as can be seen from the general expression for $\arg\max_{a\in[0,1]\cup\{a^*\}} E_{\omega\sim q}[u(a,\omega)]$ above. Furthermore, by the same expression, $a_{\beta|s_i} \notin \arg\max_{a\in[0,1]\cup a^*} E_{\omega\sim p|s_i}[u(a,\omega)]$. Therefore, in each case, $\exists s_i \in \mathcal{S}$ s.t. $d(p|s_i, \beta|s_i) = E_{\omega\sim p|s_i}[u(a^*,\omega)] - E_{\omega\sim p|s_i}\left[u\left(a_{\beta|s_i},\omega\right)\right] > 0$, which implies $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) < 0$.

Let's continue assuming that the agent does not interpret the signal as noise and let's now prove the sufficient conditions; i.e., if $p|s_2(\omega_1) \leq \beta|s_2(\omega_1) \leq \beta|s_1(\omega_1) \leq p|s_1(\omega_1)$ then $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq 0$ in all decision problems $\mathcal{D}(\mathcal{A}, u)$. Notice that the fact that the agent does not interpret the signal as noise, together with Assumption 1, implies $p|s_2(\omega_1) \leq \beta|s_2(\omega_1) < \beta(\omega_1) < \beta|s_1(\omega_1) \leq p|s_1(\omega_1)$.

As shown above, $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$ can be rewritten as $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = E_{s\sim m}\left[d\left(p|s, \beta\right) - d\left(p|s, \beta|s\right)\right]$. In order to show that $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \geq 0$, it is sufficient to show that $\forall s_i \in \mathcal{S}$, $d\left(p|s, \beta\right) - d\left(p|s, \beta|s\right) \geq 0$. This is true because $E_{\omega\sim q}\left[u\left(a_q, \omega\right)\right]$ is weakly convex, which implies that, in each direction, $d\left(p|s, q\right)$ weakly increases the further $q\left(\omega_1\right)$ is from $p|s\left(\omega_1\right)$, and because $\forall s_i \in \mathcal{S}$ $|\beta|s_i\left(\omega_1\right) - p|s_i\left(\omega_1\right)| < |\beta\left(\omega_1\right) - p|s_i\left(\omega_1\right)|$ by assumption. □

Lemma. *$\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \leq 0$ in all decision problems $\mathcal{D}\left(\mathcal{A}, u\right)$ if and only if either the agent interprets the signal as noise or $\beta\left(\omega_1\right) \in \left[p|s_2\left(\omega_1\right), p|s_1\left(\omega_1\right)\right]$ and $\beta|s_1\left(\omega_1\right) \leq \beta|s_2\left(\omega_1\right)$.*

If the agent interprets the signal as noise, then $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = 0$ in all decision problems $\mathcal{D}\left(\mathcal{A}, u\right)$, because the agent takes the same action at β, $\beta|s_1$, and $\beta|s_2$.

Suppose the agent does not interpret the signal as noise. I will first show that if it is not the case that $\beta\left(\omega_1\right) \in \left[p|s_2\left(\omega_1\right), p|s_1\left(\omega_1\right)\right]$ and $\beta|s_1\left(\omega_1\right) \leq \beta|s_2\left(\omega_1\right)$, then there exists a decision problem $\mathcal{D}\left(\mathcal{A}, u\right)$ in which $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) > 0$.

Since it is not the case that $\beta\left(\omega_1\right) \in \left[p|s_2\left(\omega_1\right), p|s_1\left(\omega_1\right)\right]$ and $\beta|s_1\left(\omega_1\right) \leq \beta|s_2\left(\omega_1\right)$, one of three things must be true: i) $\beta\left(\omega_1\right) \notin \left[p|s_2\left(\omega_1\right), p|s_1\left(\omega_1\right)\right]$ and $\beta|s_1\left(\omega_1\right) \geq \beta|s_2\left(\omega_1\right)$, ii) $\beta\left(\omega_1\right) \notin \left[p|s_2\left(\omega_1\right), p|s_1\left(\omega_1\right)\right]$ and $\beta|s_1\left(\omega_1\right) \leq \beta|s_2\left(\omega_1\right)$, iii) $\beta\left(\omega_1\right) \in \left[p|s_2\left(\omega_1\right), p|s_1\left(\omega_1\right)\right]$ and $\beta|s_1\left(\omega_1\right) > \beta|s_2\left(\omega_1\right)$.

Consider case i): $\beta\left(\omega_1\right) \notin \left[p|s_2\left(\omega_1\right), p|s_1\left(\omega_1\right)\right]$ and $\beta|s_1\left(\omega_1\right) \geq \beta|s_2\left(\omega_1\right)$. First, suppose $\beta\left(\omega_1\right) < p|s_2\left(\omega_1\right)$. By Assumption 1, we know $\beta|s_2\left(\omega_1\right) < \beta\left(\omega_1\right) < p|s_2\left(\omega_1\right)$. Construct a decision problem as follows: $\mathcal{A} = \{a_1, a_2\}$. $u\left(a_1, \omega_1\right) = -1$, $u\left(a_1, \omega_2\right) = 0$, $u\left(a_2, \omega_1\right) = u\left(a_2, \omega_2\right) = -\frac{\beta(\omega_1)+\min\{p|s_2(\omega_1), \beta|s_1(\omega_1)\}}{2}$. Then, the agent picks action a_2 whenever she believes state of the world ω_1 occurs with probability $q\left(\omega_1\right) \geq \frac{\beta(\omega_1)+\min\{p|s_2(\omega_1), \beta|s_1(\omega_1)\}}{2}$. But then, $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) > 0$, because

$$\begin{aligned}
&\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = \\
&= m\left(s_1\right)\left[-\frac{\beta\left(\omega_1\right) + \min\left\{p|s_2\left(\omega_1\right), \beta|s_1\left(\omega_1\right)\right\}}{2}\right] + m\left(s_2\right)\left[p|s_2\left(\omega_1\right)\left(-1\right) + p|s_2\left(\omega_2\right)\left(0\right)\right] + \\
&- \left\{m\left(s_1\right)\left[p|s_1\left(\omega_1\right)\left(-1\right) + p|s_1\left(\omega_2\right)\left(0\right)\right] + m\left(s_2\right)\left[p|s_2\left(\omega_1\right)\left(-1\right) + p|s_2\left(\omega_2\right)\left(0\right)\right]\right\} = \\
&= m\left(s_1\right)\left[-\frac{\beta\left(\omega_1\right) + \min\left\{p|s_2\left(\omega_1\right), \beta|s_1\left(\omega_1\right)\right\}}{2}\right] - m\left(s_1\right)\left[-p|s_1\left(\omega_1\right)\right] = \\
&= m\left(s_1\right)\left[-\frac{\beta\left(\omega_1\right) + \min\left\{p|s_2\left(\omega_1\right), \beta|s_1\left(\omega_1\right)\right\}}{2} + p|s_1\left(\omega_1\right)\right] > 0
\end{aligned}$$

which is true because $m\left(s_2\right) > 0$ and $p|s_1\left(\omega_1\right) > \frac{\beta(\omega_1)+\min\{p|s_2(\omega_1), \beta|s_1(\omega_1)\}}{2}$. Therefore, when $\beta\left(\omega_1\right) < p|s_2\left(\omega_1\right)$ and $\beta|s_1\left(\omega_1\right) \geq \beta|s_2\left(\omega_1\right)$, there exist decision problems $\mathcal{D}\left(\mathcal{A}, u\right)$ such that $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) > 0$. A similar line of reasoning shows that, when $\beta\left(\omega_1\right) > p|s_1\left(\omega_1\right)$ and $\beta|s_1\left(\omega_1\right) \geq \beta|s_2\left(\omega_1\right)$, there also exist decision problems in which $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) > 0$.

Let's now consider case ii): $\beta\left(\omega_1\right) \notin \left[p|s_2\left(\omega_1\right), p|s_1\left(\omega_1\right)\right]$ and $\beta|s_1\left(\omega_1\right) \leq \beta|s_2\left(\omega_1\right)$. First, suppose $\beta\left(\omega_1\right) < p|s_2\left(\omega_1\right)$. By Assumption 1, we know $\beta|s_1\left(\omega_1\right) < \beta\left(\omega_1\right) < p|s_2\left(\omega_1\right)$. Construct a decision problem as follows: $\mathcal{A} = \{a_1, a_2\}$. $u\left(a_1, \omega_1\right) = -1$, $u\left(a_1, \omega_2\right) = 0$, $u\left(a_2, \omega_1\right) = u\left(a_2, \omega_2\right) = -\frac{\beta(\omega_1)+\min\{p|s_2(\omega_1), \beta|s_2(\omega_1)\}}{2}$. Then, the agent picks action a_2 whenever she believes state of the world ω_1 occurs with probability $q\left(\omega_1\right) \geq \frac{\beta(\omega_1)+\min\{p|s_2(\omega_1), \beta|s_2(\omega_1)\}}{2}$. But then, $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) >$

0, because

$$\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)=$$
$$=m\left(s_1\right)\left[p|s_1\left(\omega_1\right)\left(-1\right)+p|s_1\left(\omega_2\right)\left(0\right)\right]+m\left(s_2\right)\left[-\frac{\beta\left(\omega_1\right)+\min\left\{p|s_2\left(\omega_1\right),\beta|s_2\left(\omega_1\right)\right\}}{2}\right]+$$
$$-\left\{m\left(s_1\right)\left[p|s_1\left(\omega_1\right)\left(-1\right)+p|s_1\left(\omega_2\right)\left(0\right)\right]+m\left(s_2\right)\left[p|s_2\left(\omega_1\right)\left(-1\right)+p|s_2\left(\omega_2\right)\left(0\right)\right]\right\}=$$
$$=m\left(s_2\right)\left[-\frac{\beta\left(\omega_1\right)+\min\left\{p|s_2\left(\omega_1\right),\beta|s_2\left(\omega_1\right)\right\}}{2}\right]-m\left(s_2\right)\left[-p|s_2\left(\omega_1\right)\right]=$$
$$=m\left(s_2\right)\left[-\frac{\beta\left(\omega_1\right)+\min\left\{p|s_2\left(\omega_1\right),\beta|s_2\left(\omega_1\right)\right\}}{2}+p|s_2\left(\omega_1\right)\right]>0$$

which is true because $m\left(s_2\right)>0$ and $p|s_2\left(\omega_1\right)>-\frac{\beta(\omega_1)+\min\{p|s_2(\omega_1),\beta|s_2(\omega_1)\}}{2}$. Therefore, when $\beta\left(\omega_1\right)<p|s_2\left(\omega_1\right)$ and $\beta|s_1\left(\omega_1\right)\leq\beta|s_2\left(\omega_1\right)$, there exist decision problems $\mathcal{D}\left(\mathcal{A},u\right)$ such that $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)>0$. A similar line of reasoning shows that, when $\beta\left(\omega_1\right)>p|s_1\left(\omega_1\right)$ and $\beta|s_1\left(\omega_1\right)\leq\beta|s_2\left(\omega_1\right)$, there also exist decision problems $\mathcal{D}\left(\mathcal{A},u\right)$ such that $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)>0$.

Let's now consider case iii): $\beta\left(\omega_1\right)\in\left[p|s_2\left(\omega_1\right),p|s_1\left(\omega_1\right)\right]$ and $\beta|s_1\left(\omega_1\right)>\beta|s_2\left(\omega_1\right)$. We will consider four subcases: a) $p|s_2\left(\omega_1\right)\leq\beta|s_2\left(\omega_1\right)<\beta\left(\omega_1\right)<\beta|s_1\left(\omega_1\right)\leq p|s_1\left(\omega_1\right)$, b) $\beta|s_2\left(\omega_1\right)<p|s_2\left(\omega_1\right)\leq\beta\left(\omega_1\right)<\beta|s_1\left(\omega_1\right)\leq p|s_1\left(\omega_1\right)$, c) $p|s_2\left(\omega_1\right)\leq\beta|s_2\left(\omega_1\right)\leq\beta\left(\omega_1\right)\leq p|s_1\left(\omega_1\right)<\beta|s_1\left(\omega_1\right)$, and d) $\beta|s_2\left(\omega_1\right)<p|s_2\left(\omega_1\right)\leq\beta\left(\omega_1\right)\leq p|s_1\left(\omega_1\right)<\beta|s_1\left(\omega_1\right)$.

Consider subcase a): $p|s_2\left(\omega_1\right)\leq\beta|s_2\left(\omega_1\right)<\beta\left(\omega_1\right)<\beta|s_1\left(\omega_1\right)\leq p|s_1\left(\omega_1\right)$. We will consider three sub-subcases: $p\left(\omega_1\right)\in\left[p|s_2\left(\omega_1\right),\beta|s_2\left(\omega_1\right)\right]$, $p\left(\omega_1\right)\in\left[\beta|s_1\left(\omega_1\right),p|s_1\left(\omega_1\right)\right]$, $p\left(\omega_1\right)\in\left[\beta|s_2\left(\omega_1\right),\beta|s_1\left(\omega_1\right)\right]$. If $p\left(\omega_1\right)\in\left[p|s_2\left(\omega_1\right),\beta|s_2\left(\omega_1\right)\right]$, let $\mathcal{A}=\{a_1,a_2\}$, $u\left(a_1,\omega_1\right)=0$, $u\left(a_1,\omega_2\right)=-1$, $u\left(a_2,\omega_1\right)=u\left(a_2,\omega_2\right)=-1+\frac{1}{2}\left[\beta\left(\omega_1\right)+\beta|s_1\left(\omega_1\right)\right]$. Then, the agent picks action a_1 whenever she believes state of the world ω_1 occurs with probability $q\left(\omega_1\right)\geq\frac{1}{2}\left[\beta\left(\omega_1\right)+\beta|s_1\left(\omega_1\right)\right]$. But then, $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)>0$, because

$$\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)=$$
$$=m\left(s_1\right)\left[p|s_1\left(\omega_1\right)\left(0\right)+p|s_1\left(\omega_2\right)\left(-1\right)\right]+m\left(s_2\right)\left[-1+\frac{1}{2}\left[\beta\left(\omega_1\right)+\beta|s_1\left(\omega_1\right)\right]\right]+$$
$$-\left\{m\left(s_1\right)\left[-1+\frac{1}{2}\left[\beta\left(\omega_1\right)+\beta|s_1\left(\omega_1\right)\right]\right]+m\left(s_2\right)\left[-1+\frac{1}{2}\left[\beta\left(\omega_1\right)+\beta|s_1\left(\omega_1\right)\right]\right]\right\}=$$
$$=m\left(s_1\right)\left[p|s_1\left(\omega_1\right)\left(0\right)+p|s_1\left(\omega_2\right)\left(-1\right)\right]-m\left(s_1\right)\left[-1+\frac{1}{2}\left[\beta\left(\omega_1\right)+\beta|s_1\left(\omega_1\right)\right]\right]=$$
$$=m\left(s_1\right)\left\{-p|s_1\left(\omega_2\right)+1-\frac{1}{2}\left[\beta\left(\omega_1\right)+\beta|s_1\left(\omega_1\right)\right]\right\}=m\left(s_1\right)\left[p|s_1\left(\omega_1\right)-\frac{1}{2}\left[\beta\left(\omega_1\right)+\beta|s_1\left(\omega_1\right)\right]\right]>0$$

If $p\left(\omega_1\right)\in\left[\beta|s_1\left(\omega_1\right),p|s_1\left(\omega_1\right)\right]$, a symmetric argument shows that there exists a decision problem $\mathcal{D}\left(\mathcal{A},u\right)$ for which $\mathcal{V}\left(p,\{p|s\}_{s\in\mathcal{S}},\beta,\{\beta|s\}_{s\in\mathcal{S}}\right)>0$. Lastly, if $p\left(\omega_1\right)\in\left[\beta|s_2\left(\omega_1\right),\beta|s_1\left(\omega_1\right)\right]$, let there be three actions $\mathcal{A}=\{a_1,a_2,a_3\}$. Let $u\left(a_1,\omega_1\right)=-\frac{1}{\min\{\beta(\omega_1),p(\omega_1)\}}$, $u\left(a_1,\omega_2\right)=0$, $u\left(a_2,\omega_1\right)=0$, $u\left(a_2,\omega_2\right)=-\frac{1}{1-\max\{\beta(\omega_1),p(\omega_1)\}}$, $u\left(a_3,\omega_1\right)=u\left(a_3,\omega_2\right)=-1$. Therefore, the agent picks action a_1 whenever she believes state of the world ω_1 occurs with probability $q\left(\omega_1\right)\leq\min\left\{\beta\left(\omega_1\right),p\left(\omega_1\right)\right\}$, action a_2 whenever she believes state of the world ω_1 occurs with probability

$q(\omega_1) \geq \max\{\beta(\omega_1), p(\omega_1)\}$ and action a_3 otherwise. But then,

$$\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) =$$

$$= m(s_1)\left[-\frac{1}{1-\max\{\beta(\omega_1), p(\omega_1)\}} p|s_1(\omega_2)\right] + m(s_2)\left[-\frac{1}{\min\{\beta(\omega_1), p(\omega_1)\}} p|s_2(\omega_1)\right] - (-1) =$$

$$= m(s_1)\left[1 - \frac{1}{1-\max\{\beta(\omega_1), p(\omega_1)\}} p|s_1(\omega_2)\right] + m(s_2)\left[1 - \frac{1}{\min\{\beta(\omega_1), p(\omega_1)\}} p|s_2(\omega_1)\right] > 0$$

which is true, because $1 - \frac{1}{1-\max\{\beta(\omega_1),p(\omega_1)\}} p|s_1(\omega_2) > 0$, $1 - \frac{1}{\min\{\beta(\omega_1),p(\omega_1)\}} p|s_2(\omega_1) > 0$.

Consider subcase b): $\beta|s_2(\omega_1) < p|s_2(\omega_1) \leq \beta(\omega_1) < \beta|s_1(\omega_1) \leq p|s_1(\omega_1)$. Let there be two actions: $\mathcal{A} = \{a_1, a_2\}$. Let $\mathcal{A} = \{a_1, a_2\}$, $u(a_1, \omega_1) = u(a_1, \omega_2) = -1 + \frac{1}{2}[\beta(\omega_1) + \beta|s_1(\omega_1)]$, $u(a_2, \omega_1) = 0$, $u(a_2, \omega_2) = -1$. Therefore, the agent picks action a_2 whenever she believes state of the world ω_1 occurs with probability $q(\omega_1) \geq \frac{1}{2}[\beta(\omega_1) + \beta|s_1(\omega_1)]$. But then,

$$\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) =$$

$$= m(s_1)\left[p|s_1(\omega_1)(0) + p|s_1(\omega_2)(-1)\right] + m(s_2)\left[-1 + \frac{1}{2}[\beta(\omega_1) + \beta|s_1(\omega_1)]\right] - \left\{-1 + \frac{1}{2}[\beta(\omega_1) + \beta|s_1(\omega_1)]\right\} =$$

$$= m(s_1)\left[-p|s_1(\omega_2) + 1 - \frac{1}{2}[\beta(\omega_1) + \beta|s_1(\omega_1)]\right] > 0$$

which is true, because $m(s_1) > 0$ and $p|s_1(\omega_1) > \frac{1}{2}[\beta(\omega_1) + \beta|s_1(\omega_1)]$.

Consider subcase c): $p|s_2(\omega_1) \leq \beta|s_2(\omega_1) \leq \beta(\omega_1) \leq p|s_1(\omega_1) < \beta|s_1(\omega_1)$. A similar line of reasoning as in the previous subcase shows that there exists a decision problems $\mathcal{D}(\mathcal{A}, u)$ such that $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) > 0$.

Lastly, consider subcase d): $\beta|s_2(\omega_1) < p|s_2(\omega_1) \leq \beta(\omega_1) \leq p|s_1(\omega_1) < \beta|s_1(\omega_1)$. Let there be three actions $\mathcal{A} = \{a_1, a_2, a_3\}$. Let $u(a_1, \omega_1) = -\frac{1}{\min\{\beta(\omega_1),p(\omega_1)\}}$, $u(a_1, \omega_2) = 0$, $u(a_2, \omega_1) = 0$, $u(a_2, \omega_2) = -\frac{1}{1-\max\{\beta(\omega_1),p(\omega_1)\}}$, $u(a_3, \omega_1) = u(a_3, \omega_2) = -1$. Therefore, the agent picks action a_1 whenever she believes state of the world ω_1 occurs with probability $q(\omega_1) \leq \min\{\beta(\omega_1), p(\omega_1)\}$, action a_2 whenever she believes state of the world ω_1 occurs with probability $q(\omega_1) \geq \max\{\beta(\omega_1), p(\omega_1)\}$ and action a_3 otherwise. But then,

$$\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) =$$

$$= m(s_1)\left[-\frac{1}{1-\max\{\beta(\omega_1), p(\omega_1)\}} p|s_1(\omega_2)\right] + m(s_2)\left[-\frac{1}{\min\{\beta(\omega_1), p(\omega_1)\}} p|s_2(\omega_1)\right] - (-1) =$$

$$= m(s_1)\left[1 - \frac{1}{1-\max\{\beta(\omega_1), p(\omega_1)\}} p|s_1(\omega_2)\right] + m(s_2)\left[1 - \frac{1}{\min\{\beta(\omega_1), p(\omega_1)\}} p|s_2(\omega_1)\right] > 0$$

which is true, because $1 - \frac{1}{1-\max\{\beta(\omega_1),p(\omega_1)\}} p|s_1(\omega_2) \geq 0$, $1 - \frac{1}{\min\{\beta(\omega_1),p(\omega_1)\}} p|s_2(\omega_1) \geq 0$, and one of the two inequalities has to be strict.

Therefore, if it is not the case that $\beta(\omega_1) \in [p|s_2(\omega_1), p|s_1(\omega_1)]$ and $\beta|s_1(\omega_1) \leq \beta|s_2(\omega_1)$, then there exists a decision problem $\mathcal{D}(\mathcal{A}, u)$ in which $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) > 0$.

Next, I will show that if $\beta(\omega_1) \in [p|s_2(\omega_1), p|s_1(\omega_1)]$ and $\beta|s_1(\omega_1) \leq \beta|s_2(\omega_1)$, then $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \leq 0$ in all decision problems $\mathcal{D}(\mathcal{A}, u)$. Since the agent does not in-

terpret the signal as noise, $\beta|s_1(\omega_1) < \beta(\omega_1) \leq p|s_1(\omega_1)$ and $p|s_2(\omega_1) \leq \beta(\omega_1) < \beta|s_2(\omega_1)$. As shown in the proof of the previous lemma, $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$ can be rewritten as $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = E_{s\sim m}\left[d\left(p|s, \beta\right) - d\left(p|s, \beta|s\right)\right]$. In order to show that $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \leq 0$, it is sufficient to show that $\forall s_i \in \mathcal{S}$, $d\left(p|s, \beta\right) - d\left(p|s, \beta|s\right) \leq 0$. This is true because $E_{\omega\sim q}\left[u\left(a_q, \omega\right)\right]$ is weakly convex, which implies that, in each direction, $d\left(p|s, q\right)$ weakly increases the further $q(\omega_1)$ is from $p|s(\omega_1)$, and because $\forall s_i \in \mathcal{S}$ $|\beta|s_i(\omega_1) - p|s_i(\omega_1)| > |\beta(\omega_1) - p|s_i(\omega_1)|$ by assumption. □

Proof of Proposition 6

I will start by showing that statement 3. implies statement 2. Consider encoding mapping $\{m|\omega\}_{\omega\in\Omega}$. Since $\{m|\omega\}_{\omega\in\Omega}$ is an encoding mapping, we know that $\forall\omega \in \Omega$, $\exists s \in \mathcal{S}$ such that $m|\omega(s) > 0$. Since information is fully encoded, we know that $\forall\omega' \in \Omega$ with $\omega' \neq \omega$, $m|\omega'(s) = 0$. Since information is correctly decoded, we know $\beta|s(\omega) = \frac{m|\omega(s)\beta(\omega)}{\sum_{\omega'\in\Omega} m|\omega'(s)\beta(\omega')}$. Therefore, we can rewrite $\beta|s(\omega)$ as:

$$\beta|s(\omega) = \frac{m|\omega(s)\beta(\omega)}{\sum_{\omega'\in\Omega} m|\omega'(s)\beta(\omega')} = \frac{m|\omega(s)\beta(\omega)}{m|\omega(s)\beta(\omega) + \sum_{\omega'\in\Omega,\omega'\neq\omega} m|\omega'(s)\beta(\omega')} = \frac{m|\omega(s)\beta(\omega)}{m|\omega(s)\beta(\omega)} = 1$$

Therefore, the biased decoder has completely accurate posterior beliefs.

I will now show that statement 2. implies statement 3.

Step 1: if the agent has completely accurate posterior beliefs, then information must be fully encoded. Assume, aiming towards contradiction, that the agent has completely accurate posterior beliefs but that information is not fully encoded. Then, $\exists\omega, \omega' \in \Omega$ with $\omega' \neq \omega$ and $s \in \mathcal{S}$, such that $m|\omega(s) > 0$ and $m|\omega'(s) > 0$. Since the agent has completely accurate beliefs, $\beta|s(\omega) = 1$ and $\beta|s(\omega') = 1$, which is a contradiction. Therefore, if full information is transmitted, then information must be fully encoded.

Step 2: if the agent has completely accurate posterior beliefs, then information must be correctly decoded. Suppose, aiming towards contradiction, that the agent has completely accurate posterior beliefs and that $\exists\omega \in \Omega$ and $s \in \mathcal{S}$ with $m|\omega(s) > 0$ such that $\beta|s(\omega) \neq \frac{m|\omega(s)\beta(\omega)}{\sum_{\omega'\in\Omega} m|\omega'(s)\beta(\omega')}$. Since the agent has completely accurate posterior beliefs, we know $\beta|s(\omega) = 1$. $\beta|s(\omega) \neq \frac{m|\omega(s)\beta(\omega)}{\sum_{\omega'\in\Omega} m|\omega'(s)\beta(\omega')}$ implies

$$\frac{m|\omega(s)\beta(\omega)}{\sum_{\omega'\in\Omega} m|\omega'(s)\beta(\omega')} < 1$$

which in turn implies

$$0 < \sum_{\omega'\in\Omega,\omega'\neq\omega} m|\omega'(s)\beta\left(\omega'\right)$$

Since we assumed $\beta(\omega') > 0 \ \forall \omega' \in \Omega$ and since $m|\omega'(s) \geq 0 \ \forall \omega' \in \Omega$, there must exist $\omega'' \neq \omega \in \Omega$ s.t. $m|\omega''(s) > 0$. But then, information is not fully encoded. Since, by step 1, the agent has completely accurate beliefs implies the fact that information is fully encoded, we have a contradiction. Therefore, statement 2. implies statement 3.

Lastly, I will show that statement 2. and statement 1. are equivalent, starting with 2. implying 1. Since the agent has completely accurate posterior beliefs,

$$\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) = E_{\omega\sim p}\left[u\left(a_{\delta_\omega}, \omega\right)\right] - E_{\omega\sim p}\left[u\left(a_\beta, \omega\right)\right] = \overline{\mathcal{V}}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$$

In order to show that statement 1. implies statement 2., I will show the contrapositive. If the agent does not have completely accurate beliefs, then there exists a decision problem $\mathcal{D}(\mathcal{A}, u)$ in which $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) \neq \overline{\mathcal{V}}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$. Let $\mathcal{A} = \Delta(\Omega)$, denote an action $a \in \mathcal{A}$ by vector $(a_1, ..., a_n)$, and, $\forall i \in \{1, ..., n\}$ let $u(a, \omega_i) = \ln(a_i)$. Then, $E_{\omega\sim q}[u(a, \omega)] = \sum_{i=1}^{n} q(\omega_i)\ln(a_i)$. We know that $\arg\max_{a\in\mathcal{A}} E_{\omega\sim q}[u(a, \omega)] = \{(q(\omega_1), ..., q(\omega_n))\}$ (Aczél and Daróczy, 1975). Therefore,

$$\max_{a\in\mathcal{A}} E_{\omega\sim q}[u(a, \omega)] = E_{\omega\sim q}[u(a_q, \omega)] = \sum_{\omega\in\Omega}[q(\omega)\ln(q(\omega))]$$

Notice $E_{\omega\sim q}[u(a_q, \omega)]$ is a strictly convex function in q and notice that $E_{\omega\sim p}[u(a_{\delta_\omega}, \omega)] = 0$. Since the agent does not have completely accurate posterior beliefs, $\exists\tilde{\omega} \in \Omega$, $\tilde{s} \in \mathcal{S}$ s.t. $m|\tilde{\omega}(\tilde{s}) > 0$ and $\beta|\tilde{s}(\tilde{\omega}) < 1$. There are two cases to consider: $\beta|\tilde{s}(\tilde{\omega}) = p|\tilde{s}(\tilde{\omega})$ and $\beta|\tilde{s}(\tilde{\omega}) \neq p|\tilde{s}(\tilde{\omega})$. If $\beta|\tilde{s}(\tilde{\omega}) = p|\tilde{s}(\tilde{\omega}) < 1$, then

$$E_{\omega\sim p|\tilde{s}}\left[u\left(a_{p|\tilde{s}}, \omega\right)\right] = \sum_{\omega\in\Omega}[p|\tilde{s}(\omega)\ln(p|\tilde{s}(\omega))] = p|\tilde{s}(\tilde{\omega})\ln(p|\tilde{s}(\tilde{\omega})) + \sum_{\omega\in\Omega,\omega\neq\tilde{\omega}}[p|\tilde{s}(\omega)\ln(p|\tilde{s}(\omega))] < 0$$

Therefore, $E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s}, \omega\right)\right]\right] < E_{\omega\sim p}[u(a_{\delta_\omega}, \omega)]$ and

$$\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) < \overline{\mathcal{V}}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$$

If $\beta|s(\omega) \neq p|s(\omega)$, then consider the decomposition of $\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$ from Proposition 1. The fact that $E_{\omega\sim q}[u(a_q, \omega)]$ is a strictly convex function implies $d(p|\tilde{s}, \beta|\tilde{s}) > 0$. Therefore, $E_{s\sim m}[d(p|s, \beta|s)] > 0$ and

$$\mathcal{V}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right) =$$

$$= \left\{E_{s\sim m}\left[E_{\omega\sim p|s}\left[u\left(a_{p|s}, \omega\right)\right]\right] - E_{\omega\sim p}[u(a_p, \omega)]\right\} + d(p, \beta) - E_{s\sim m}[d(p|s, \beta|s)] <$$

$$< \{E_{\omega\sim p}[u(a_{\delta_\omega}, \omega)] - E_{\omega\sim p}[u(a_p, \omega)]\} + d(p, \beta) = \overline{\mathcal{V}}\left(p, \{p|s\}_{s\in\mathcal{S}}, \beta, \{\beta|s\}_{s\in\mathcal{S}}\right)$$

Therefore, statement 1. implies statement 2.

This completes the proof of the proposition. □

B Relation to Literature Outside of Economics

The definition of B-EVSI builds on the statistics literature on proper scoring rules and their connections to generalized measures of uncertainty and divergences (Dawid 1998, 2007; Dawid and Sebastiani, 1999; Gneiting and Raftery, 2007). The basic model of communication explored in this paper builds on the seminal work of Claude Shannon (Shannon, 1948). By explicitly modeling the receiver's beliefs and interpretations, this paper takes a step forward in applying the ideas of information theory to human communication as discussed below (Shannon and Weaver, 1949; Shannon, 1956). The literature in artificial intelligence has been struggling with testing for and measuring the emergence of communication in multi-agent reinforcement learning settings (see Lowe et al. 2019 for a review). The measures of meaningful information transmission introduced in this paper should: i) help test for the emergence of communication in sender-receiver games featuring reinforcement learning, and ii) help measure the degree to which impairments to communication are due to partial encodings vs. incorrect decodings.[35] Lastly, the measures of meaningful information transmission introduced in this paper can be used as stepping stones in the development of theories of language evolution in mathematical biology (Nowak, 1999; Corominas-Murtra, Fortuny, and Solé, 2014).

Shannon (1948) In his introduction to Shannon's 1948 seminal paper "A mathematical theory of communication", mathematician and science communicator Warren Weaver identified three fundamental questions in the study of communication: a) "how accurately can the symbols of communication be transmitted?" b) "How precisely do the transmitted symbols convey the desired meaning?" and c) "How effectively does the received meaning affect conduct in the desired way?" Weaver referred to the three questions above as the technical problem, the semantic problem, and the effectiveness problem, respectively (Shannon and Weaver, 1949). Shannon's paper tackled the technical problem, but Weaver speculated that his insights could be useful in addressing the other two problems as well (Shannon and Weaver, 1949).

In this paper, I ignore the technical problem of communication altogether by assuming that the symbols of communication can be transmitted without any interference and focus on the other two problems. I begin by asking a broad question, namely "how does the observation of a signal affect the actions and, therefore, the welfare of a subjective expected utility maximizer in the context of a decision problem under uncertainty?" Of course, the answer depends on: i) the nature of the decision problem, ii) the information environment, and iii) the agent's perceptions of the information environment.

The framework I develop to tackle the question above helps me formally address the semantic problem of communication. In the context of one-way communication between a sender and a re-

[35] Lowe et al. (2019) refer to partial encodings as failures in "positive signaling" and incorrect decodings as failures in "positive listening."

ceiver, the information environment that the receiver encounters is in part controlled by the sender. In particular, the sender's encoding mapping determines the association between states of the world or concepts and arbitrary symbols. Such association lays the pre-conditions for referential meaning to be conveyed to the receiver in the process of communication. In order for referential meaning to actually be conveyed in the process of communication, the receiver must interpret the symbols of communication correctly.[36] Therefore, the degree to which the symbols of communication convey meaning depends on the joint properties of the sender's encoding mapping and the receiver's decoding strategy. The measures of meaningful information transmission introduced in Section 4 – which include a generalization of Shannon's mutual information – capture the degree to which the symbols transmitted in the context of one-way communication convey the referential meaning encoded by the sender. Thus, they provide a formal answer to the question that, according to Weaver, defines the semantic problem.

The framework developed in this paper also provides a formal answer to the effectiveness problem.[37] Specifically, one can measure the degree to which the sender's messages affect the receiver's conduct in the desired way by endowing the sender with a payoff function that depends on the receiver's actions. The appropriate payoff function will, of course, depend on context. For instance, the sender might want the receiver to develop as accurate beliefs as possible about the realized state of the world. In that case, the sender's payoff function might simply be one of the measures of meaningful information transmission introduced in Section 4. Conversely, the sender might want the receiver to develop inaccurate beliefs about the state of the world; in that case, as shown in Section 5, the sender might attempt to deceive the receiver.

C Empirical Measurement of Meaningful Information

The measures of meaningful information transmission introduced in Section 4 can be easily estimated on empirical data. In this section, I present two examples of how such estimation can be done. The first example uses the dataset from Ambuehl and Li (2018) to study the degree to which, in the context of a simple belief updating experiment, meaningful information transmission is impaired as a result of subjects failing to update their beliefs according to Bayes' rule. The second example uses the dataset from Braghieri (2021) to study the degree to which, in the context of an experiment about political correctness, meaningful information transmission is impaired as a result of subjects having incorrect beliefs about an endogenously determined signal structure.

Measures of meaningful information transmission can be deployed in many other settings, ranging from discrimination (Arrow 1974; Bohren, Haggag, Imas, Pope, 2021), to biased memory (Zimmermann, 2020; Gödker, Jiao, and Smeets, 2022; Bordalo, 2023), to cursed behavior (Eyster

[36] See Lewis (1969) for a philosophical perspective on how referential meaning arises in the context of communication.

[37] In fact, the entire game theoretic approach to communication can be seen addressing the effectiveness problem.

and Rabin, 2005; Deversi, Ispano, and Schwardmann, 2021), to lying experiments (Gneezy, 2005), to information cascades (Banerjee, 1992; Bikchandani, Hirshleifer, and Welch, 1992; Cipriani and Guarino, 2005), to cheap talk experiments (see Blume, Lai, and Lim, 2020 for a review), to motivated reasoning (Kunda, 1990), to experiments eliciting preferences over information structures (Matsalioglu, Orhun, and Raymond, 2017), etc.

Ambuehl and Li (2018) Ambuehl and Li (2018) present the results of a laboratory experiment aimed at studying how individuals value noisy information. As part of the experiment, the authors ask subjects to perform a simple belief updating task in a setting with a binary state of the world and a binary signal. Specifically, after being informed that the two states of the world have equal ex-ante probability of occurring and after being shown a signal structure, subjects are asked to report their posterior beliefs about the state of the world conditional on each possible signal realization. Such posteriors are incentivized for accuracy. If subjects are familiar with Bayes' rule, they can trivially calculate the correct posteriors and extract full meaningful information from the signal. Conversely, if subjects are not familiar with Bayes' rule, they might hold subjective posteriors that differ from the Bayesian benchmark and, as a result, they might extract less than full meaningful information from the signal. In fact, if subjects' posteriors are very inaccurate, subjects might even extract a negative amount of meaningful information from the signal.

Using the notation from Section 3, the setting can be summarized as follows. The set of states of the world is $\Omega = \{\omega_1, \omega_2\}$. Subjects face ten different signal structures in random order, but, for the purpose of this paper, it is sufficient to focus the four simplest structures $\left(\mathcal{S}, \left\{m^j|\omega\right\}_{\omega\in\Omega}\right)$ for $j \in \{1,2,3,4\}$. The signal space is the same for each signal structure, namely $\mathcal{S} = \{s_1, s_2\}$. The four simplest structures are symmetric – meaning that $\forall j \in \{1,2,3,4\}$ $m^j|\omega_1(s_1) = m^j|\omega_2(s_2)$ – and differ in their precision. Specifically $m^1|\omega_1(s_1) = 0.6$, $m^2|\omega_1(s_1) = 0.7$, $m^3|\omega_1(s_1) = 0.8$ and $m^4|\omega_1(s_1) = 0.9$. Subjects are informed about the prior distribution of states of the world. Since the two states have an equal probability of being drawn, $\beta(\omega_i) = p(\omega_i) = \frac{1}{2}$ $\forall i \in \{1,2\}$. Subjects are informed about the signal structure; therefore, for every $\omega \in \Omega$, $\mu|\omega = m|\omega$. The posterior distribution over states of the world conditional on each signal realization $\{p|s\}_{s\in\mathcal{S}}$ can be recovered using Bayes' rule. However, since experimental subjects might not employ Bayes' rule when updating their beliefs, their subjective posteriors $\{\beta|s\}_{s\in\mathcal{S}}$ need not equal $\{p|s\}_{s\in\mathcal{S}}$.

Using Shannon's entropy function as the baseline measure of uncertainty, Figure A1 computes the amount of meaningful information that subjects extract from each of the four signal structures. Since subjects are informed about the prior distribution of states of the world, the expression for meaningful information boils down to $I(s,\omega) - E_{s\sim m}\left[D_{KL}(p|s, \beta|s)\right]$, where $I(\cdot)$ denotes mutual information. As discussed in Section 4, when the baseline measure of uncertainty is Shannon's entropy function, mutual information – denoted by the solid red line in the figure – is an upper bound for the amount of meaningful information transmitted. In the context of the experiment

by Ambuehl and Li (2018), such upper bound is reached whenever a subject updates her beliefs according to Bayes' rule.

Figure A1 exhibits four features worth highlighting. First, for each signal structure, a majority of subjects fails to extract full meaningful information from the signal. A plurality of subjects does extract close to full meaningful information, but such plurality is relatively narrow and ranges between 15% and 30% depending on the signal structure. Second, for the least informative signal structure according to the Blackwell order, namely Signal Structure 1, more than 50% of subjects extract a negative amount of meaningful information from the signal. Therefore, when faced with Signal Structure 1 and when the accuracy of beliefs is measured according to Shannon's entropy function, more than 50% of subjects update their beliefs in such an erratic way as to develop less accurate beliefs about the state of the world after sampling from the signal. Third, average accuracy – measured as the difference between mutual information and meaningful information averaged across subjects – improves the more informative the signal structure. This suggests that subjects find it particularly hard to update their beliefs based on relatively uninformative signal structures. Fourth, consistent with Proposition 3, the more informative the signal structure, the smaller the fraction of individuals who extract a negative amount of meaningful information from the signal.

Overall, the analysis of the dataset from Ambuehl and Li (2018) shows that, even in the context of an extremely simple belief updating task, subjects might not be able to extract full meaningful information from a signal. In fact, they might be so unskilled at Bayesian updating as to extract a negative amount of information, especially for signal structures that are not very informative to begin with.

Braghieri (2021) The second example employs the dataset from Braghieri (2021). Braghieri (2021) presents the results of two experiments aimed at studying the degree to which political correctness impairs information transmission among college students. In the first experiment, subjects are randomized into a private and a public treatment. Subjects assigned to the private treatment are given a guarantee of anonymity, whereas subjects assigned to the public treatment are given hints suggesting that their answers will be shared with other participants in the study. Next, subjects are asked to report the extent to which they agree or disagree with a set of politically sensitive statements.

One of the statements is about whether the U.S. government should provide reparations for slavery. On average, subjects assigned to the public treatment are more likely to report that they agree with the statement than subjects assigned to the private treatment. The average effect is highly heterogeneous across students with different political ideologies. Specifically, among students who self-identify as liberal, the fraction agreeing with the statement is approximately equal across treatments: 78% in the private treatment and 76% in the public treatment. Conversely, among

students who self-identify as moderate or conservative, the fraction agreeing with the statement is quite different across treatments: 27% in the private treatment and 50% in the public treatment. As a result, the degree to which agreeing with the statement about reparations for slavery is diagnostic of self-reported political ideology is much higher in the private than in the public treatment.

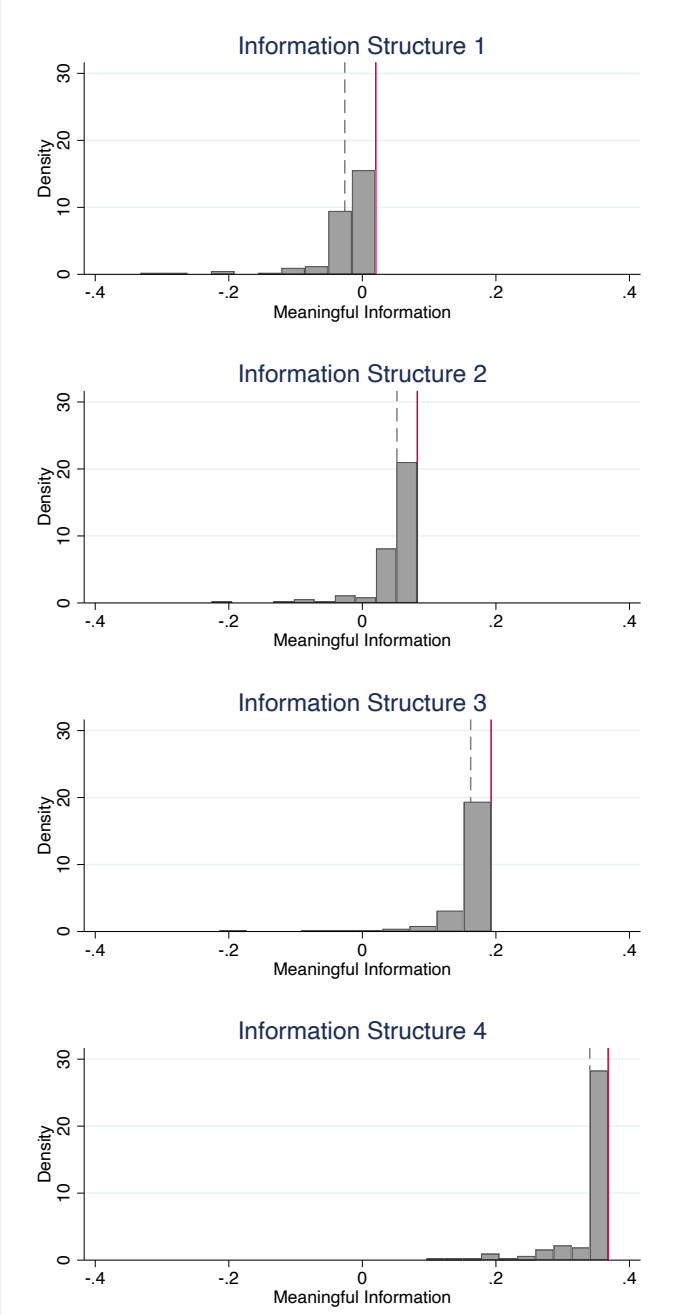
The second experiment involves recruiting a new set of subjects from the same pool and and asking them to forecast the behavior of participants in the first experiment. Specifically, one of the questions entails: i) showing subjects in the second experiment the fraction of self-identified liberals and moderates/conservatives who, in the private treatment of the first experiment, agreed with the statement about reparations for slavery, and ii) asking those subjects to forecast the corresponding fractions for the public treatment of the first experiment. Since agreeing or disagreeing with the statement about reparations for slavery is a diagnostic signal for self-reported political ideology, the question is designed to isolate the beliefs of subjects in the second experiment about the signal structure generated by the answers of participants in the public treatment of the first experiment.

Using Shannon's entropy function as the baseline measure of uncertainty, Figure A2 computes the amount of meaningful information implied by subjects' beliefs about the signal structure generated by the answers of participants in the public treatment of the first experiment. In order to isolate impairments to meaningful information transmission due to the subjects' imperfect understanding of the signal structure, I assume that subjects know the prior distribution of states of the world and, for each subject, I compute posterior distributions according to Bayes' rule. Therefore, the only possible impairments to meaningful information transmission come from the subjects' imperfect understanding of the encoding mapping.

The figure shows that the average amount of meaningful information extracted by subjects is close to zero. Furthermore, it shows that about half of subjects extract a negative amount of meaningful information from the signal.

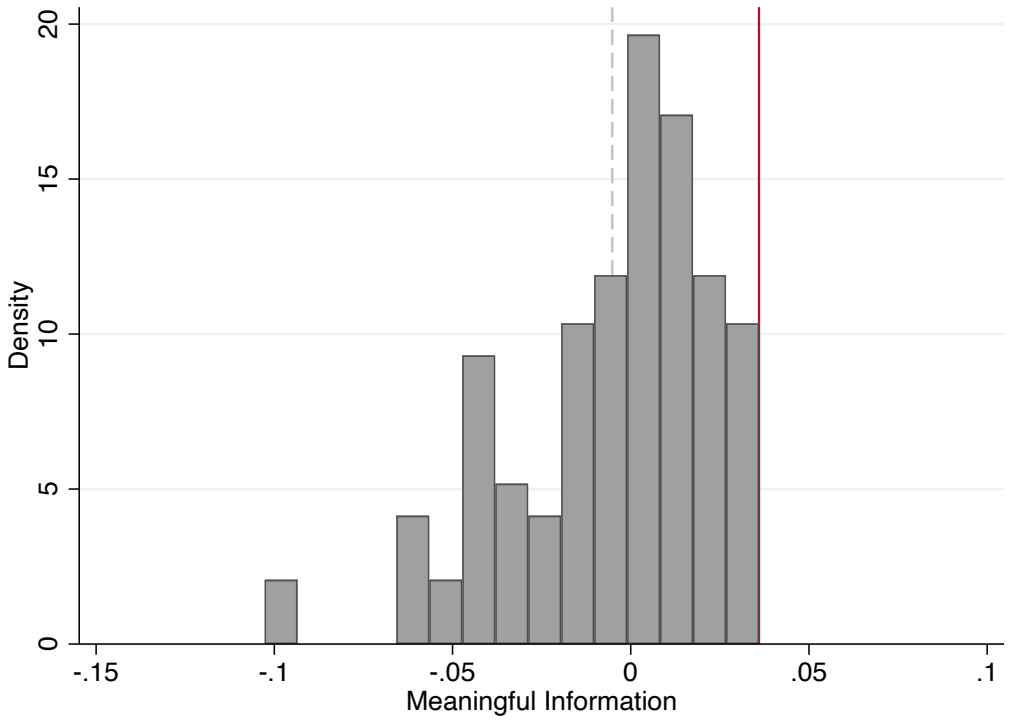
Overall, the analysis of the dataset from Braghieri (2021) shows that, when the signal structure is endogenously determined by the behavior of a sender or a set of senders, receivers might have incorrect beliefs about the encoding mapping and, as a result, might not be able to extract full meaningful information from the signal.

Figure A1: **Distribution of Meaningful Information from Ambuehl and Li (2018)**



Notes: The figure shows, for each of the four symmetric signal structures from Ambuehl and Li (2018), the empirical distribution of the amount of meaningful information extracted by experimental subjects. The solid red line describes mutual information, which is the upper bound on the amount of meaningful information that subjects can extract. The dashed gray line marks the average amount of meaningful information extracted. The baseline measures of uncertainty is Shannon's entropy function. In this figure, subjects know the prior distribution of states of the world and the signal structure, but have to calculate posterior probabilities.

Figure A2: **Distribution of Meaningful Information from Braghieri** (2021)



Notes: The figure shows the empirical distribution of the amount of meaningful information extracted by experimental subjects in the experiment from Braghieri (2021). The solid red line describes mutual information, which is the upper bound on the amount of meaningful information that subjects can extract. The dashed gray line marks the average amount of meaningful information extracted. The baseline measures of uncertainty is Shannon's entropy function. In this figure, subjects know the prior distribution of states of the world and their posteriors are calculated mechanically according to Bayes' rule, but subjects might have incorrect beliefs about the signal structure.

D Thwarted Deception

The last cell of Table 1 relates to settings in which the sender has dishonest intentions but in which, nevertheless, the receiver interprets messages correctly. This might occur, for instance, when the sender tries to deceive the receiver and fails.

Definition 19. The receiver thwarts the sender's attempt at deception in realized state of the world $\omega \in \Omega$ if $s \in \mathcal{S}$ is intended to be deceptive in state ω and $\beta|s(\omega) = p|s(\omega)$.

From the point of view of the receiver, the consequences of successfully thwarting the sender's attempt at deception are the same as the consequences of successful communication described in Corollary 2. Specifically, whenever the receiver thwarts the sender's attempt at deception, the value of communication to the receiver is positive in all decision problems $\mathcal{D}(\mathcal{A}, u)$ and the amount of meaningful information transmitted to the receiver is positive for all measures of meaningful information characterized in Section 4.