Noisy persuasion*

Elias Tsakas†
Maastricht University

Nikolas Tsakas‡
University of Cyprus

First draft: March, 2017
This draft: December, 2020

Abstract

We study the effect of noise due to exogenous information distortions in the context of Bayesian persuasion. In particular, we first provide a full characterization of the optimal signal in a standard special case that has attracted a lot of attention in the literature. Then, we ask whether more noise (à la Blackwell) is always harmful for the information designer, i.e., the sender. We show that in general this is not the case. We provide a necessary and sufficient condition for the sender to always be worse off when noise increases in a binary noisy channel. There are two ways to read our result: (a) the sender always dislikes additional noise if and only if we start with little noise in the first place, (b) the sender always dislikes additional noise if and only if this additional noise is modelled by a sufficiently symmetric channel. Then, we provide sufficient conditions that extend this result to channels of arbitrary cardinality. Finally, we show that in every noisy persuasion game, increased complexity of the message space makes the sender weakly better off, while for a rather rich class of games the improvement is strict. This is in contrast to the noiseless case, where the sender's maximum expected utility can always be achieved with a bounded number of messages.

Keywords: Bayesian persuasion; noisy channel; data distortions; garbling; complexity.
JEL codes: C72, D72, D82, D83, K40, M31.

1. Introduction

Information distortions are among the most common and widely-studied phenomena in many areas within economics. The interest in the subject stems primarily from the fact that such noise often leads to inefficiencies. In this paper we study the effect of exogenous data distortions in the context of the recently surging literature on Bayesian persuasion.

(Bayesian) persuasion games are sender-receiver games with commitment (Rayo and Segal, 2010; Kamenica and Gentzkow, 2011). In particular, an information designer (viz., the female sender) chooses an experiment which is commonly known; the decision-maker (viz., the male receiver) observes an outcome of the experiment (viz., a message) and subsequently takes an action that affects

*We are indebted to Matthew Gentzkow and Stefan Terstiege for valuable comments and discussions. We would also like to thank anonymous referees and an associate editor for their insightful reports that helped us improve the paper. Elias thanks the Department of Economics at UC Davis for its hospitality while working on this project.

†Department of Economics (MPE), Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands; Homepage: http://www.elias-tsakas.com/home.html; E-mail: e.tsakas@maastrichtuniversity.nl

‡Department of Economics, University of Cyprus, P.O. Box 20537, 1678, Nicosia, Cyprus; Homepage: https://sites.google.com/site/nikolastsakas/; E-mail: tsakas.nikolaos@ucy.ac.cy

both agents. Now, an issue that can arise in such processes is that the message observed by the receiver may often be different from the one that was actually realized during the experiment, i.e., the experimental data are often distorted. Such distortions can be attributed to errors, often appearing in some of the following instances:

- Data gathering: The experiment is run by an agent, henceforth called the data collector, who could in principle be the sender or the receiver or even a third party. The data collector observes a noisy version of the realized message, due to *measurement errors* in the experimental implementation.

- Data processing: The raw data is gathered by the data collector and is processed before being used by the receiver. Processing can take the form of storage (either in the collector's memory or in some external device) and retrieval at a later time, in which case noise is attributed to *memory constraints*. Alternatively, processing errors can be due to the collector's *lack of expertise* which precludes him/her from correctly encoding or interpreting the actual message.

- Data transmission: The data collector is some agent other than the receiver who (truthfully) communicates the observed data to the receiver. The receiver observes a noisy version of the transmitted message, due to *communication errors* or *language barriers* that lead to misunderstanding of the communicated data.[1]

The obvious restriction that noise poses to the sender is that it restricts the set of signals from which she can effectively choose. The result for the sender is twofold: First, the noise limits the posteriors that can be formed by the receiver upon observing some outcome of the experiment, as there are posteriors that no experiment can induce. For instance, if the receiver observes the actual message with probability $1 - \varepsilon$ and every other message with a small positive probability, there is no experiment that can reveal to the receiver the true state with certainty. Second, it restricts the possible combinations of posteriors that the receiver can form upon observing different messages associated with the same experiment. This brings us naturally to our first research question. Namely, *how does an optimal signal look like in the presence of noise?*

From the previous discussion, it becomes clear that adding noise to an otherwise noiseless environment would be necessarily harmful for the sender. However, the previous argument applies only when we compare the noiseless case with a noisy one. This yields a natural question, as to whether the argument extends to cases where we compare any two types of data distortions with one being noisier than the other. Formally, *is the sender's expected utility in equilibrium increasing with respect to the (Blackwell) informativeness of the noisy channel that describes the respective distortions?*

Our third and final question focuses on the complexity of the noisy channel. In particular, recall that in the standard noiseless case, it suffices for the sender to use a given number of signals in order to achieve her maximum expected utility. *Is this still true in the presence of noise?*

Answering the previous questions can have important implications on the choice of communication channels, delegations or mediators. For instance, if a pharmaceutical company is seeking regulator's approval for some drug, should they choose to perform the clinical trials inhouse or should they delegate the implementation of the experiment to some clinical trial experts? Also, should they communicate the outcome of the medical experiment directly or through mediators (e.g., lawyers)? Similarly, when a local politician tries to attract visitors to her province by persuading them that water quality in the local beach is high, what system should she use to communicate to the public the results of water control? In particular, should she use a system with many different grades/colors? Likewise, how much detail should be contained in the report provided by a physician to a patient

[1] For related stylized facts from a leading medical journal, see Flores (2006).

who has limited understanding of medical terms or faces other communication barriers (e.g., they don't share a common mother tongue)?

Starting with our first general question, not surprisingly, the answer turns out to be very complex to allow us to provide a general characterization for all persuasion games and all noisy channels. Nevertheless, we fully solve the problem for the standard application that widely appears in the literature, i.e., the prosecutor-judge example of Kamenica and Gentzkow (2011). In particular, suppose that a municipality (viz., sender) announces that an environmental experiment will be carried out to test the water quality at the local beach. The outcomes of the experiment are presented to the potential swimmer (viz., receiver) in the form of flag colors. The municipality wants to persuade the swimmer visit the beach irrespective of the water quality, whereas the swimmer wants to swim in these waters if and only if the quality is good. Now, the caveat compared to earlier works in this literature, is that the swimmer may confuse the different colors, i.e., he attaches the correct meaning to each flag with probability $1-\varepsilon$ and misinterprets all other colors with the same (small) probability. Then, our characterization result states that in equilibrium all but one colors will (just) persuade the swimmer to visit the beach (Proposition 1). Moreover, the total probability of the swimmer being persuaded is proportional to the error probability, and inversely-proportional to the number of colors that the grading system uses.

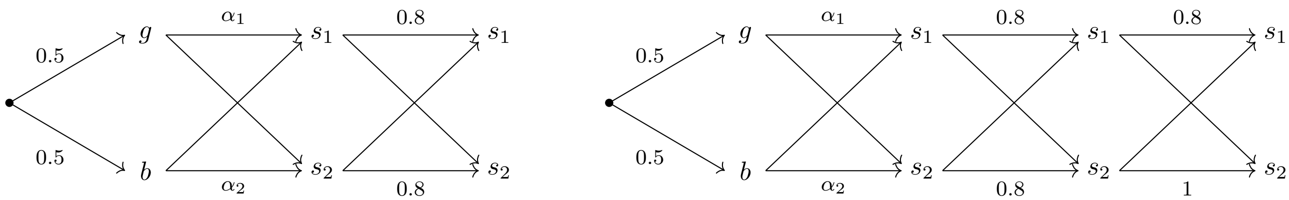
Let us now turn to our second question. As it turns out, surprisingly, the answer is in general negative, i.e., there exist pairs of noisy channels, one being a garbling of the other (Blackwell, 1951, 1953), such that the sender's expected utility under the garbled channel (viz., the more noisy one) is strictly larger than the expected utility under the original channel (viz., the less noisy of the two). Thus, *adding more noise to an already noisy environment can be beneficial for the sender*.

Let us consider the following example that will be used repeatedly throughout the paper. Suppose that a pharmaceutical company (viz., sender) wants to submit an application to the regulator (viz., receiver) for an experimental drug to be approved for commercial use. It is ex-ante commonly known that the drug is effective at the good state (which occurs with probability 0.5) and ineffective at the bad state. An application consists of a clinical trial (viz., an experiment), which is designed by the company and produces evidence (viz., one of two possible messages) that leads to some updated probability of the state being good. This evidence is sent to the regulator, who then truthfully announces the updated probability in a press release. The company has reputation concerns in the sense that her utility is increasing in the regulator's announcement, with a jump at 0.8, which is the probability at which the drug is being approved. This jump is sufficiently large to ensure that the company always prefers trials that can lead to the drug's approval (i.e., it can yield evidence that would lead to an updated probability at least as high as 0.8) compared to trials that cannot.[2] The standard research question in the noiseless persuasion model of Kamenica and Gentzkow (2011) is to characterize the optimal experiment for the sender, and it is typically answered by identifying a pair of posteriors (i.e., a good one above 0.5 and a bad one below 0.5) which in expectation are equal to the prior. In the current setup – with the preferences we describe above – the optimal experiment of the standard noiseless case would yield either a (good) posterior equal to 0.8 or (a bad) one equal to 0.3.

However, communication is noisy, thus the regulator might observe a different message than the one that was actually transmitted. In the first scenario – presented in Figure 1(a) – noise is such that either message is wrongly transmitted with probability 0.2. In this case, the only experiment that can yield a good posterior as high as 0.8 is the fully informative one, i.e. one which would reveal the true state. This, however, necessarily yields a bad posterior equal to 0.2. Hence, the sender is strictly worse off compared to the noiseless case. In the second scenario – presented in Figure 1(b) – on top

[2]For a more detailed form of the company's utility function see Figure 5.

of the potential mistakes in evidence transmission, evidence leading to the good posterior might also be misinterpreted as being bad with probability 0.2. It turns out that the addition of this new type of distortion is not detrimental for the sender, as the fully informative experiment now yields the same pair of posteriors as the optimal experiment of the noiseless case, i.e. a good posterior equal to 0.8, or a bad one equal to 0.3. Thus, in this case, more noise is actually beneficial for the sender.



(a) The first noisy channel p is such that the regulator observes the opposite message than the one that was sent with probability 0.2, i.e. $p(s_2|s_1) = p(s_1|s_2) = 0.2$.

(b) The second noisy channel q is a garbling of the first one, as the message passes also from a second channel r that further distorts the message, after the original distortion induced by the channel p. The probabilities of having such additional distortions are $r(s_2|s_1) = 0.2$ and $r(s_1|s_2) = 0$.

Figure 1: Garbling with a binary state space: The channel q (right) is more noisy than the channel p (left).

But then we naturally ask: *is it possible to identify conditions under which monotonicity (of the sender's expected utility with respect to the channel's informativeness) holds?* Our next result (Proposition 2) provides necessary and sufficient conditions for such a monotonic relationship in binary noisy channels. Subsequently, we partially extend our result by providing sufficient conditions for channels of arbitrary cardinality (Propositions 3 and 4).

There are two ways to read our first proposition. Both become clear once the result has been formally stated, but let us already give a preview. According to the first interpretation, the more noisy the channel we start with is, the easier it becomes for monotonicity to be violated. A lot of initial noise means that the set of feasible posteriors has already shrunk significantly, and therefore even a little additional noise suffices for distributions of posteriors that could not be reached originally to become feasible. Loosely speaking, if the receiver does not trust the accuracy of the data he observes in the first place, then additional mistrust is not necessarily detrimental for the sender. According to the second interpretation, more noise is always harmful when the second channel, which we use to garble the original one, is sufficiently symmetric. Intuitively, this makes it less likely for the receiver to observe the actual message, without the errors being distributed in a way that would favor some specific messages whose frequent observation could be potentially more beneficial for the sender (see our leading example). Loosely speaking, if the channel that we use to garble is sufficiently symmetric, we only increase the amount of noise without changing much the structure of the first noisy channel.

Finally, let us turn to the third question, reagarding the complexity of the noisy channel. For starters, recall that in the usual noiseless case, the number of messages that are needed for the sender to maximize her expected utility is bounded from above by the number of states. For instance, in our first example, there are only two states (i.e., the water quality is either good or bad). Thus, if the swimmer does not confuse colors, the municipality only needs two flags (viz., blue and red) to maximize the probability of persuading the swimmer. However, when we introduce noise this is no longer the case. In fact, already Proposition 1 suggests that, whenever noise has a nice symmetric structure, the sender strictly benefits from adding more colors. As it turns out, this is not an artefact of this particular game or of this particular noisy channel. Indeed, we can take any game, any channel and many different ways of increasing the complexity of said channel, and it will still be the case that making the channel more complex will weakly benefit the sender (Proposition 5). And for a rich class of games, the improvement is strict. In particular, the sender becomes better off, not only by introducing additional colors (e.g., yellow and green on top of blue and red), but also by introducing different shades of the same color (e.g., splitting blue into dark blue and light blue, and likewise for red).

While there are several earlier (e.g., Glazer and Rubinstein, 2004; Milgrom and Roberts, 1986) as well as contemporary (e.g., Rayo and Segal, 2010) influential papers on persuasion, we view Kamenica and Gentzkow (2011) as the natural predecessor of our work. Their paper has triggered wide interest in Bayesian persuasion. For a recent overview, we refer to Kamenica (2019).

Close to our work lies Le Treust and Tomala (2019) in which the authors study Bayesian persuasion in the presence of noise. In their paper, they consider information distortions, similarly to our work, but they allow for multiple experiments that are conducted sequentially. Then, they study the effect of noise on the sender's expected utility as the number of experiments increases. Issues pertaining to computational aspects of the Bayesian persuasion problem with noise are studied by Dughmi (2017), Dughmi et al. (2016) and Dughmi and Xu (2016).

Related is also the work of Kosenko (2018) who studies persuasion in the presence of mediators. In this case, noise becomes endogenous and strategic, unlike our case where it is exogenously given and independent of the experiment the sender chooses. The role of endogenous sender-imposed distortions with commitment is also studied in Frankel and Kartik (2020). In another recent paper, Guo and Shmaya (2020) study a setup in which the sender can provide inaccurate information, but this bears a direct cost for the sender, which captures possible reputation concerns.

In the context of strategic information transmission, Blume et al. (2007) introduce noisy communication to a standard cheap talk game á la Crawford and Sobel (1982) in an analogous way to our variant of Kamenica and Gentzkow's (2011) persuasion game. They show that noise may lead to increases of aggregate welfare, similarly to our leading example (though for different reasons). Related are also the papers of Blume and Board (2013, 2014), who study noise due to language barriers and intentional vagueness. The general problem of strategic information transmission through noisy channels is studied in Le Treust and Tomala (2018).

Finally, our model is also related to the one of Perez-Richet and Skreta (2018). In their work, a principal designs a test (which technically corresponds to the noisy channel in our case), a persuader chooses a manipulation technology (which technically corresponds to the experiment in our case), and the receiver decides whether to approve a technology or not. Our model takes the test as an exogenous parameter and focuses on the choice of the persuader, while in their case the focus is on how to design the test in order to avoid manipulation. Of course, while the two models papers bear similarities in the analysis and the results (see discussion after Proposition 1), the applications that they address are very different.

The paper is structured as follows: Section 2 introduces our model, and presents the equilibrium analysis and our leading example. Section 3 contains our main results on monotonicity of the sender's expected utility with respect to the informativeness of the noisy channel. Section 4 discusses our analysis on the importance of the complexity of the message space. Finally, Section 5 contains a concluding discussion. All proofs are relegated to the Appendices.

2. Persuasion game with noise

2.1. Signalling structures

Let $\Omega = \{\omega_1, \ldots, \omega_N\}$ be a (finite) set of states and A be a compact action space. There are two players, a (female) sender and a (male) receiver, with a common full-support prior $\mu_0 \in \Delta(\Omega)$, and continuous utility functions, $v : A \times \Omega \to \mathbb{R}$ and $u : A \times \Omega \to \mathbb{R}$ respectively. Whenever there are only two states in Ω, we identify the prior with the probability it attaches to ω_1, in which case with a slight abuse of notation we write $\mu_0 \in [0, 1]$.

Let $S = \{s_1, \dots, s_K\}$ be the finite set of messages that can be encoded with the available technology. A *(noisy) signalling structure* consists of an *experiment* $\pi : \Omega \to \Delta(S)$ chosen by the sender, and an exogenously given *(noisy) channel* $p : S \to \Delta(S)$ that may distort the message that was realized during the experiment.[3] Thus, $p(s'|s)$ denotes the probability that the receiver observes s' when s has been realized.

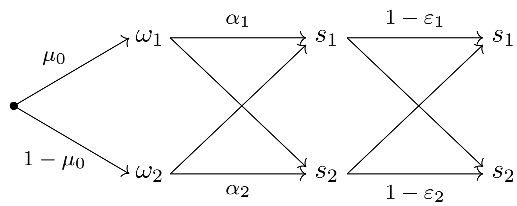


Figure 2: The sender chooses the experiment in the form of conditional probabilities, $\alpha_1 := \pi(s_1|\omega_1)$ and $\alpha_2 := \pi(s_2|\omega_2)$. A message $s \in S$ is first realized. Then, it is (possibly) distorted, with exogenous and commonly known error probabilities $\varepsilon_1 := p(s_2|s_1)$ and $\varepsilon_2 := p(s_1|s_2)$.

A channel is called binary whenever $K = 2$ (Figure 2). A channel is called noiseless whenever $p(s|s) = 1$ for all $s \in S$. Throughout the paper we assume that error probabilities are relatively small, suggesting that we depart relatively little from the original Bayesian persuasion game, i.e., formally, $p(s|s) > 1/2$ for all $s \in S$.

It is important to stress that *messages are not ex ante attached to a particular meaning*. Instead, meaning is acquired via the experiment and then distorted by the noisy channel. Notably, the error probabilities do not depend on the meaning that a message carries, but on the underlying technology, i.e., on how easy it is to confuse messages during gathering/processing/transmitting information.

Example 1. Suppose that a municipality (the sender) announces that an environmental experiment will be carried out to test the water quality at the local beach. The quality of the water is either good (state ω_1) or bad (state ω_2). The set of messages corresponds to the different flag colors that the beach can be awarded, viz., $\{$blue (s_1), red $(s_2)\}$. An experiment is identified by the conditional probabilities of obtaining each flag color given each quality level. However, with small probability, a swimmer (the receiver) may forget the meaning that is attached to each color. Such error probabilities depend on the choice of the messages, e.g., if the municipality had chosen to use $\{$orange (s_1), red $(s_2)\}$ instead of $\{$blue (s_1), red $(s_2)\}$, the error probabilities would have been larger, as it would have been easier to confuse orange with red than it is to confuse blue with red. On the other hand, if the municipality had decided to use the messages $\{$safe (s_1), dangerous $(s_2)\}$, the error probabilities would have been even lower. $\triangleleft$

Throughout the paper, we regularly focus on some special cases of noisy channels that we find interesting for studying certain applications. A channel p is called *symmetric* whenever $p(s|t) = p(t|s)$ for all $s, t \in S$. Such channels can be interpreted by means of an underlying metric that measures the

[3]In their recent paper, Le Treust and Tomala (2019) consider a sender who chooses an experiment $\pi : \Omega^n \to \Delta(S^k)$, and each message in the realized sequence $(s^1, \dots, s^k) \in S^k$ goes through the same noisy channel $p : S \to \Delta(S)$. The idea is that the sender designs n independent experiments that yield k data points that are independently distorted before being observed by the receiver. In this sense, our model can be viewed as a special case of theirs with $n = k = 1$.

distance between any two messages, and the error probability depends on said distance.[4] For instance, in the previous example, the probability of confusing blue with red is the same as the probability of confusing red with blue. The simplest form of a symmetric channel appears when the receiver hears the true message with probability $1 - \varepsilon$ and every other message with equal probability $\varepsilon/(K-1)$. These last channels are called *strongly symmetric*. In practice, we can assume a strongly symmetric channel when the messages cannot be bundled into similarity classes, e.g., in the previous example, if we use three primary colors {blue (s_1), red (s_2), yellow (s_3)}, the probability of confusing red with yellow is equal to the probability of confusing red with blue.

A signalling structure (π, p) induces a *signal* $\sigma : \Omega \to \Delta(S)$ such that

$$\sigma(s|\omega) = \sum_{t \in S} p(s|t)\pi(t|\omega). \tag{1}$$

With a slight abuse of terminology, we will often say that the sender chooses the signal σ rather than the experiment π. The set of experiments is denoted by Π, whereas the set of feasible signals (given the channel p) is denoted by $\Sigma_p \subseteq \Pi$, with equality holding if and only if p is noiseless. Each feasible signal $\sigma \in \Sigma_p$ is associated with a mapping from Ω to $\Delta(B_p) \subsetneq \Delta(S)$, where $B_p := \{p(\cdot|s_1), \ldots, p(\cdot|s_K)\} \subseteq \Delta(S)$ (see Figure 3). Representing Σ_p with $(\Delta(B_p))^\Omega$ will turn out to be quite useful in our proofs. Whenever it is clear which is the noisy channel that we have in mind, we

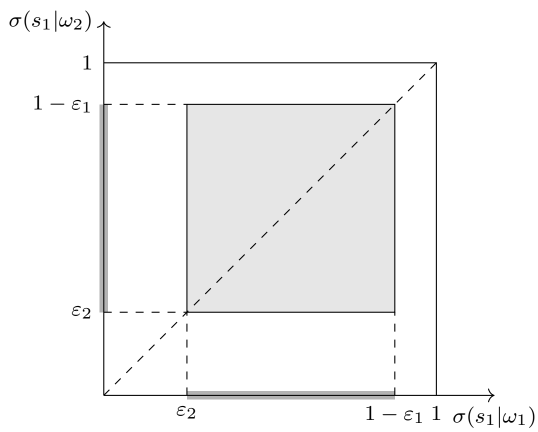


Figure 3: The set of feasible signals Σ_p for the binary channel p of Figure 2: The unit square represents the entire $\Pi = (\Delta(S))^\Omega$, with each element of $\Delta(S)$ being identified by the probability it attached to s_1. The shaded interval on each of the two axes is the convex hull of $B_p = \{\varepsilon_2, 1-\varepsilon_1\}$, where $\sigma(s_1|\omega) = 1 - \varepsilon_1$ if $\pi(s_1|\omega) = 1$ and $\sigma(s_1|\omega) = \varepsilon_2$ if $\pi(s_1|\omega) = 0$ for each $\omega \in \Omega$. Then, $\Sigma_p = (\Delta(B_p))^\Omega$ is the set of feasible signals under p, and it is represented by the shaded square.

omit reference to the subscript p, thus simply writing Σ.

2.2. Equilibrium analysis

After the sender having chosen some signal $\sigma \in \Sigma_p$ and the receiver having heard some message $s \in S$, the receiver forms a posterior belief $\mu_s \in \Delta(\Omega)$ via Bayes rule, viz., for each $\omega \in \Omega$,

$$\mu_s(\omega) = \frac{\mu_0(\omega)\sigma(s|\omega)}{\langle \mu_0, \sigma(s|\cdot) \rangle}, \tag{2}$$

[4]Well-known examples of symmetric channels contain different versions of noisy typewriters (Cover and Thomas, 2006) and different versions of circulant matrices, which constitute a special case of Latin squares (Marshall et al., 2011).

where $\langle \cdot, \cdot \rangle$ denotes the inner product, as usual. Each signal $\sigma \in \Sigma_p$ induces a profile of posteriors $(\mu_1, \ldots, \mu_K)$, where $\mu_k := \mu_{s_k}$ is the posterior given the message s_k. The set of feasible profiles of posteriors is denoted by $\mathcal{M}_p$, and it is a compact subset of $(\Delta(\Omega))^S$. Each posterior $\mu \in \{\mu_1, \ldots, \mu_K\}$ occurs with probability $\langle \mu_0, \sigma(\{s \in S : \mu_s = \mu\}|\cdot)\rangle$, implying that μ_0 belongs to the convex hull of $\{\mu_1, \ldots, \mu_K\}$.

Example 2. Whenever the state space and the channel are both binary, $\mathcal{M}_p$ takes the form that is illustrated in Figure 4 (borrowed from Le Treust and Tomala, 2019, Figure 5). Clearly, it will either

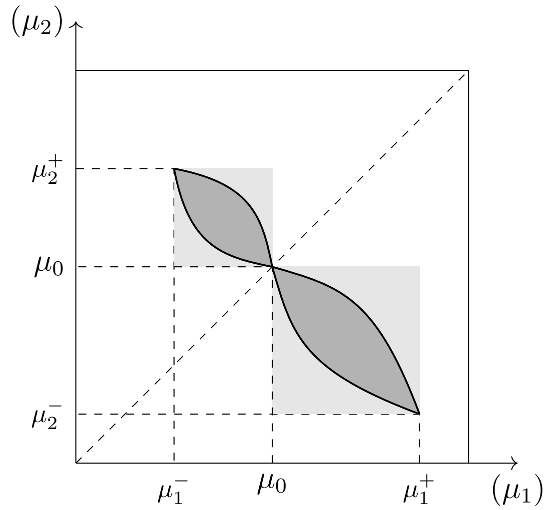


Figure 4: THE SET OF FEASIBLE PROFILES OF POSTERIORS, $\mathcal{M}_p$, for the channel of Figure 2: The posterior μ_1 (resp., μ_2) is obtained when the message s_1 (resp., s_2) is realized. Every signal $\sigma \in \Sigma_p$ leads to a unique pair (μ_1, μ_2) in the shaded area, which is denoted by $\mathcal{M}_p$.

be the case that $\mu_2 \leq \mu_0 \leq \mu_1$ (bottom-right leaf) or $\mu_1 \leq \mu_0 \leq \mu_2$ (top-left leaf). The extreme points at the bottom-right corner

$$(\mu_1^+, \mu_2^-) = \left(\frac{\mu_0(1-\varepsilon_1)}{\mu_0(1-\varepsilon_1) + (1-\mu_0)\varepsilon_2} \, , \, \frac{\mu_0\varepsilon_1}{\mu_0\varepsilon_1 + (1-\mu_0)(1-\varepsilon_2)} \right), \tag{3}$$

and at the top-left corner

$$(\mu_1^-, \mu_2^+) = \left(\frac{\mu_0\varepsilon_2}{\mu_0\varepsilon_2 + (1-\mu_0)(1-\varepsilon_1)} \, , \, \frac{\mu_0(1-\varepsilon_2)}{\mu_0(1-\varepsilon_2) + (1-\mu_0)\varepsilon_1} \right), \tag{4}$$

correspond to the profiles of posteriors induced by the two perfectly informative experiments, i.e., when $(\alpha_1, \alpha_2) = (1, 1)$ and $(\alpha_1, \alpha_2) = (0, 0)$ respectively. Notably, $\mathcal{M}_p$ does not have a product structure. This means that noise does not just restrict the posterior beliefs that can be achieved, but also the way feasible posteriors can be combined with each other, e.g., although the sender can achieve every $\mu_1 \in [\mu_0, \mu_1^+]$ and every $\mu_2 \in [\mu_2^-, \mu_0]$, it is not necessarily the case that she can simultaneously achieve every pair $(\mu_1, \mu_2) \in [\mu_0, \mu_1^+] \times [\mu_2^-, \mu_0]$. As it will become apparent later on, this is exactly the reason why the sender sometimes prefers the channel to be more noisy. $\triangleleft$

Once the receiver has formed some posterior $\mu \in \Delta(\Omega)$, he chooses an action that maximizes his expected utility,

$$u_\mu(a) := \sum_{\omega \in \Omega} \mu(\omega) u(a, \omega).$$

Since u_μ is continuous over the compact set A, a maximum always exists. If there are multiple maxima, the receiver chooses the one that maximizes the sender's expected utility (given μ). If there are multiple sender-preferred maxima, the receiver picks an arbitrary one. We denote the receiver's optimal action, given the posterior μ, by $\hat{a}(\mu)$.

Then, the sender's expected utility from $\sigma \in \Sigma$ (given that she anticipates the receiver to choose optimally) is equal to

$$\hat{v}(\sigma) := \sum_{\omega \in \Omega} \mu_0(\omega) \sum_{s \in S} \sigma(s|\omega) v(\hat{a}(\mu_s), \omega). \tag{5}$$

An *optimal signal* for the sender is one from $\arg\max_{\sigma \in \Sigma} \hat{v}(\sigma)$. We denote the *(sender's) value* of her optimal signal by

$$\hat{v}_p^* := \max_{\sigma \in \Sigma_p} \hat{v}(\sigma). \tag{6}$$

Throughout the paper, we call a channel p trivial whenever the completely uninformative signal is optimal, implying that persuasion attempts are actually useless. We find such cases uninteresting, and thus we only consider non-trivial channels.

Remark 1. It is standard to show that an optimal signal always exists. This follows directly from the sender's expected utility being an upper semi-continuous function over a compact set. $\triangleleft$

The fact that $\mathcal{M}_p$ does not have a product structure (e.g., see Example 2) implies that we cannot use the concavification technique to compute the optimal signal in general. So, let us focus on some special cases of economic interest.[5]

We begin with a binary state space $\Omega = \{\omega_1, \omega_2\}$ and a binary action space $A = \{a_1, a_2\}$, such that $u(a_1, \omega_1) > u(a_2, \omega_1)$ and $u(a_2, \omega_2) > u(a_1, \omega_2)$, where without loss of generality we normalize $u(a_1, \omega_2) = u(a_2, \omega_1) = 0$. That is, the receiver wants to match the state, and therefore he chooses a_1 if and only if his posterior attaches to ω_1 probability larger or equal than the cutoff

$$\bar{\mu} := \frac{u(a_2, \omega_2)}{u(a_1, \omega_1) + u(a_2, \omega_2)},$$

which is obviously larger than μ_0. On the other hand, the sender has state-independent preferences such that $v(a_1, \omega) = 1$ and $v(a_2, \omega) = 0$ for both $\omega \in \Omega$, i.e., she wants to persuade the sender to choose a_1. We refer to this game as the *noisy prosecutor-judge game*, due to its resemblance to the original example of Kamenica and Gentzkow (2011).

Proposition 1. *Consider the noisy prosecutor-judge game together with a strongly symmetric channel with error probability ε and K messages. Then, a signal is optimal, if and only if, there is some $\tilde{s} \in S$ such that $\mu_s = \bar{\mu}$ for all $s \in S \setminus \{\tilde{s}\}$. Moreover, the value of the optimal signal is equal to*

$$\hat{v}_p^* = \frac{\mu_0}{\bar{\mu}} \Big(1 - \frac{\varepsilon}{K-1}\Big). \tag{7}$$

The structure of the optimal signal is similar to the one in Perez-Richet and Skreta (2018), where exactly one message leads to the undesirable action, and all other messages make the receiver indifferent between the two actions. Moreover, note that the sender's value is strictly increasing in K, although the error probability remains constant. In other words, *in the presence of noise, more complex communication channels are strictly beneficial for the sender.* In the context of Example 1, the municipality becomes strictly better off by introducing more colors in the flag system, and

[5]We thank an anonymous referee for suggesting this approach.

eventually it will become possible to overcome the restrictions that stem from swimmers getting possibly confused. The latter is in contrast with the usual setting of Kamenica and Gentzkow (2011), where the complexity of the optimal signal is bounded by the number number of states. We further elaborate on the role of complexity in Section 4.

2.3. Leading example: Persuading the regulator

Let us formalize our leading example from the introduction. Accordingly, a pharmaceutical company (sender) wants to submit an application to the regulator (receiver) for an experimental drug to be approved for commercial use. It is common knowledge that the drug is effective at state ω_1 which occurs with probability $\mu_0 = 0.5$, and it is ineffective at ω_2. An application consists of a clinical trial, which is modelled in its reduced form as an experiment over a binary signalling structure. The error probabilities $\varepsilon_1 = \varepsilon_2 = 0.2$ capture noise in the implementation of the trial. Therefore, we obtain $\mu_1^- = \mu_2^- = 0.2$ and $\mu_1^+ = \mu_2^+ = 0.8$.

Upon receiving some message from $\{s_1, s_2\}$, the regulator announces in a press release an updated probability of the drug being effective, i.e., formally, the regulator's set of actions is $A = [0, 1]$. The regulator's utility function is such that the unique optimal action is to report truthfully. Indeed, his utility is given by $u(a, \omega_1) = -(1-a)^2$ and $u(a, \omega_2) = -a^2$ for each $a \in [0, 1]$, implying that his expected utility $u_\mu(a) = -\mu(1-a)^2 - (1-\mu)a^2$ is maximized at $a = \mu$ for every $\mu \in [0, 1]$.

The company's utility depends solely on the regulator's report, and it is assumed to be increasing in the reported probability, with a jump at 0.8 which is the probability threshold for the drug to be approved. Intuitively, the sender has reputation concerns in the sense that she cares about the reported belief being as high as possible, but at the same time he enjoys some bonus utility if the drug is approved. Thus, we henceforth refer to the posterior that attaches to ω_1 probability larger than 0.5 (resp., smaller than 0.5) as the good posterior (resp., bad posterior).

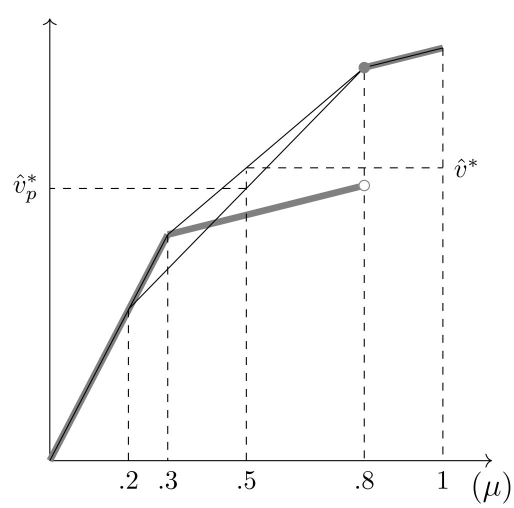


Figure 5: Persuading the regulator: The sender's expected utility is a function of the receiver's posterior beliefs, and it is depicted by the thick grey line. The respective concave closure is depicted by the thin black line. The sender's value without noise is equal to $\hat{v}^*$, whereas his value with noise being given by the channel p is equal to $\hat{v}_p^*$.

Let us first observe that the sender will benefit from a signal only if the good posterior is 0.8. In order to achieve the good posterior of 0.8, it must necessarily be the case that the bad posterior is 0.2 (see Figure 4). Hence, the value of the optimal signal is equal to $\hat{v}_p^*$ (see Figure 5).

Following the analysis of Kamenica and Gentzkow (2011), the optimal signal in the noiseless case is the one that combines the good posterior 0.8 with the bad posterior 0.3, thus yielding value $\hat{v}^* > \hat{v}_p^*$. That is, ideally the sender wants to increase the probability of the drug's success under the bad posterior, without trading off approval of the drug under the good posterior. Interestingly, the presence of noise leads to a more informative optimal signal. Thus, given that the receiver's expected utility function is convex, more noise turns out to be beneficial for the receiver. This conclusion is similar to the one of Blume et al. (2007) for the noisy cheap-talk game, although in their case the analysis is different.

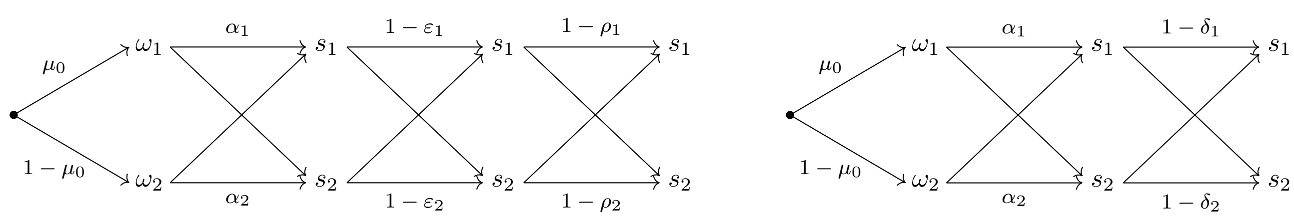
3. Monotonicity

Let us now turn to the main research question of the paper, viz., *is more noise always harmful or are there cases where it is beneficial for the sender?* We study this question both for binary and for some more general signalling structures.

In order to tackle this question in a systematic way, we first recall Blackwell's informativeness relation over the set of noisy channels (Blackwell, 1951, 1953). We say that q is a *garbling* of p (viz., q is *more noisy* than p) whenever there is a channel $r : S \to \Delta(S)$ such that

$$q(t|s) = \sum_{u \in S} p(u|s) r(t|u) \tag{8}$$

for each $s, t \in S$. In this case we write $p \succeq q$ and $q = p \circ r$. Intuitively, a garbling is obtained by adding another channel to the right of the original channel, e.g., a garbling of the channel p (from Figure 2) is illustrated below (in Figure 6).



(a) There exists a second channel r that further distorts the message, after the original distortion induced by the channel p. The probabilities to have such additional distortions are $\rho_1 := r(s_2|s_1)$ and $\rho_2 := r(s_1|s_2)$.

(b) The garbled channel q is obtained by combining the two channels, p and r, thus obtaining error probabilities $\delta_1 := (1-\varepsilon_1)\rho_1 + \varepsilon_1(1-\rho_2)$ and $\delta_2 := \varepsilon_2(1-\rho_1) + (1-\varepsilon_2)\rho_2$.

Figure 6: Garbling with a binary state space: The channel q is more noisy than the channel p.

Then, our question is formalized as follows: *does $p \succeq q$ always imply $\hat{v}_p^* \geq \hat{v}_q^*$?* Our intuition says that most probably this will have to be the case. In the most obvious special case, where p is noiseless, noise is trivially harmful for the sender, as $\Sigma_q \subseteq \Pi = \Sigma_p$. Moreover, when the sender and receiver have aligned preferences, it is again quite clear that more noise is harmful for the sender. This follows directly from Blackwell's well-known theorem (Blackwell, 1951, 1953). In particular, since the receiver's optimal expected utility is convex on $\Delta(\Omega)$, so will be the sender's expected utility. Then, by the fact that the posteriors under the more informative channel p are more dispersed than the posteriors under the less informative channel q, the value of the sender decreases as we add more noise. However, it turns out that our initial intuition is not correct in general. Namely, the sender's value is not always increasing in the channel's informativeness, i.e., more noise may be beneficial for the sender.

Leading example (continued). Recall the example from Section 2.3, where $\varepsilon_1 = \varepsilon_2 = 0.2$, and suppose that $0 < \rho_1 < 0.5$ and $\rho_2 = 0$. Then, the garbled channel q will be such that $\delta_1 = 0.2 + 0.8\rho_1$ and $\delta_2 = 0.2(1 - \rho_1)$. For instance, if $\rho_1 = 0.2$, garbling p with r leads the set of feasible profiles of posteriors to change from the points in the black graph to those in the gray graph in Figure 7. This means that we can now achieve a larger bad posterior (viz., $\mu_2 = 0.3$ instead of $\mu_2 = 0.2$) without reducing the good posterior (viz., $\mu_1 = 0.8$). As a result, we obtain $\hat{v}_q^* > \hat{v}_p^*$, thus violating monotonicity.

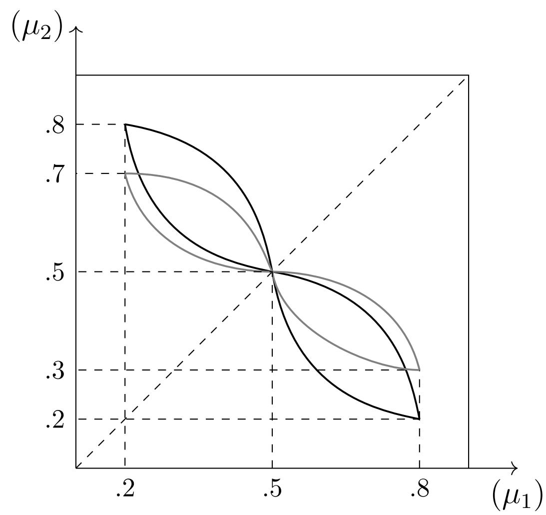


Figure 7: The effect of garbling on the set of feasible profiles of posteriors: the black graph corresponds to the profile of posteriors under p, whereas the gray graph corresponds to the profiles of posteriors under q.

Intuitively, r further distorts our data (besides the distortion due to p) only if the bad message is realized, but not if the good message is realized. This is convenient for the sender, as additional data distortions occur only in the bad case (which can only get better), but not in the good case. Overall, more noise leads to a less informative signal, which makes the sender better off and the receiver worse off. $\triangleleft$

Remark 2. The fact that in the previous example, more noise leads to a less informative optimal signal is mere coincidence. For instance, in the noisy judge-prosecutor game, the optimal signal under q would have been more informative than the one under p, as the sender would have been able to combine μ_2^* with some $\mu_1 < \mu_1^*$, and both agents would have been better off. In general, as we have already mentioned, the receiver's expected utility is a convex function, and therefore it is increasing in the informativeness of the optimal signal, which in turn depends not only on the channel but also on the underlying preferences of the two agents. $\triangleleft$

The question then becomes: *can we identify the garblings that always make the sender worse off, irrespective of the preference profile and the prior?* For now, we focus on binary channels in order to be able to provide a full characterization, which will, in turn, allow us to get good intuition. Later on, we provide extensions to channels of arbitrary cardinality. Let us stress that in all cases the state space is of an arbitrary finite cardinality.

Proposition 2. *Let p and q be binary channels such that $q = p \circ r$, as in Figure 6. Then, $\hat{v}_p^* \geq \hat{v}_q^*$ for all pairs of utility functions and all priors, if and only if, the following condition holds:*

$$\frac{\varepsilon_1}{1 - \varepsilon_1} \leq \frac{\rho_1}{\rho_2} \leq \frac{1 - \varepsilon_2}{\varepsilon_2}. \tag{9}$$

There are two ways to read the previous result. According to our first interpretation, if we start with a channel p and we mix it with another channel r, the resulting (more noisy) channel $q = p \circ r$ will always make the sender worse off whenever r is "sufficiently symmetric", i.e., if ρ_1 and ρ_2 are sufficiently close to each other, in which case ρ_1/ρ_2 is sufficiently close to 1. The idea is that, in this case, q will be more noisy than p, but nonetheless "similar" in terms of the underlying structure. In Figure 7, this would correspond to the gray leaves being contained in the black leaves, i.e., the black leaves shrink inwards, without rotating much around (μ_0, μ_0). Conversely, if ρ_1 is sufficiently far from ρ_2, the leaves will rotate, thus leading to new combinations of posterior beliefs. In this case, there will always exist utility functions that make the sender prefer q over p.

According to the second interpretation, if we start with a very noisy channel p then it becomes easier to break monotonicity, i.e., if the error probabilities ε_1 and ε_2 are large in the first place (close to 0.5), then there are many channels r that will violate condition (9), whereas if we start with a channel that does not induce a lot of noise (i.e., both error probabilities are close to 0), fewer channels r lead to such violations. The idea here is that, if the receiver has little trust in his source in the first place, there are more instances of even more mistrust where the sender is better off. In Figure 7, this would be illustrated by black leaves that are small and thin in the first place, and therefore it would be easier for the gray leaves to rotate slightly outside the black ones.

So far, our monotonicity analysis has focused on binary channels. Do our conclusions carry on to more general channels? Let us first stress that obtaining a full characterization result by means of a simple formula à la condition (9) would be too ambitious given the complexity of the problem. Hence, we are going to establish sufficient conditions for some special cases, along the lines of the first interpretation that we provided above. Namely, we will show that within classes of similar channels, more noise always makes the sender worse off.

We first focus on the class of symmetric channels. We show that between two Blackwell-comparable symmetric channels, the sender is always better off with the more informative one.

Proposition 3. *Let p and q be symmetric channels such that $p \succeq q$. Then, $\hat{v}_p^* \geq \hat{v}_q^*$ for all pairs of utility functions and all priors.*

We now focus on cases where the channel r is strongly symmetric. This is a generalization of condition (9) when $\rho_1 = \rho_2$ to any two channels p and q. Again it turns out that such garblings always make the sender worse off in equilibrium.

Proposition 4. *Let r be a strongly symmetric channel such that $q = p \circ r$. Then, $\hat{v}_p^* \geq \hat{v}_q^*$ for all pairs of utility functions and all priors.*

Both previous results rely on the fact that the set of signals shrinks (i.e., $\Sigma_q \subseteq \Sigma_p$) and as a result the sender maximizes his expected utility function over a smaller set, thus leading to $\hat{v}_p^* \geq \hat{v}_q^*$ irrespective of the prior and the players' preferences. Let us briefly explain why we obtain $\Sigma_q \subseteq \Sigma_p$ in each of the previous results.

Both results make use of the matrix representation of a noisy channel. In particular, p is typically represented by a stochastic transition matrix P such that $P_{k,\ell} := p(s_\ell|s_k)$. Each row of P is an element of $B_p = \{p(\cdot|s_1), \ldots, p(\cdot|s_K)\} \subseteq \Delta(S)$, which identifies Σ_p. It is easy to see that the garbling condition $q = p \circ r$ is equivalent in matrix notation to $Q = PR$.

In Proposition 3, we show that every column of Q is a convex combination of the columns of P, implying by symmetry that every row of Q is a convex combination of the rows of P. Hence, $\Delta(B_q) \subseteq \Delta(B_p)$, which in turn implies $\Sigma_q \subseteq \Sigma_p$.

In Proposition 4, we show that garbling p with a strongly symmetric channel r leads to shrinking of $\Delta(B_p)$. The underlying idea is that each $q(\cdot|s)$ will lie on the linear segment that connects $p(\cdot|s)$

with the center of $\Delta(S)$. Then, given that the center lies in the interior of $\Delta(B_p)$, every $q(\cdot|s)$ lies inside $\Delta(B_p)$. Therefore, $\Delta(B_q) \subseteq \Delta(B_p)$, and consequently $\Sigma_q \subseteq \Sigma_p$.

4. Complexity

As it is well-known, in standard persuasion games the sender can always achieve her value with a given number of messages (Kamenica and Gentzkow, 2011). In other words, there exists an upper bound on the complexity of the signal that the sender needs to use in order to maximize her expected utility, and this maximum complexity depends merely on the complexity of the state space. However, as Proposition 1 suggests, in the usual judge-prosecutor game, if the experiment is distorted by a strongly symmetric channel, a more complex message space always benefits the sender. In this section, we generalize this insight to every underlying persuasion game and many different ways of increasing the complexity of a channel.[6]

To do so, we first need to be precise on how the complexity of the channel can increase. First of all, it is clearly the case that if we increase the complexity of the channel by adding messages – or groups of messages – that do not interact with each other, we will eventually trivially get a noiseless channel. For instance, if next to the messages {blue, red} we add the messages {good, bad}, it will most likely be the case that the former will not be confused with the later, and as a consequence we will be able to achieve all signals over a binary message space. So let us focus on cases where we can only add messages that interact with each other. In principle, this can be done in many different ways. Let us illustrate this point by means of our earlier example.

Example 1 (continued). Suppose that we begin with a binary strongly symmetric channel p, viz., there are two message {blue, red} with the probability of not confusing colors being $p(\text{blue}|\text{blue}) = p(\text{red}|\text{red}) = 0.9$. Now, consider the following two ways of replacing p with a more complex channel:

- The municipality introduces additional distinct colors, thus obtaining an expanded message space {blue, red, yellow, green}. Nevertheless, the probability of not confusing colors remains the same, i.e., $q(\text{blue}|\text{blue}) = q(\text{red}|\text{red}) = q(\text{yellow}|\text{yellow}) = q(\text{green}|\text{green}) = 0.9$.

- The municipality introduces different shades for some of the existing colors, thus obtaining the message space {blue, dark red, light red}. The total probability of not confusing colors remains the same, i.e., $\bar{p}(\{\text{dark red, light red}\}|\text{dark red}) = \bar{p}(\{\text{dark red, light red}\}|\text{light red}) = 0.9$ and $\bar{p}(\text{blue}|\text{blue}) = 0.9$. At the same time, different shades of the same color are almost indistinguishable, i.e., $\bar{p}(\text{light red}|\text{dark red}) = \bar{p}(\text{dark red}|\text{light red}) = 0.4$ and $\bar{p}(\text{dark red}|\text{dark red}) = \bar{p}(\text{light red}|\text{light red}) = 0.5$.

Notice that both ways of expanding the message space keep certain elements of the structure of p, e.g., in both cases the probability of not confusing colors remains the same. ◁

Let, us now introduce a general procedure to make a channel more complex. First, take each message $s_k \in S$ and create an arbitrary number of duplicates, thus obtaining an entire set S_k of duplicate messages with typical element $\bar{s}_k$. Note that not all messages in S have necessarily the same number of duplicates. Thus, we obtain an enlarged set of messages $\bar{S} := S_1 \cup \cdots \cup S_k$. Then, we define the channel $\bar{p} : \bar{S} \to \Delta(\bar{S})$ as follows. For every pair of messages $\bar{s}_k \in S_k$ and $\bar{s}_\ell \in S_\ell$:

(a) If $\bar{s}_k$ and $\bar{s}_\ell$ are not duplicates of the same original message, i.e., if $k \neq \ell$, then

$$\bar{p}(\bar{s}_\ell|\bar{s}_k) = \frac{p(s_\ell|s_k)}{|S_\ell|}.$$

[6] Again, we would like thank an anonymous referee for raising this interesting question

That is, each duplicate of s_k distributes uniformly across the duplicates of s_ℓ the error probability that s_k assigned to s_ℓ under the original channel.

(b) If $\bar{s}_k$ and $\bar{s}_\ell$ are duplicates of the same original message, i.e., if $k = \ell$, then

$$\bar{p}(\bar{s}_\ell|\bar{s}_k) = \bar{p}(\bar{s}_k|\bar{s}_\ell).$$

That is, the probability $p(s_k|s_k)$ of correctly observing s_k in the original channel, is distributed in a symmetric way within S_k.

Whenever the previous two conditions are satisfied we write $\bar{p} \trianglerighteq p$.

Example 1 (continued). Recall the second way of expanding channel p, which was done by introducing different shades. We started with the messages {blue, red}, which we expanded to {blue, dark red, light red}, i.e., the only duplicate of blue is blue itself, while the duplicates of red are dark red and light red.

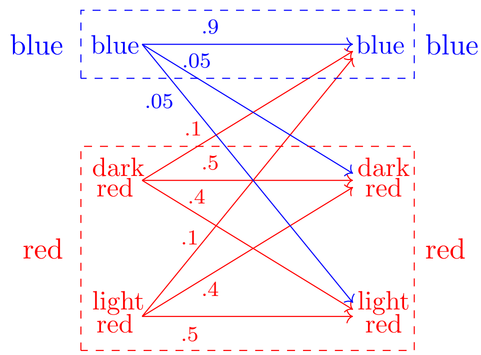


Figure 8: Increasing the complexity of the binary symmetric channel by introducing different shades.

Notice that the total error probability of one set of duplicates agrees with the original probabilities, e.g., $\bar{p}(\{\text{dark red, light red}\}|\text{blue}) = p(\text{red}|\text{blue}) = 0.1$. Moreover, conditional on blue each shade of red receives the same probability, e.g., $\bar{p}(\text{dark red}|\text{blue}) = \bar{p}(\text{light red}|\text{blue}) = 0.05$. And finally, within the red shades, noise is symmetric, e.g., $\bar{p}(\text{light red}|\text{dark red}) = \bar{p}(\text{dark red}|\text{light red}) = 0.4$. ◁

Obviously, there are multiple channels $\bar{p}$ that we can obtain from p using the previous procedure, which differ in the number of duplicates we introduce for each of the original messages as well as on the probability of correctly observing each duplicate message. But as the following result shows, this way will always be beneficial for the sender.

Proposition 5. *If $\bar{p} \trianglerighteq p$, then $\hat{v}^*_{\bar{p}} \geq \hat{v}^*_p$ for all pairs of utility functions and all priors.*

Let us briefly sketch the proof of the previous result in the context of our earlier example. For any experiment over the message space {blue, red}, consider the experiment over the expanded message space {blue, dark red, light red} such that the total conditional probability of red is distributed uniformly across dark red and light red, while the conditional probability of blue remains the same (given each state). Then, we show that the two experiments (when distorted by the corresponding channels) yield the same distribution of posteriors. Hence, whatever expected utility the sender can achieve under p, she will also be able to achieve under $\bar{p}$.

Remark 3. Of course, quite likely, this inequality will be strict, i.e., the sender will typically strictly benefit from expanding the message space in such a way. This is because under the more complex channel $\bar{p}$ there will always exist distributions of posteriors that cannot be achieved under p. Hence, there will always be preference profiles that will make the sender strictly better off under $\bar{p}$ compared to the situation under p (similarly to Proposition 2). In fact, the class of such games is quite rich. ◁

Now, if we combine the previous result with Proposition 4, we can essentially generalize the conclusion of Proposition 1 to any persuasion game.

Corollary 1. *Let $p : S \to \Delta(S)$ and $q : \bar{S} \to \Delta(\bar{S})$ be two strongly symmetric channels with $|\bar{S}| = M \cdot |S|$ for some integer $M > 1$. Moreover, we assume that the corresponding total error probabilities are denoted by $\varepsilon := p(s'|s)$ and $\delta := q(\bar{s}'|\bar{s})$ for any two distinct $s, s' \in S$ and any two distinct $\bar{s}, \bar{s}' \in \bar{S}$. Then, if the error probabilities satisfy*

$$\frac{|S|}{|S|-1}\varepsilon \geq \frac{|\bar{S}|}{|\bar{S}|-1}\delta,$$

it will be the case that $\hat{v}_q^* \geq \hat{v}_p^*$ for all pairs of utility functions and all priors.

The underlying idea is that q is more beneficial to the sender due to the fact that it has more available messages, even in cases where it has strictly larger total error probability. In fact, we identify the trade-off between the channel complexity (measured by the cardinality of the message space) and the size of the total error probability, i.e., multiplying the number of messages by M allows us to decrease the total error probability by a factor of $\frac{|S|-1}{|S|} \cdot \frac{|\bar{S}|}{|\bar{S}|-1}$, without making the sender worse off. For instance, if p and q are strongly symmetric channels with 2 and 4 messages respectively, q will make the sender better off whenever the error $\delta \leq \frac{3}{2}\varepsilon$, i.e., even in cases where δ is larger than ε. Of course, a direct consequence is that, if we increase the number of messages by some factor while maintaining the total error probability fixed (i.e., while taking $\varepsilon = \delta$), the sender does become better off. Let us illustrate this last point in our working example.

Example 1 (continued). Recall the first way of expanding a channel, which was done by introducing new colors. In particular, we begin with the strongly symmetric channel p over {blue, red} with probability of not confusing colors equal to 0.9. We want to compare this channel (from the sender's point of view) with a strongly symmetric channel q over {blue, red, yellow, green}, with the same probability of 0.9 of not confusing colors. We do this into steps. First, by Proposition 5, it follows that p is worse for the sender compared to the (strongly symmetric) channel of Figure 9. Now, observe that this is a strongly symmetric channel with 4 messages and probability of seeing

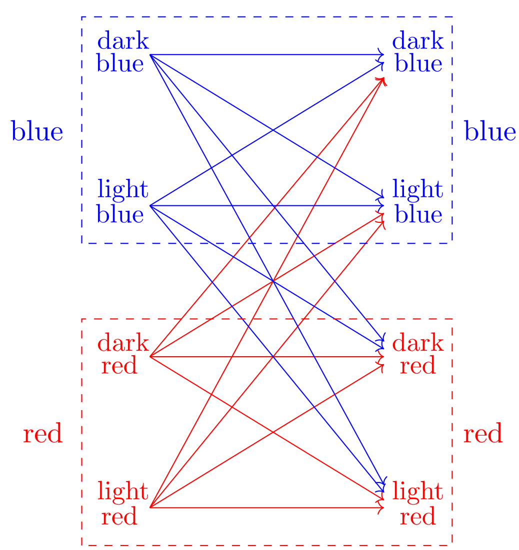


Figure 9: Increasing the complexity of the binary symmetric channel by introducing new colors: All errors occur with probability 0.05, while the probability of seeing the correct message is equal to 0.85.

the correct message equal to 0.85. This is clearly less desired by the sender compared to q, which is also a strongly symmetric channel with 4 messages but higher probability of not confusing colors. ◁

Finally, notice that this last comparison (between two strongly symmetric channels that differ only in the error probability) will often make the sender to strictly prefer the less noisy one. This is because the distributions of posteriors that can be achieved with the more noisy channel form a strict subset of the set of distributions that can be achieved with the less noisy one. Hence, quite often, the sender becomes strictly better off by increasing the complexity of the channel.

5. Discussion

5.1. Endogenous noise

An interesting extension of our model is to consider experiment-dependent noisy channels. Consider the following example.[7] There are two different species of blue birds and one species of red birds. A company commissions an environmental report on the numbers of the different species, which is necessary in order for a proposed project to be approved by the corresponding agency. It is probably easier to make mistakes when trying to distinguish two blue birds, as opposed to when trying to distinguish a blue from a red bird. Here, it seems natural to assume that the noisy channel depends on the experiment that the sender has chosen.

Such an extension would be relevant for various applications, which would be certainly interesting to study in follow-up papers. It is important to stress that the way noise is endogenized often depends on its source. For instance, noise can be increasing in the informativeness of the experiment when it captures mistakes in the implementation of the experiment, i.e., more informative experiments require larger and more complex datasets, and are therefore more prone to mistakes. On the other hand, noise can also be decreasing in the informativeness of the experiment if it captures the communication errors or mistakes due to limited understanding, i.e., data leading to more clear-cut conclusions are misunderstood less often. Along these lines, related is also the work of Kosenko (2018) on Bayesian persuasion with mediators, where noise becomes endogenous through the preferences of the mediator.

5.2. Correlation in message realization

In noisy persuasion games with multiple receivers, the optimal signal often depends on whether the realizations of the noisy channel are independent or correlated across receivers. For instance, in a voting environment à la Alonso and Câmara (2016), the politician chooses an experiment π, then a message $s \in S$ is drawn from $\pi(\cdot|\omega)$, and finally each voter observes the distorted message $t \in S$ which is drawn from the error distribution $p(\cdot|s)$. If we then assume that all voters observe the same distorted message (i.e., the errors are perfectly correlated across voters), noise can capture mistakes in the politician's campaign or in the media coverage, which are perceived symmetrically across voters. If on the other hand, we wanted to model mistakes due to the voters' limited ability to understand the message, or due to the fact that the voters do not fully trust the message they hear, we would assume that a different distorted message t is drawn from $p(\cdot|s)$ independently for each voter. And of course, one could study intermediate cases with different correlation structures. Overall, it is not just the model of the distortions but also their source that affects the optimal signal. Understanding the role of correlation of distortions across receivers is an interesting problem for future research.

[7] We are greatly indebted to an anonymous referee for suggesting this extension and the corresponding example.

A. Proofs

Proof of Proposition 1. Fix an optimal experiment π, writing for simplicity $x_k := \pi(s_k|\omega_1)$ and $y_k := \pi(s_k|\omega_2)$, and additionally $\alpha_k := (1-\varepsilon)x_k + \frac{\varepsilon}{K-1}(1-x_k)$ and $\beta_k := (1-\varepsilon)y_k + \frac{\varepsilon}{K-1}(1-y_k)$. Note that the posterior belief given s_k will be equal to

$$\mu_k := \frac{\mu_0 \alpha_k}{\mu_0 \alpha_k + (1-\mu_0)\beta_k}.$$

Without loss of generality, we assume that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$, with at least one inequality being strict.

Step 1. Let us first prove that $\mu_1 = \bar{\mu}$. We proceed by contradiction, assuming that $\mu_1 > \bar{\mu}$. Since $\mu_1 > \mu_0$, it will be the case that $\mu_0 > \mu_K$, implying that $y_K > 0$. Now, for any $\lambda \in (0, y_K)$, define the new experiment π_λ which is exactly the same as π except for the fact that $y_1^\lambda := y_1 + \lambda$ and $y_K^\lambda := y_K - \lambda$. The new posteriors that we will obtain (for the respective messages) will be $\mu_1^\lambda, \ldots, \mu_K^\lambda$. Obviously, we have $\mu_k^\lambda = \mu_k$ for all $k \in \{2, \ldots, K-1\}$. Moreover, note that μ_1^λ is continuously decreasing in λ, and likewise μ_K^λ is continuously increasing in λ. So, we can select some λ small enough such that $\mu_K < \bar{\mu} \leq \mu_1^\lambda < \mu_1$. Hence, the receiver chooses a_1 under π, if and only if, he chooses a_1 under π_λ. Then, letting $M := \max\{k = 1, \ldots, K | \mu_k \geq \bar{\mu}\}$, the expected probability of the receiver choosing a_1 (and a fortiori the sender's expected utility) is equal to

$$\mu_0 \sum_{k=1}^{M} \alpha_k + (1-\mu_0) \sum_{k=1}^{M} \beta_k$$

under the experiment π, and equal to

$$(1-\mu_0)\Big(1 - \varepsilon - \frac{\varepsilon}{K-1}\Big)\lambda + \mu_0 \sum_{k=1}^{M} \alpha_k + (1-\mu_0) \sum_{k=1}^{M} \beta_k$$

under the experiment π_λ. Note that $\varepsilon + \frac{\varepsilon}{K-1} < 2\varepsilon < 1$, implying that π_λ is better than π for the sender, thus reaching a contradiction.

Step 2. By the previous step, we have

$$\beta_k = \frac{\mu_0(1-\bar{\mu})}{(1-\mu_0)\bar{\mu}} \alpha_k$$

for every $k = 1, \ldots, M$. Thus, the sender's expected utility is equal to

$$\frac{\mu_0}{\bar{\mu}} \sum_{k=1}^{M} \alpha_k = \frac{\mu_0}{\bar{\mu}} \Big(1 - \varepsilon - \frac{\varepsilon}{K-1}\Big) \sum_{k=1}^{M} x_k + \frac{\mu_0}{\bar{\mu}} \frac{\varepsilon}{K-1} M.$$

This is strictly increasing in $\sum_{k=1}^{M} x_k$ and in M. Hence, it must be the case that $\sum_{k=1}^{M} x_k = 1$ and $M = K-1$, implying that the value of the optimal signal is

$$\hat{v}_p^* = \frac{\mu_0}{\bar{\mu}} \Big(1 - \frac{\varepsilon}{K-1}\Big),$$

which completes the proof. □

Proof of Proposition 2. Recall that the state space is $\Omega = \{\omega_1, \ldots, \omega_N\}$. Thus, we can identify each experiment $\pi \in \Pi$ with the vector $(\pi_1, \ldots, \pi_N) \in [0,1]^N$, where $\pi_n := \pi(s_1|\omega_n)$ for every $n = 1, \ldots, N$. Now, consider a binary channel p with error probabilities $\varepsilon_1 := p(s_2|s_1)$ and $\varepsilon_2 := p(s_1|s_2)$. Similarly to an experiment, each signal $\sigma \in \Sigma_p$ is identified by the vector $(\sigma_1, \ldots, \sigma_N) \in [0,1]^N$, where $\sigma_n := \sigma(s_1|\omega_n)$ for every $n = 1, \ldots, N$. So, the set of feasible signals is

$$\Sigma_p := \Big\{ (\sigma_1, \ldots, \sigma_N) \in [0,1]^N \; : \; \text{there is a vector } (\pi_1, \ldots, \pi_N) \in [0,1]^N \text{ such that } \sigma_n = \pi_n(1-\varepsilon_1) + (1-\pi_n)\varepsilon_2 \text{ for every } n = 1, \ldots, N \Big\} .$$

The latter can be rewritten as

$$\begin{aligned} \Sigma_p &= \left\{ (\sigma_1, \ldots, \sigma_N) \in [0,1]^N : 0 \leq \frac{\sigma_n - \varepsilon_2}{1 - \varepsilon_1 - \varepsilon_2} \leq 1 \right\} \\ &= \left\{ (\sigma_1, \ldots, \sigma_N) \in [0,1]^N : \varepsilon_2 \leq \sigma_n \leq 1 - \varepsilon_1 \right\}, \end{aligned}$$

which is obviously nonempty, as $\varepsilon_1 < 1/2$ and $\varepsilon_2 < 1/2$.

Now consider a garbling $q = p \circ r$, where $\delta_1 := q(s_2|s_1)$ and $\delta_2 := q(s_1|s_2)$ are the error probabilities of q, while $\rho_1 := r(s_2|s_1)$ and $\rho_2 := r(s_1|s_2)$ are the error probabilities of r. Then, it is easy to see that the following equivalences hold

$$\begin{aligned} \Sigma_q \subseteq \Sigma_p &\Leftrightarrow \varepsilon_1 \leq \delta_1 \text{ and } \varepsilon_2 \leq \delta_2 \\ &\Leftrightarrow \frac{\varepsilon_1}{1-\varepsilon_1} \leq \frac{\rho_1}{\rho_2} \leq \frac{1-\varepsilon_2}{\varepsilon_2}, \end{aligned} \tag{A.1}$$

where the last pair of inequalities is our condition (9).

Sufficiency: It follows directly from (A.1) combined with the fact that $\Sigma_q \subseteq \Sigma_p$ implies $\hat{v}_p^* \geq \hat{v}_q^*$.

Necessity: Suppose that condition (9) does not hold, implying by (A.1) that either $\delta_1 < \varepsilon_1$ or $\delta_2 < \varepsilon_2$. For the time being, consider a binary state space $\Omega = \{\omega_1, \omega_2\}$ with prior $\mu_0 = 1/2$. Then, under the more noisy channel q, if the sender chooses the signal $\sigma' \in \Sigma_q$ with $\sigma'_1 = 1 - \delta_1$ and $\sigma'_2 = \delta_2$, the receiver's profile of posterior beliefs will become

$$(\mu'_1, \mu'_2) = \left(\frac{1-\delta_1}{1-\delta_1+\delta_2} \; , \; \frac{\delta_1}{1+\delta_1-\delta_2} \right) \in \mathcal{M}_q. \tag{A.2}$$

Likewise, if she chooses the signal $\sigma'' \in \Sigma_q$ with $\sigma''_1 = \delta_1$ and $\sigma''_2 = 1 - \delta_2$, the receiver's profile of posterior beliefs will become

$$(\mu''_1, \mu''_2) = \left(\frac{\delta_1}{1+\delta_1-\delta_2} \; , \; \frac{1-\delta_1}{1-\delta_1+\delta_2} \right) \in \mathcal{M}_q, \tag{A.3}$$

noticing that obviously $(\mu'_1, \mu'_2) = (\mu''_2, \mu''_1) = (\mu^-, \mu^+)$. In order for $\{(\mu'_1, \mu'_2), (\mu''_1, \mu''_2)\} \cap \mathcal{M}_p \neq \emptyset$, there must exist some $\sigma = (\sigma_1, \sigma_2) \in \Sigma_p$ such that

$$\left(\frac{\sigma_1}{1+\sigma_1-\sigma_2} \; , \; \frac{1-\sigma_1}{1-\sigma_1+\sigma_2} \right) \in \{ \, (\mu'_1, \mu'_2), (\mu''_1, \mu''_2) \, \} \, . \tag{A.4}$$

Simple algebra yields that $(\mu'_1, \mu'_2) = \left(\frac{\sigma_1}{1+\sigma_1-\sigma_2} \; , \; \frac{1-\sigma_1}{1-\sigma_1+\sigma_2} \right)$ holds if and only if the system

$$\begin{bmatrix} \delta_2 & 1-\delta_1 \\ 1-\delta_2 & \delta_1 \end{bmatrix} \cdot \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} = \begin{bmatrix} 1-\delta_1 \\ 1-\delta_2 \end{bmatrix} \tag{A.5}$$

has a solution in Σ_p. Note that the determinant of the coefficient matrix is equal to $\delta_1 + \delta_2 - 1 \neq 0$, implying that the system has a unique solution. But, then by construction this solution is $(1-\delta_1, 1-\delta_2)$ which does not belong to Σ_p. Likewise, $(\mu_1'', \mu_2'') = \big(\frac{\sigma_1}{1+\sigma_1-\sigma_2}\ ,\ \frac{1-\sigma_1}{1-\sigma_1+\sigma_2}\big)$ holds if and only if

$$\begin{bmatrix} 1-\delta_2 & \delta_1 \\ \delta_2 & 1-\delta_1 \end{bmatrix} \cdot \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \tag{A.6}$$

has a solution in Σ_p. The determinant of the coefficient matrix is equal to $1 - \delta_1 + \delta_2 \neq 0$, implying that once again the system has a unique solution. And again by construction this solution is (δ_1, δ_2) which does not belong to Σ_p. Now, take utility functions such that the receiver's unique optimal choice is to choose the action $A = [0,1]$ that matches his posterior belief, while the sender prefers the receiver to choose either μ^- or μ^+, i.e., formally, let $v(\mu^-, \omega) = v(\mu^+, \omega) = 1$ for both $\omega \in \Omega$, while $v(a, \omega) = 0$ for all other $a \in A$ and all $\omega \in \Omega$. This implies that $\hat{v}_q^* = 1 > \hat{v}_p^*$.

Finally, consider the case of a finite state space $\tilde{\Omega} = \{\tilde{\omega}_1, \ldots, \tilde{\omega}_N\}$ and let $\mathcal{E} = \{E_1, E_2\}$ be a partition of $\tilde{\Omega}$ together with a prior such that $\mu_0(E_1) = \mu_0(E_2) = 1/2$. Then, suppose that both agents have $\mathcal{E}$-measurable utility functions. In particular, the utility of each agent at each $\tilde{\omega} \in E_1$ is the same as the utility at ω_1 in the binary case above, and likewise for every $\tilde{\omega} \in E_2$ and ω_2. This implies that the analysis will be identical to the one of the binary case, and therefore once again we will obtain $\hat{v}_q^* = 1 > \hat{v}_p^*$, thus completing the proof. □

Lemma A1. *Take two channels p and q such that every row of Q can be written as a convex combination of the rows of P. Then, $\Sigma_q \subseteq \Sigma_p$.*

Proof. If every row of Q can be written as a convex combination of the rows of P, it will be the case that $B_q \subseteq \Delta(B_p)$. Therefore, $\Sigma_q \subseteq \Sigma_p$. □

Lemma A2. *Let P and Q be two doubly stochastic matrices such that P is nonsingular. Furthermore, let R be some stochastic matrix such that $Q = PR$. Then, R is doubly stochastic.*

Proof. Since P is nonsingular, there exists a square (inverse) matrix B such that $PB = BP = I$, where I is the identity matrix. By $BP = I$ it follows that $\sum_{k=1}^K B_{n,k} P_{k,n} = 1$ and also $\sum_{k=1}^K B_{n,k} P_{k,m} = 0$ for all $m \neq n$. Thus, using the fact that P is doubly stochastic, we obtain

$$1 = \sum_{m=1}^K \sum_{k=1}^K B_{n,k} P_{k,m} = \sum_{k=1}^K B_{n,k}. \tag{A.7}$$

Now, multiply both sides of $Q = PR$ with B (from the left) to obtain $R = BQ$, thus implying

$$R_{n,m} = \sum_{k=1}^K B_{n,k} Q_{k,m}. \tag{A.8}$$

Since R is by hypothesis (row) stochastic, it suffices to prove that the entries of each column sum up to 1. Indeed,

$$\sum_{n=1}^K R_{n,m} = \sum_{n=1}^K \sum_{k=1}^K B_{n,k} Q_{k,m} = \sum_{k=1}^K Q_{k,m} = 1, \tag{A.9}$$

with (A.9) following from (A.7) and the fact that Q is doubly stochastic. □

Proof of Proposition 3. By symmetry both P and Q are doubly stochastic. Moreover, by $p(s|s) > 1/2$ it follows that P is diagonally dominant, and therefore it is nonsingular (Levy-Desplanques Theorem). Hence, by Lemma A2, it follows that R is doubly stochastic too. Therefore, every column of Q can be written as a convex combination of columns of P. But then, since P and Q are symmetric, every column vector is also a row vector in each of them. Hence, every row of Q can also be written as a convex combination of rows of P. Therefore, by Lemma A1 we obtain $\Sigma_q \subseteq \Sigma_p$. Thus, $\hat{v}_q^* \leq \hat{v}_p^*$. $\square$

Proof of Proposition 4. Since r is strongly symmetric, we obtain $r(s|s) = 1 - \rho$ and $r(t|s) = \frac{\rho}{K-1}$ for every $t \neq s$. Therefore, for every $s \in S$, we obtain

$$q(\cdot|s) = (1-\rho)p(\cdot|s) + \frac{\rho}{K-1}(1 - p(\cdot|s)). \tag{A.10}$$

Hence, $q(\cdot|s)$ lies on the straight line that connects $p(\cdot|s)$ and $(\frac{1}{K}, \ldots, \frac{1}{K})$. But then, by $p(s|s) > 1/2$ for all $s \in S$ it follows that $(\frac{1}{K}, \ldots, \frac{1}{K})$ belongs to the interior of $\Delta(\{p(\cdot|s_1), \ldots, p(\cdot|s_K)\})$. In other words, all rows of Q can be written as convex combinations of the rows of P. Hence, by Lemma A1, we obtain $\Sigma_q \subseteq \Sigma_p$, and therefore $\hat{v}_p^* \geq \hat{v}_q^*$. $\square$

Proof of Proposition 5. For any fixed experiment $\pi : \Omega \to \Delta(S)$, take the experiment $\bar{\pi} : \Omega \to \Delta(\bar{S})$ such that

$$\bar{\pi}(\bar{s}_k|\omega) := \frac{\pi(s_k|\omega)}{|S_k|}$$

for each $\bar{s}_k \in S_k$, each $k = 1, \ldots, K$ and each $\omega \in \Omega$. Then, we obtain

$$\begin{aligned}
\bar{\sigma}(\bar{s}_k|\omega) &= \sum_{\ell=1}^{K} \sum_{\bar{s}_\ell \in S_\ell} \bar{\pi}(\bar{s}_\ell|\omega)\bar{p}(\bar{s}_k|\bar{s}_\ell) \\
&= \frac{\pi(s_k|\omega)}{|S_k|} \sum_{\bar{s}'_k \in S_k} \bar{p}(\bar{s}_k|\bar{s}'_k) + \sum_{\ell \neq k} \frac{\pi(s_\ell|\omega)}{|S_\ell|} \sum_{\bar{s}_\ell \in S_\ell} \bar{p}(\bar{s}_k|\bar{s}_\ell) \\
&= \frac{\pi(s_k|\omega)}{|S_k|} \sum_{\bar{s}'_k \in S_k} \bar{p}(\bar{s}'_k|\bar{s}_k) + \sum_{\ell \neq k} \frac{\pi(s_\ell|\omega)}{|S_\ell|} \sum_{\bar{s}_\ell \in S_\ell} \frac{p(s_k|s_\ell)}{|S_k|} \\
&= \frac{\pi(s_k|\omega)}{|S_k|} \cdot \bar{p}(S_k|\bar{s}_k) + \sum_{\ell \neq k} \frac{\pi(s_\ell|\omega)}{|S_\ell|} \cdot |S_\ell| \cdot \frac{p(s_k|s_\ell)}{|S_k|} \\
&= \frac{\pi(s_k|\omega)}{|S_k|} \cdot p(s_k|s_k) + \sum_{\ell \neq k} \frac{\pi(s_\ell|\omega)}{|S_\ell|} \cdot |S_\ell| \cdot \frac{p(s_k|s_\ell)}{|S_k|} \\
&= \frac{1}{|S_k|} \sum_{\ell=1}^{K} \pi(s_\ell|\omega)p(s_k|s_\ell) \\
&= \frac{\sigma(s_k|\omega)}{|S_k|}.
\end{aligned}$$

Therefore, the posterior probability of ω upon observing $\bar{s}_k$ is equal to

$$\bar{\mu}_k(\omega) = \frac{\mu_0(\omega)\bar{\sigma}(\bar{s}_k|\omega)}{\langle \mu_0, \bar{\sigma}(\bar{s}_k|\cdot)\rangle} = \frac{\mu_0(\omega)\frac{\sigma(s_k|\omega)}{|S_k|}}{\langle \mu_0, \frac{\sigma(s_k|\cdot)}{|S_k|}\rangle} = \mu_k(\omega),$$

i.e., all duplicates $\bar{s}_k \in S_k$ under $\bar{\pi}$ yield the same belief as s_k itself under π. Finally, notice that

$$\sum_{\omega\in\Omega} \bar{\sigma}(S_k|\omega) = \sum_{\omega\in\Omega}\sum_{\bar{s}_k\in S_k} \bar{\sigma}(\bar{s}_k|\omega) = \sum_{\omega\in\Omega}\sum_{\bar{s}_k\in S_k} \frac{\sigma(s_k|\omega)}{|S_k|} \sum_{\omega\in\Omega}\sigma(s_k|\omega),$$

i.e., the total probability of obtaining a duplicate in S_k under $\bar{\pi}$ is equal to the probability of obtaining s_k under π. Hence, the distribution of posteriors induced by $\bar{\pi}$ is the same as the distribution of posteriors induced by π. Therefore, the set of distributions of posteriors that can be achieved via the channel p is a subset of the set of distributions of posteriors that can be achieved via the channel $\bar{p}$, which completes the proof. □

Proof of Corollary 1. For convenience, we use the notation $K := |S|$ and $L := |\bar{S}|$. Then, we begin by partitioning the message space $\bar{S}$ into K equivalence classes, $\{S_1, \ldots, S_K\}$, each containing M messages. By strong symmetry of q, for every $\bar{s}_k, \bar{s}'_k \in S_k$ and every S_ℓ, we obtain $q(S_\ell|\bar{s}_k) = q(S_\ell|\bar{s}'_k)$. Moreover, again by strong symmetry of q, for every $\bar{s}_k, \bar{s}'_k \in S_k$ we obtain $q(\bar{s}'_k|\bar{s}_k) = q(\bar{s}_k|\bar{s}'_k)$. Hence, if we take the message space $S = \{s_1, \ldots, s_K\}$ together with the auxiliary channel $\tilde{p} : S \to \Delta(S)$ defined by $\tilde{p}(s_\ell|s_k) := q(S_\ell|\bar{s}_k)$, it will be the case that $q \trianglerighteq \tilde{p}$. Therefore, by Proposition 5, we get

$$\hat{v}^*_q \geq \hat{v}^*_{\tilde{p}}. \tag{A.11}$$

Now, notice that $\tilde{p}$ is a strongly symmetric channel with total error probability

$$\tilde{p}(s'|s) = \frac{L(K-1)}{(L-1)K}\delta.$$

On the other hand, p is also strongly symmetric with the same number of messages as $\tilde{p}$ and total error probability ε, which by hypothesis is (weakly) larger than $\frac{L(K-1)}{(L-1)K}\delta$. Hence, there exists a strongly symmetric channel r such that $p = \tilde{p} \circ r$, and by Proposition 4, we obtain

$$\hat{v}^*_{\tilde{p}} \geq \hat{v}^*_p. \tag{A.12}$$

Finally, combining (A.11) and (A.12) completes the proof. □

References

Aliprantis, C. & Border, K. (1994). *Infinite dimensional analysis*. Springer Verlag, Berlin.

Alonso, R. & Câmara, O. (2016). Persuading voters. *American Economic Review* 106, 35-90–3605.

Blackwell, D. (1951). The comparison of experiments. *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability* 93–102.

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

Blume, A. & Board, O.J. (2013). Language barriers. *Econometrica* 81, 781–812.

——— (2014). Intentional vagueness. *Erkenntnis* 79, 855–899.

Blume, A., Board, O.J. & Kawamura, K. (2007). Noisy talk. *Theoretical Economics* 2, 395–440.

Cover, T.M. & Thomas, J.A. (2006). *Elements of information theory.* John Wiley & Sons.

Crawford, V. & Sobel, J. (1982). Strategic information transmission. *Econometrica* 50, 1431–1451.

Dughmi, S. (2017). Algorithmic information structure design: A survey. *ACM SIGecom Exchanges* 15, 2–24.

Dughmi, S., Kempe, D. & Qiang, R. (2016). Persuasion with limited communication. *Proceedings of the 17th ACM Conference on Economics and Computation (ACM EC'17).*

Dughmi, S. & Xu, H. (2016). Algorithmic Bayesian persuasion. *Proceedings of the 47th ACM Symposium on Theory of Computing (STOC).*

Flores, G. (2006). Language barriers to health care in the United States. *New England Journal of Medicine* 355, 229–331.

Frankel, A. & Kartik, N. (2020). Improving Information from Manipulable Data. *Working Paper.*

Glazer, J. & Rubinstein, A. (2004). On optimal rules of persuasion. *Econometrica* 72, 1715–1736.

Guo, Y. & Shmaya, E. (2020). Costly Miscalibration. *Theoretical Economics* forthcoming.

Kamenica, E. (2019). Bayesian persuasion and information design. *Annual Review of Economics* (forthcoming).

Kamenica, E. & Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review* 101, 2590–2615.

Kosenko, A. (2018). Mediated persuasion. *Working Paper*

Le Treust, M. & Tomala, T. (2018). Information-theoretic limits of strategic communication. *Working paper.*

——— (2019). Persuasion with limited communication capacity. *Journal of Economic Theory* 184, 104940.

Marshall, A.W., Olkin, I. & Arnold, B.C. (2011). *Inequalities: Theory of majorization and its applications*. Springer Series in Statistics, 2nd Edition.

Milgrom, P. & Roberts, J. (1986). Relying on the information of interested parties. *RAND Journal of Economics* 17, 18–32.

Rayo, L. & Segal, I. (2010). Optimal information disclosure. *Journal of Political Economy* 118, 949–987.

Perez-Richet, E. & Skreta, V. (2018). Test design under falsification. *Working Paper.*