Games and Economic Behavior 130 (2021) 44–61

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Noisy persuasion ☆

Elias Tsakas [a,*], Nikolas Tsakas [b]

[a] Department of Economics (MPE), Maastricht University, P.O. Box 616, 6200 MD, Maastricht, the Netherlands
[b] Department of Economics, University of Cyprus, P.O. Box 20537, 1678, Nicosia, Cyprus

ARTICLE INFO

Article history:
Received 19 November 2019
Available online 11 August 2021

JEL classification:
C72
D72
D82
D83
K40
M31

Keywords:
Bayesian persuasion
Noisy channel
Data distortions
Garbling
Complexity

ABSTRACT

We study the effect of noise due to exogenous information distortions in the context of Bayesian persuasion. We first characterize the optimal signal in the prosecutor-judge game from Kamenica and Gentzkow (2011) with a noisy and strongly symmetric communication channel and show that the sender's payoff increases in the number of messages. This implies that, with exogenous noise, the sender prefers to complicate communication. Then, we establish necessary and sufficient conditions for the sender's payoff to weakly increase in the Blackwell-informativeness of the noise channel when the message space and the channel are binary. The reason why a sender may benefit from additional noise is that a garbling may alter the noise structure. Subsequently, we provide sufficient conditions that extend this result to channels of arbitrary cardinality. Finally, we introduce a procedure of making a communication channel more complex and prove that increased complexity benefits the sender.

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 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 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:

☆ 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.

* Corresponding author.
E-mail addresses: e.tsakas@maastrichtuniversity.nl (E. Tsakas), tsakas.nikolaos@ucy.ac.cy (N. Tsakas).
URLs: http://www.elias-tsakas.com/home.html (E. Tsakas), https://sites.google.com/site/nikolastsakas/ (N. Tsakas).

https://doi.org/10.1016/j.geb.2021.08.001
0899-8256/

- 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]

We can now state our three main research questions. The first question arises quite naturally. Namely, we intend to answer what does an optimal signal look like in the presence of noise. To do so, we need to understand the restrictions that noise imposes to the sender. The obvious restriction is that the noise restricts the set of signals from which the sender 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.[2] Second, it restricts the possible combinations of posteriors that the receiver can form upon observing different messages associated with the same experiment.

The second question we seek to answer is, in words, whether a more noisy channel is always more harmful for the sender than a less noisy one. Note that if one of the two channels is noiseless, then the noisy channel is trivially harmful for the sender. Yet, the argument cannot be directly extended to cases where both channels under comparison are noisy. More formally, we are asking whether or not, among two noisy channels that can be ordered with respect to their (Blackwell) informativeness, the sender achieves a higher expected utility in equilibrium under the more informative channel.

Our third and final question focuses on the complexity of the noisy channel. More specifically, we aim to understand whether there exists an upper bound on the number of messages that needs to be used for the sender to maximize her expected utility in equilibrium. Recall that, according to the well-known result of Kamenica and Gentzkow (2011), in the standard noiseless case with a relatively small number of messages –equal to the number of states– the sender can always guarantee her maximum expected utility. Yet, 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 a regulator's approval for some drug, should they choose to perform the clinical trials in-house 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 who has limited understanding of medical terms or faces other communication barriers (e.g., they don't share a common mother tongue)?

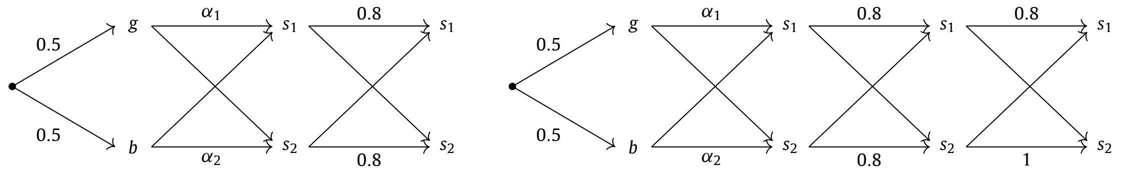
Starting with our first general question, not surprisingly, the answer turns out to be too complex to allow us to provide a general characterization for all persuasion games and all noisy channels. Nevertheless, we fully solve the problem for the standard application that widely appears in the literature, 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 to 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., one can construct 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 ap-

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

[2] 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.



(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 $r(s_1|s_2) = 0$.

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

proved 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.[3] 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 Fig. 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 Fig. 1(b)– on top 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.

But then we naturally ask whether 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 to hold when the message space and the noisy channels are binary. 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 Proposition 2. 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, regarding 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 artifact 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

[3] For a more detailed form of the company's utility function see Fig. 4.

games, the improvement is strict. In particular, the sender becomes better off by introducing additional colors (e.g., yellow and green on top of blue and red). Interestingly, he becomes better off even by introducing different shades of the same color that are hardly distinguishable with each other (e.g., splitting blue into dark blue and light blue, and likewise for red).

Since the seminal contribution of Kamenica and Gentzkow (2011), the Bayesian persuasion literature has surged.[4] Within this body of work, there is a recent growing interest in the role of distortions. Said distortions are typically seen as restrictions, either exogenously imposed (like in our paper) or endogenously emerging.

Starting with Bayesian persuasion with exogenous constraints, there is recent work within both economics and computer science. The idea is that not all signals can be used by the sender –and, a fortiori, not all distributions of posteriors are feasible– due to, for instance, communication being coarse (Dughmi et al., 2016; Aybas and Turkel, 2020), or to the presence of privacy and discrimination considerations (Babichenko et al., 2021). Within this part of the literature, there is also a stream of papers that consider a rationally inattentive receiver (Bloedel and Segal, 2020; Lipnowski et al., 2020), where although the constraints are not hard (i.e., all signals are in principle feasible), the receiver will nonetheless avoid to process certain complex –and, a fortiori, excessively costly– signals.

Within this literature, Le Treust and Tomala (2019) is the closest paper to our work. The authors study Bayesian persuasion with distortions similar to the ones we consider, 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 and the number of reported messages increases. Their main result relies on characterizing the feasible distributions of posterior beliefs given a constraint. This last result is reviewed and generalized (to multiple constraints) by Doval and Skreta (2018), thus providing a general toolbox for studying constrained Bayesian persuasion.

In their recent paper, Babichenko et al. (2021) classify exogenous constraints into ex-ante and ex-post. In both cases posteriors are first mapped on the real line via a continuous function. Then, ex-post constraints restrict the image of each posterior to be below a certain value (see also Volund, 2018), while ex-ante constraints do so only in expectation. The distortions that we consider in our paper induce ex-ante restrictions. The same is true for the related paper of Le Treust and Tomala (2019), as well as papers that study Bayesian persuasion with constraints on the cardinality of the message space (Dughmi et al., 2016; Aybas and Turkel, 2020). Models of Bayesian persuasion with exogenous constraints have been applied in the context of auctions (Dughmi et al., 2014) and bilateral trade (Dughmi et al., 2016).

Let us now focus on Bayesian persuasion models with endogenous constraints. The underlying idea in large part of this literature is that the sender can strategically distort information, sometimes albeit (lying) costs. This way, although the set of feasible signals is unrestricted for the sender, the commitment assumption is weakened. There are different degrees to which commitment can be weakened: in Nguyen and Tan (2021) the sender privately observes the realized signal and can send any message to the receiver, while in Lipnowski et al. (2019) and Min (2020) the realized signal is always sent to the receiver with some positive probability. Guo and Shmaya (2021) study a model similar to the one of Nguyen and Tan (2021), with the difference that the sender does not reveal ex-ante the experiment to the receiver.

Endogenous constraints can also emerge in settings where information is transmitted via mediators (Kosenko, 2018). In this paper, the sender communicates the signal realization to the mediator, who subsequently communicates it to the receiver via a noisy channel of his choice. That way, the sender's information is distorted.

There is also related literature in the context of cheap talk games: Blume et al. (2007) introduce noisy communication to a standard 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 recent work on endogenous information distortions. For instance, Frankel and Kartik (2021) study data manipulation, focusing on the problem of a designer who allocates resources across the (data-generating) agents. In another recent paper, Perez-Richet and Skreta (2021) study a model where 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 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 Appendix A.

[4] For a recent overview, we refer to Kamenica (2019).

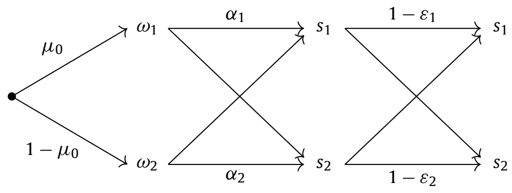


Fig. 2. The sender chooses the experiment in the form of conditional probabilities, $\alpha_1 := \pi(s_1|\omega_1)$ and $\alpha_2 := \pi(s_2|\omega_2)$. A message $s \in S$ is first realized. Then, it is (possibly) distorted, with exogenous and commonly known error probabilities $\varepsilon_1 := p(s_2|s_1)$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

2. Persuasion game with noise

2.1. Noisy information 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 $S = \{s_1, \ldots, s_K\}$ be the finite set of messages that can be encoded with the available technology. A *(noisy) information 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.[5] Thus, $p(s'|s)$ denotes the probability that the receiver observes s' when s has been realized.

A channel is called binary whenever $K = 2$ (Fig. 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 a 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. ◁

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 distance between any two messages, and the error probability depends on said distance.[6] 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 observes 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.

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

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

[5] 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 $(s^1, \ldots, s^k) \in S^k$ 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 $n = k = 1$.

[6] 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).

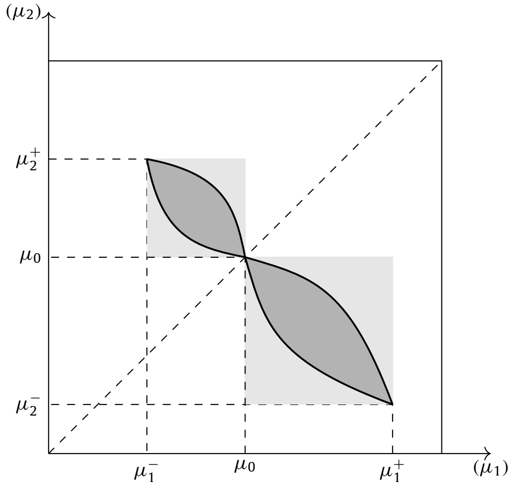


Fig. 3. THE SET OF FEASIBLE PROFILES OF POSTERIORS for the channel of Fig. 2: The posterior μ_1 (resp., μ_2) is obtained when the message s_1 (resp., s_2) is realized. Every signal $\sigma \in \Sigma_p$ leads to a unique pair (μ_1, μ_2) in the inner shaded area (i.e., in the union of the two leaves).

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. Whenever it is clear which is the noisy channel that we have in mind, we omit reference to the subscript p, thus simply writing Σ.

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

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

where $\langle \cdot, \cdot \rangle$ denotes the inner product, as usual. Each signal $\sigma \in \Sigma_p$ induces a profile of posteriors $(\mu_1, \dots, \mu_K)$, where $\mu_k := \mu_{s_k}$ is the posterior given the message s_k. Each posterior $\mu \in \{\mu_1, \dots, \mu_K\}$ occurs with probability $\langle \mu_0, \sigma(\{s \in S : \mu_s = \mu\}|\cdot)\rangle$.

Example 2. Whenever the state space and the channel are both binary, the set of feasible profiles of posteriors takes the form that is illustrated in Fig. 3 (borrowed from Le Treust and Tomala, 2019, Figure 5). Clearly, it will either be the case that $\mu_2 \le \mu_0 \le \mu_1$ (bottom-right leaf) or $\mu_1 \le \mu_0 \le \mu_2$ (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, the set of feasible profiles of posteriors 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. ◁

2.2. Equilibrium analysis

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 the (noisy persuasion) game *trivial* whenever the completely uninformative signal is optimal, implying that persuasion attempts cannot really help the sender. We find such cases uninteresting, and thus we only consider non-trivial games.

Remark 1. Using standard tools, one can easily 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 the set of feasible posterior profiles 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.[7]

We begin with a binary state space $\Omega = \{\omega_1, \omega_2\}$ and a binary action space $A = \{a_1, a_2\}$. 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. On the other hand, the receiver's utility function is 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

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

We impose two assumptions to avoid the game being trivial. First, we assume that the threshold is above the prior: otherwise, the receiver would anyway choose the sender's desired action a_1, implying that the completely uninformative signal would be trivially optimal. Second, we assume that there exists some signal yielding with positive probability larger or equal than the threshold: otherwise, the receiver would always choose a_2 (irrespective of the signal), implying that the sender would remain indifferent across all signal, and a fortiori the completely uninformative signal would again be trivially optimal. 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 resembles to some extent the one in Perez-Richet and Skreta (2021), where exactly one message leads to the undesirable action, and all other messages lead to the sender's preferred action.[8] Our result emphasizes that the sender's value is strictly increasing in K and approaches the value of standard noiseless game as K goes to infinity, i.e., intuitively, *in the presence of noise, more complex communication channels are strictly beneficial for the sender*, and in the limit *infinite complexity removes the restrictions imposed by the presence of noise*. This follows from the fact that all messages that lead to the sender's preferred action, do so by inducing exactly the threshold posterior belief, and the total probability of this posterior belief being realized increases with the number of messages. In the context of Example 1, the optimal signal will be such that the swimmer will visit the beach upon seeing any color except red, and the municipality always becomes strictly better off by complicating the flag system, i.e., by introducing more and more colors. 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. Interestingly, our conclusion is aligned with recent work on persuasion with coarse communication, where increased complexity always benefits the sender (Aybas and Turkel, 2020). We further elaborate on the role of complexity in Section 4.

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

[8] Of course, in their result the optimal signal will use uncountably many messages in equilibrium, as opposed to our case, where the number of messages is exogenously restricted to K. Furthermore, in their case, different messages induce different posteriors (above the threshold), as opposed to our result where all "good messages" induce the same posterior, viz., exactly the threshold.

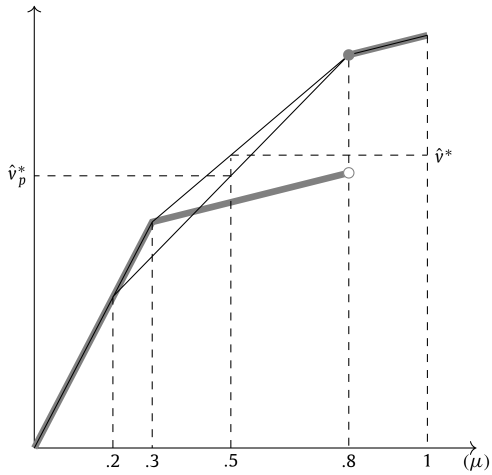


Fig. 4. PERSUADING THE REGULATOR: The sender's expected utility is a function of the receiver's posterior beliefs, and it is depicted by the thick grey line. The respective concave closure is depicted by the thin black line. The sender's value without noise is equal to $\hat{v}^*$, whereas his value with noise being given by the channel p is equal to $\hat{v}_p^*$.

Finally, it is worthwhile pointing out that the size of K is inconsequential for the receiver. This is because the optimal signal leads to a distribution of posteriors that puts positive probability to some posterior below the prior and the remaining probability is exactly at the threshold posterior. This means that the sender's non-preferred action is optimal for the receiver for both of these posteriors, because for the high posterior the receiver is indifferent between the two actions. As a result, the receiver's expected utility will always remain equal to the one he would have received if the sender had chosen the completely uninformative signal (irrespective of K). So, overall, increased complexity leads to a Pareto improvement.

2.3. Leading example: persuading the regulator

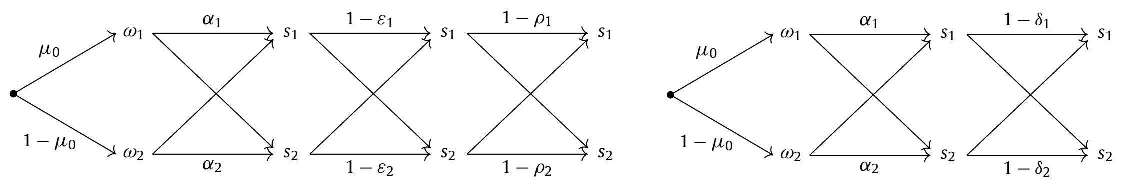
Let us formalize our leading example from the introduction. Accordingly, a pharmaceutical company (sender) wants to submit an application to the regulator (receiver) for an experimental drug to be approved for commercial use. It is common knowledge that the drug is effective at state ω_1 which occurs with probability $\mu_0 = 0.5$, and it is ineffective at ω_2. An application consists of a clinical trial, which is modeled in its reduced form as an experiment over a binary information 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).

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 Fig. 3). Hence, the value of the optimal signal is equal to $\hat{v}_p^* = 1/2$ (see Fig. 4). In fact, this profile of posteriors can be achieved if the sender designs a fully informative experiment.

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 $\hat{v}^* = 5/8$, thus $\hat{v}^* > \hat{v}_p^*$. 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.



(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 $\delta_2 := \varepsilon_2(1-\rho_1) + (1-\varepsilon_2)\rho_2$.

Fig. 5. Garbling with a binary state space: The channel q is more noisy than the channel p.

As shown in Proposition 1, a richer message space could increase value the sender could obtain from an optimally designed experiment compared to $\hat{v}_p^*$. Namely, with a message space $\{s_1, s_2, s_3\}$ and a strongly symmetric noisy channel p' with total error probability $\varepsilon = 0.2$, an optimal experiment could lead the receiver to form the good posterior 0.8 upon observing two of the three messages, say s_1 and s_2, and to form a bad posterior of approximately 0.11 upon observing s_3.[9] In this case, an optimal experiment would yield to the sender value $\hat{v}_{p'}^* = 9/16$, for which it holds that $\hat{v}^* > \hat{v}_{p'}^* > \hat{v}_p^*$.

3. Does more noise harm the sender?

Let us now turn to one of the main research questions of the paper, viz., *is more noise always harmful or are there cases where it is beneficial for the sender?* Formally speaking, *is the sender's value increasing with respect to the informativeness of the noisy channel?* We study this question both for binary and for some more general information 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 Fig. 2) is illustrated below (in Fig. 5).

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 Fig. 6. 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.

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. ◁

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, 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. ◁

[9] In this case, there are several optimal experiments for the sender, all of which yield the same value for the receiver. One of those would be $\pi(s_1|\omega_1) = \pi(s_2|\omega_1) = 1/2$ and $\pi(s_1|\omega_2) = \pi(s_2|\omega_2) = 1/56$.

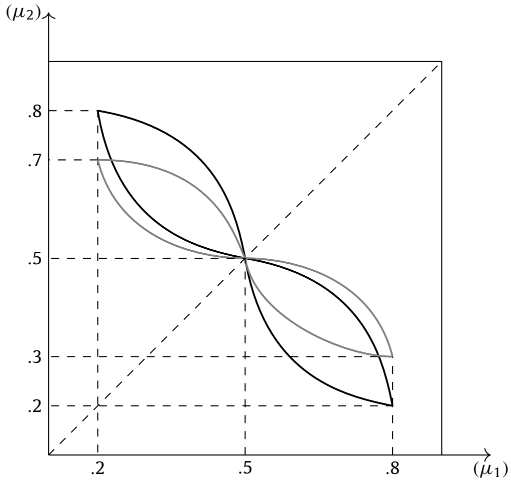


Fig. 6. The effect of garbling on the set of feasible profiles of posteriors: the black graph corresponds to the profile of posteriors under p, whereas the gray graph corresponds to the profiles of posteriors under q.

The question then becomes: *can we identify the garblings that always make the sender worse off, irrespective of the preference profile and the prior?* This would essentially account to finding the garblings that shrink the set of feasible profiles of posteriors. 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 Fig. 5. 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:*

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

The interpretation of the previous result can be twofold. 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 Fig. 6, 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 profiles 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 Fig. 6, 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 set of feasible profiles of posteriors shrink too. Hence, the sender maximizes his expected utility function over a smaller domain, 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 fact that each signal $\sigma \in \Sigma_p$ is identified by a mapping from Ω to $\Delta(B_p) \subseteq \Delta(S)$, implying that $(\Delta(B_p))^\Omega$ represents Σ_p. Hence, $\Delta(B_q) \subseteq \Delta(B_p)$ implies $\Sigma_q \subseteq \Sigma_p$.

In Proposition 3, we start with the matrix representation of a noisy channel. Accordingly, p is represented by the stochastic transition matrix P, where $P_{k,\ell} := p(s_\ell|s_k)$. Then, 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 $\Sigma_q \subseteq \Sigma_p$.

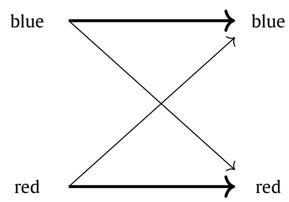
In Proposition 4, we show that garbling p with a strongly symmetric channel r leads to shrinking of $\Delta(B_p)$. The underlying idea is that each $q(\cdot|s)$ will lie on the linear segment that connects $p(\cdot|s)$ with the center of $\Delta(S)$. Then, given that the center lies in the interior of $\Delta(B_p)$, 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. Does the sender like complex information structures?

As it is well-known, in standard persuasion games the sender can always achieve her value with a given –relatively small– number of messages, viz., there exists always an optimal signal (for the sender) that uses no more messages than the number of states (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.[10]

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.

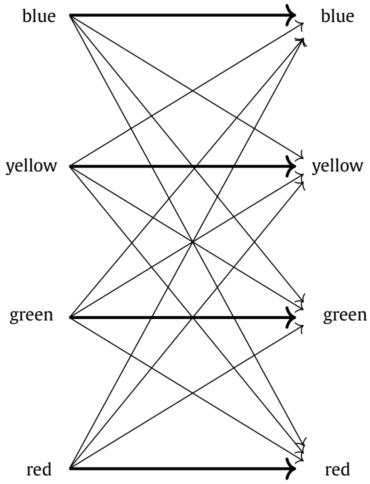
Example 1 (continued). Suppose that we begin with a binary strongly symmetric channel p:



The bold arrows represent probability 0.85 and the thin ones represent probability 0.15, i.e., we have $p(\text{blue}|\text{blue}) = p(\text{red}|\text{red}) = 0.85$. Now, consider the following two ways of replacing p with a more complex channel:

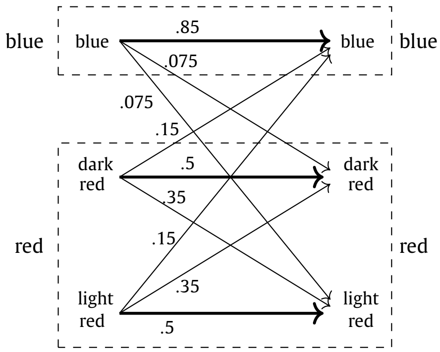
- The municipality introduces additional distinct colors, thus obtaining an expanded message space {blue, red, yellow, green}. Nevertheless, the probability of not confusing colors remains the same, i.e., $q(\text{blue}|\text{blue}) = q(\text{red}|\text{red}) = q(\text{yellow}|\text{yellow}) = q(\text{green}|\text{green}) = 0.85$.

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



Once again, the thick arrows represent probability 0.85, while the thin arrows now stand for probability 0.05.

- 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., $\bar{p}(\{\text{dark red, light red}\}|\text{dark red}) = \bar{p}(\{\text{dark red, light red}\}|\text{light red}) = 0.85$ and $\bar{p}(\text{blue}|\text{blue}) = 0.85$. At the same time, different shades of the same color are almost indistinguishable, i.e., $\bar{p}(\text{light red}|\text{dark red}) = \bar{p}(\text{dark red}|\text{light red}) = 0.35$ and $\bar{p}(\text{dark red}|\text{dark red}) = \bar{p}(\text{light red}|\text{light red}) = 0.50$.



Notice that both ways of expanding the message space keep certain elements of the structure of p unchanged, 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

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

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. Notice that the total error probability of one set of duplicates agrees with the original probabilities, e.g., $\bar{p}(\{\text{dark red, light red}\}|\text{blue}) = p(\text{red}|\text{blue}) = 0.15$. Moreover, conditional on blue each shade of red receives the same probability, e.g., $\bar{p}(\text{dark red}|\text{blue}) = \bar{p}(\text{light red}|\text{blue}) = 0.075$. And finally, within the red shades, noise is symmetric, e.g., this means that $\bar{p}(\text{light red}|\text{dark red}) = \bar{p}(\text{dark red}|\text{light red}) = 0.35$. ◁

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 becomes better off. Let us illustrate this last point in our working example.

Example 1 (continued). We will proceed into steps.

First, we take the strongly symmetric channel p with two colors (viz., blue and red) and probability of seeing the correct color equal to 0.85. Then, we construct the more complex channel $\bar{p}$ as follows. First, we duplicate each color by taking the respective dark and light shades, i.e., we now have four messages in total (viz., light blue, dark blue, light red and dark red). The total probability of seeing a shade of the incorrect color remains the same and is split uniformly across the two shades, e.g., $\bar{p}(\text{light red}|\text{light blue}) = \bar{p}(\text{dark red}|\text{light blue}) = 0.15/2 = 0.075$. The probability of seeing the wrong shade of the correct color is also set equal to 0.075, thus implying that every mistake occurs the same probability 0.075, while the probability of seeing the correct color is equal to 0.775. Of course, by construction $\bar{p}$ is more complex than p, and therefore by Proposition 5 it always makes the sender better off.

In the second step, take the strongly symmetric channel q with the same message space as $\bar{p}$ (viz., light blue, dark blue, light red and dark red) and probability of seeing the correct color equal to 0.85, which is of course higher than 0.775. Hence, by Proposition 3, the sender will always be better off with q than with $\bar{p}$. Therefore, by transitivity, q will always be better for the sender than p. ◁

Notice that this last comparison (between two strongly symmetric channels that differ only in the number of messages, while having the same total error probability) will often make the sender to strictly prefer the more complex. This is because the distributions of posteriors that can be achieved in one case are a strict superset of the distributions that can be achieved in the other. Hence, for a large family of utility functions this additional complexity will lead to a strict improvement for the sender.

Finally, it is worthwhile pointing out that, unlike the specific case of the judge-prosecutor example, no general conclusion can be drawn regarding the effect of complexity on the receiver's expected utility in equilibrium. Depending on the specific case in which we make the channel more complex, the resulting optimal signal could be more informative, but it could also be less informative. Then, given that the receiver's indirect utility function is always convex, such increased complexity could respectively be beneficial or harmful for the receiver.

5. Discussion

5.1. Endogenous noise

An interesting extension of our model is to consider experiment-dependent noisy channels. Consider the following example.[11] 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.

Appendix 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 $\xi_k := (1-\varepsilon)x_k + \frac{\varepsilon}{K-1}(1-x_k)$ and $\psi_k := (1-\varepsilon)y_k + \frac{\varepsilon}{K-1}(1-y_k)$. Note that the posterior belief given s_k will be equal to

$$\mu_k := \frac{\mu_0\xi_k}{\mu_0\xi_k + (1-\mu_0)\psi_k}.$$

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

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

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 $\mu_K < \bar{\mu} \le \mu_1^\lambda < \mu_1$. Hence, the receiver chooses a_1 under π, if and only if, he chooses a_1 under π_λ. Then, letting $M := \max\{k = 1, \dots, K | \mu_k \ge \bar{\mu}\}$, the expected probability of the receiver choosing a_1 (and a fortiori the sender's expected utility) is equal to

$$\mu_0 \sum_{k=1}^{M} \xi_k + (1-\mu_0)\sum_{k=1}^{M} \psi_k$$

under the experiment π, and equal to

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

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

$$\psi_k = \frac{\mu_0(1-\bar{\mu})}{(1-\mu_0)\bar{\mu}}\xi_k$$

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

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

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 $\Omega = \{\omega_1, \dots, \omega_N\}$. Thus, we can identify each experiment $\pi \in \Pi$ with the vector $(\pi_1, \dots, \pi_N) \in [0,1]^N$, where $\pi_n := \pi(s_1|\omega_n)$ for every $n = 1, \dots, N$. 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 $(\sigma_1, \dots, \sigma_N) \in [0,1]^N$, where $\sigma_n := \sigma(s_1|\omega_n)$ for every $n = 1, \dots, N$. So, the set of feasible signals is

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

The latter can be rewritten as

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

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

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

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 $\hat{v}_p^* \ge \hat{v}_q^*$.

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

$$(\mu_1', \mu_2') = \left(\frac{1-\delta_1}{1-\delta_1+\delta_2} \, , \, \frac{\delta_1}{1+\delta_1-\delta_2} \right) \in \mathcal{M}_q, \tag{A.2}$$

where $\mathcal{M}_q$ is the set of feasible profiles of posteriors under the channel q. 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

$$\left(\frac{\sigma_1}{1+\sigma_1-\sigma_2} \, , \, \frac{1-\sigma_1}{1-\sigma_1+\sigma_2} \right) \in \left\{ (\mu_1', \mu_2'), (\mu_1'', \mu_2'') \right\}. \tag{A.4}$$

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, $(\mu_1'', \mu_2'') = \left(\frac{\sigma_1}{1+\sigma_1-\sigma_2} \, , \, \frac{1-\sigma_1}{1-\sigma_1+\sigma_2} \right)$ 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, $\hat{v}_q^* \leq \hat{v}_p^*$. □

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 $\hat{v}_p^* \geq \hat{v}_q^*$. □

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

$$\bar{\pi}(\bar{s}_k|\omega) := \frac{\pi(s_k|\omega)}{|S_k|} \tag{A.11}$$

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

$$\bar{\sigma}(\bar{s}_k|\omega) = \sum_{\ell=1}^{K}\sum_{\bar{s}_\ell \in S_\ell} \bar{\pi}(\bar{s}_\ell|\omega)\bar{p}(\bar{s}_k|\bar{s}_\ell) \tag{A.12}$$

$$= \frac{\pi(s_k|\omega)}{|S_k|}\sum_{\bar{s}_k' \in S_k} \bar{p}(\bar{s}_k|\bar{s}_k') + \sum_{\ell \neq k}\frac{\pi(s_\ell|\omega)}{|S_\ell|}\sum_{\bar{s}_\ell \in S_\ell} \bar{p}(\bar{s}_k|\bar{s}_\ell) \tag{A.13}$$

$$= \frac{\pi(s_k|\omega)}{|S_k|}\sum_{\bar{s}_k' \in S_k} \bar{p}(\bar{s}_k'|\bar{s}_k) + \sum_{\ell \neq k}\frac{\pi(s_\ell|\omega)}{|S_\ell|}\sum_{\bar{s}_\ell \in S_\ell} \frac{p(s_k|s_\ell)}{|S_k|} \tag{A.14}$$

$$= \frac{\pi(s_k|\omega)}{|S_k|}\cdot \bar{p}(S_k|\bar{s}_k) + \sum_{\ell \neq k}\frac{\pi(s_\ell|\omega)}{|S_\ell|}\cdot |S_\ell| \cdot \frac{p(s_k|s_\ell)}{|S_k|} \tag{A.15}$$

$$= \frac{\pi(s_k|\omega)}{|S_k|}\cdot p(s_k|s_k) + \sum_{\ell \neq k}\frac{\pi(s_\ell|\omega)}{|S_\ell|}\cdot |S_\ell| \cdot \frac{p(s_k|s_\ell)}{|S_k|} \tag{A.16}$$

$$= \frac{1}{|S_k|}\sum_{s_\ell \in S_\ell} \pi(s_\ell|\omega)p(s_k|s_\ell) \tag{A.17}$$

$$= \frac{\sigma(s_k|\omega)}{|S_k|}, \tag{A.18}$$

where the Equation (A.12) follows from the definition of a signal; Equations (A.13) follows from (A.11); Equation (A.14) follows from $\bar{p}$ being more complex than p; Equation (A.15) follows from summing up over S_k and S_ℓ respectively; Equation (A.16) follows again from $\bar{p}$ being more complex than p; Equation (A.17) follows from summing up over all elements of S_ℓ; and finally Equation (A.18) follows from the definition of a signal.

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, $\{S_1, \dots, S_K\}$, 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 $S = \{s_1, \dots, s_K\}$ 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

$$\hat{v}^*_q \geq \hat{v}^*_{\tilde{p}}. \tag{A.19}$$

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

$$\hat{v}^*_{\tilde{p}} \geq \hat{v}^*_p. \tag{A.20}$$

Finally, combining (A.19) and (A.20) completes the proof. □

References

Alonso, R., Câmara, O., 2016. Persuading voters. Am. Econ. Rev. 106, 35–90–3605.
Aybas, Y., Turkel, E., 2020. Persuasion with coarse communication. Working Paper.
Babichenko, Y., Talgam-Cohen, I., Zabarnyi, K., 2021. Bayesian persuasion under ex ante and ex post constraints. Working Paper.
Blackwell, D., 1951. The comparison of experiments. In: Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability, pp. 93–102.
Blackwell, D., 1953. Equivalent comparisons of experiments. Ann. Math. Stat. 24, 265–272.
Bloedel, I., Segal, I., 2020. Persuading a rationally inattentive agent. Working Paper.
Blume, A., Board, O.J., 2013. Language barriers. Econometrica 81, 781–812.
Blume, A., Board, O.J., 2014. Intentional vagueness. Erkenntnis 79, 855–899.
Blume, A., Board, O.J., Kawamura, K., 2007. Noisy talk. Theor. Econ. 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.
Doval, L., Skreta, V., 2018. Constrained information design: Toolkit. Working Paper.
Dughmi, S., Immorlica, N., Roth, A., 2014. Constrained signalling in auction design. In: Proceedings of the 25th Symposium on Discrete Algorithms. SODA.
Dughmi, S., Kempe, D., Qiang, R., 2016. Persuasion with limited communication. In: Proceedings of the 17th ACM Conference on Economics and Computation. ACM EC'17.
Flores, G., 2006. Language barriers to health care in the United States. N. Engl. J. Med. 355, 229–331.
Frankel, A., Kartik, N., 2021. Improving information from manipulable data. J. Eur. Econ. Assoc. https://doi.org/10.1093/jeea/jvab017.
Guo, Y., Shmaya, E., 2021. Costly miscalibration. Theor. Econ. 16, 477–506.
Kamenica, E., 2019. Bayesian persuasion and information design. Annu. Rev. Econ. 11, 249–272.
Kamenica, E., Gentzkow, M., 2011. Bayesian persuasion. Am. Econ. Rev. 101, 2590–2615.
Kosenko, A., 2018. Mediated persuasion. Working Paper.
Le Treust, M., Tomala, T., 2018. Information-theoretic limits of strategic communication. Working paper.
Le Treust, M., Tomala, T., 2019. Persuasion with limited communication capacity. J. Econ. Theory 184, 104940.
Lipnowski, E., Mathevet, L., Wei, D., 2020. Attention management. Am. Econ. Rev.: Insights 2, 17–32.
Lipnowski, E., Ravid, D., Shishkin, D., 2019. Persuasion via weak institutions. Working Paper.
Marshall, A.W., Olkin, I., Arnold, B.C., 2011. Inequalities: Theory of Majorization and Its Applications, 2nd edition. Springer Series in Statistics.
Min, D., 2020. Persuasion under partial commitment. Working Paper.
Nguyen, A., Tan, T.Y., 2021. Bayesian persuasion with costly messages. J. Econ. Theory 193, 105212.
Perez-Richet, E., Skreta, V., 2021. Test design under falsification. Working Paper.
Rayo, L., Segal, I., 2010. Optimal information disclosure. J. Polit. Econ. 118, 949–987.
Volund, R., 2018. Bayesian persuasion on compact subsets. In: Theoretical Models in Behavioral Economics, pp. 64–77.