Communication with Endogenous Deception Costs*

Ran Eilat† and Zvika Neeman‡

January 28, 2021

Abstract

We study how the suspicion that communicated information might be deceptive affects the nature of what can be communicated in a sender-receiver game. Sender, who observes the state of the world, is said to *deceive* Receiver if she sends a message that induces beliefs that are different from those that should have been induced in the realized state. Deception is costly to Sender and the cost is endogenous: it increases in the distance between the induced beliefs and the beliefs that should have been induced. A message function that induces the sender to engage in deception is said to be non-credible and cannot be part of equilibrium. We study credible communication in Bayesian persuasion and in cheap-talk games. Importantly, the cost of deception parametrizes the sender's ability to commit to her strategy. Through varying this cost, our model spans the range from no commitment (cheap-talk), to full commitment (Bayesian persuasion).

Keywords: Communication games, costly deception, Bayesian persuasion, cheap talk.

*Acknowledgements: TBA

†Department of Economics, Ben Gurion University of the Negev, eilatr@bgu.ac.il
‡School of Economics, Tel-Aviv University, zvika@tauex.tau.ac.il

1 Introduction

Lying and deception are universally condemned. Philosophers and religious leaders, including Aristotle, Confucius, Saint Augustine, Saint Thomas Aquinas and Immanuel Kant, all emphasized that lying and deception are wrong as such, even with no consideration of their consequences. It comes as no surprise that lying and deception are also costly to those who engage in them (see Abeler, Nosenzo and Raymond 2019 for a survey of experimental studies that document this cost).

In this paper we take the position that a lack of integrity is costly only to the extent that it undermines beliefs. Indeed, a common distinction between lying and deception is that a lie is "a statement that the speaker believes is false" whereas deception is a "statement – or action – that induces the audience to have incorrect beliefs" (Sobel, 2020).[1] Accordingly, we assume that deception (rather than mere lying) is costly, and study how the suspicion that communicated information might be deceptive affects the nature of what can be communicated.

We consider this question in the context of a standard model of communication between an informed Sender (she) and an uninformed Receiver (he). Sender observes a certain variable and sends a message about it to Receiver who, upon receiving the message, takes an action. The payoffs of both Sender and Receiver depend both on the value of the variable and on Receiver's action. We enrich the standard model by assuming that Sender may deceive Receiver, at a cost.

We measure the cost of deception in terms of difference between beliefs. Specifically, Receiver forms beliefs about the relevant variable that depend on the prior distribution, Sender's message strategy, and the actual message sent. Sender may deceive Receiver by sending a message that is different from what she was supposed to send given her message strategy, at a cost that is increasing in the distance between the beliefs induced by the message actually sent, and the beliefs that would have been induced under the message that was supposed to be sent. Importantly, this implies that the cost of deception in our model is measured relative to Receiver's equilibrium expectations and so is endogenous to the model, because it depends on the message strategy chosen by Sender.

A message function that induces Sender to engage in deception is not credible, and cannot be part of equilibrium. We are interested in what can be communicated in equilibrium, via credible message functions.

The ability of Sender to deceive Receiver is closely related to Sender's ability to commit to her message strategy, in the sense of sending the specific message prescribed by the strategy and not a different message. A sufficiently large cost of deception implies "full commitment" of Sender to her message strategy. Such commitment is obviously very valuable. It is a stan-

[1] As emphasized by Sobel (2020), these definitions imply that a lie need not be deceptive, and deception need not involve lying.

dard assumption in the literature on Bayesian persuasion (Kamenica and Gentzkow, 2011). In contrast, costless deception implies that Sender cannot commit to follow her message strategy, and consequently, messages should be interpreted as mere "cheap-talk" (Crawford and Sobel, 1982). Both of these extreme cases serve as useful benchmarks for us. They have both been extensively studied in the literature surveyed below. Through varying the cost of deception in our model, our approach allows us to span the range from cheap-talk, or no commitment, to full commitment.

To illustrate, consider the following stylized example of a candidate for public office, who tries to persuade the public to elect her for office. The state of the world is either "business as usual" or "looming crisis," with prior probabilities two-thirds and a third, respectively. Suppose that the public would like to elect the candidate if it believes that the posterior probability that the state is looming crisis, denoted p, is larger than or equal to one-half. If it believes this probability to be smaller than one-half, then it prefers to elect another candidate. Suppose that the payoff to the candidate from being elected or not is one and zero, respectively, regardless of the state of the world.

Suppose that the candidate, who observes the state of the world, employs the following message strategy. If the state is looming crisis, then the candidate sends the message "crisis." And, if the state is business as usual, then the candidate randomizes between sending the message "crisis" and the message "usual" with equal probabilities.

Upon hearing the message "usual," a public that believes that the candidate is using this message strategy realizes that the state is business as usual ($p = 0$) and elects another candidate. But upon hearing the message "crisis", the public believes that the state is looming crisis with probability one-half ($p = \frac{1}{2}$) and elects the candidate. As famously shown by Kamenica and Gentzkow (2011), this is the message function that maximizes the expected payoff of a candidate who has the ability to fully commit to following her message strategy.

Suppose that the cost of deception to the candidate is given by the difference in induced beliefs p under the message actually sent and the message that was supposed to be sent under the message strategy described above. This means that when the state is business as usual, the cost of sending the message "crisis" (that induces the public to elect the candidate) instead of the message "usual" (that induces the public to elect another candidate) is one-half.

The public realizes that if it believes that the candidate uses the message strategy described above, then the benefit to the candidate of sending the message "crisis" instead of "usual" is one. So the public understands that the message strategy described above is not credible, in the sense that it cannot possibly be employed by the candidate in equilibrium.

As explained above, we are interested in what can be communicated through credible message functions, which can be part of an equilibrium, and that would not induce the candidate to deceive the public. In this example, because the benefit that the candidate obtains from being and not being elected is one and zero, respectively, as long as the cost of deception is strictly smaller than one, a message function that induces certain election of the candidate following some message and election of another candidate following some other message cannot be credible.

The message function in which the candidate sends the message "crisis" if the state is looming crisis, and "usual" otherwise, is credible. The ex-ante expected payoff of the candidate under this message function is one-third, compared to two-thirds under the message function the candidate could use if it were able to fully commit to it. This is the optimal credible message strategy for the candidate in this example. The fact that this strategy fully reveals the state of the world to the public suggests that the public benefits from the fact that the candidate cannot commit to her message strategy. Indeed, as we show, in certain circumstances, the mistrust that is induced by the possibility of deception can benefit the receiver.[2]

We provide a formal definition of a credibility along the lines described above, and show that no loss of optimality is implied by restricting attention to credible message function that employ no more messages than the number of states of the world.

We study the implications of credible communication in a model of Bayesian persuasion in which Receiver's optimal action depends only on the expected state, and Sender's preferences over Receiver's actions are independent of the state of the world. We geometrically characterize Sender's highest equilibrium payoff in this model. We show that this highest payoff is obtained on a partial concavification of Sender's indirect payoff function, which is based on an upper envelope of Sender's indirect payoff function with a bounded slope that depends on the cost of deception. The bound on the slope increases with the cost of deception.

The literature on Bayesian persuasion has already highlighted the link between Sender's value and the upper envelope of her indirect payoff function. Kamenica and Gentzkow (2011) famously characterized Sender's value in terms of the *concave* envelope of her indirect payoff function. Others, such as Lipnowski and Ravid (2020), and Lipnowski, Ravid and Shishkin (2020), have characterized Sender's value in terms of the *quasi-concave*, and a mixture of the concave and quasi-concave envelopes of Sender's indirect payoff function. In contrast, we characterize Sender's value in terms of the concave envelopes of her indirect payoff function, with a bounded slope.

We provide conditions that ensure that Sender's value is continuous and present examples where it is discontinuous (Lipnowski, Ravid and Shishkin, 2020, interpret discontinuity in the cost of deception as "collapse of trust"). Finally, we discuss the circumstances under which productive mistrust arises.

We then study credible communication in a discrete version of the uniform-quadratic

[2] Lipnowski, Ravid and Shishkin (2020) refer to this as "productive mistrust." We explain the circumstances under which it arises below.

example of Crawford and Sobel's (1982) model of strategic communication. An important difference between this setting and the one described above is that, in this setting, Sender's payoff does depend on the state of the world. We show that, in contrast to Crawford and Sobel's (1982) model where communication induces a partition of the state space into convex sets, in our model Sender may induce a credible partition of the state space with nonconvex elements. However, the *optimal* partition for Sender consists only of convex sets of states of the world as in Crawford and Sobel (1982). This result allows us to explicitly solve for the optimal message function for Sender and describe how it relates to the optimal message function in Crawford and Sobel (1982). The fact that deception is costly facilitates more informative communication between Sender and Receiver compared to the most informative equilibrium in Crawford and Sobel (1982), and is akin to decreasing the value of the parameter that measures the discrepancy between Sender's and receiver's payoff in Crawford and Sobel's uniform-quadratic example.

Related Literature

Sobel (2020) introduces game theoretic definitions of lying and deception. Our definition of deception is consistent with his in that, in our model too, deception involves inducing "incorrect beliefs." However, we also add a cost of deception that is not explicitly incorporated into Sobel's model. More importantly, according to our definition, deception is measured with respect to equilibrium beliefs and is therefore endogenous, whereas in Sobel's model, deception is with respect to the true state, and so is determined by an exogenous standard.

Kartik (2009) adds the possibility of costly lying into the communication game considered by Crawford and Sobel (1982). The cost of lying in Kartik's model depends only on the sender's type and the literal message he uses, which may be interpreted as an announcement about the sender's type. Equilibria in his model involve lying, but no deception. The key difference between Kartik's model and ours is that we measure the cost of deception in terms of the differences in Receiver's induced beliefs. Other models of lying consider perturbed versions of games in which, with positive probability, the sender is a behavioral type who always reports honestly; or the receiver is a behavioral type who interprets messages literally (believing that the state is m after receiving the message m) (Chen, 2011). Fischbacher and Föllmi-Heusi (2008) and Gneezy (2005) are examples of experimental papers on communication that associate the message to the state and treat messages as lies if they are not equal to the state.

As mentioned above, this paper contributes to the literature on strategic information transmission. To place our work in context, it is useful to consider the two extreme benchmarks of full commitment and cheap-talk. Full commitment is assumed in the Bayesian persuasion literature (Aumann and Maschler, 1995; Kamenica and Gentzkow, 2011; Kamenica, 2019), which studies sender-receiver games in which a sender commits to an information-transmission strategy. In contrast, in the case of cheap-talk (Crawford and Sobel, 1982) the sender has no commitment power whatsoever. Lipnowski and Ravid (2020) study cheap talk games under state-independent sender preferences.[3]

Lipnowski, Ravid and Shishkin (2020) and Guo and Shmaya (forthcoming) are similar to our work in that they both bridge the gap between cheap-talk and Bayesian persuasion models. In Lipnowski, Ravid and Shishkin (2020), a sender commissions a study to persuade a receiver, but may influence the report with some state-dependent probability. They show that increasing this probability can benefit the receiver and can lead to a discontinuous drop in the sender's payoffs. In Guo and Shmaya (forthcoming) each communicated message is a distribution of states, which the receiver trusts, and the sender faces a miscalibration cost that increases in the distance between the message and its induced equilibrium posterior belief. They show that when costs are sufficiently large, the sender attains her full-commitment payoff under any extensive-form rationalizable play.

Perez-Richet and Skreta (2020) also consider costly falsification of signals. In their model, an agent can manipulate a Blackwell experiment's input at a cost. They characterize receiver-optimal tests under different constraints in this setting. In Nguyen and Tan (2019), a sender has the opportunity to privately change the publicly observed outcome of a previously announced experiment, at a cost that depends on the outcome. They describe conditions under which the sender does not alter the experiment's outcome in the sender-optimal equilibrium.

The fact that Sender's payoff depends directly on Receiver's endogenous beliefs implies that the game we consider is a psychological game (Geanakoplos, Pearce and Stacchetti, 1989, Battigalli and Dufwenberg, 2009).[4] This literature justifies the distaste for lying through an aversion to guilt (Battigalli and Dufwenberg, 2007). Other papers consider communication between an informed Sender and an uninformed Receiver within the framework of psychological games, as we do, but with a very different focus from ours. See for example Caplin and Leahy (2004), Ottaviani and Sørensen (2006), and Ely, Frankel and Kamenica (2015).

The rest of the paper proceeds as follows. Section 2 describes the model and introduces our definition of a credible message function. In this section we also characterize the number of messages in credible message functions. In Section 3 we study credible message functions in a model of Bayesian persuasion, and in Section 4 we study credible message functions in cheap-talk games. As mentioned above, f ormally, the main difference between Sections 3 and 4 is that the former section is devote to the study of state-independent sender preferences, whereas the latter considers a case with state-dependent sender preferences. Section 5 concludes. All proofs are relegated to the appendix.

[3] Min (2018) generalizes their work and allows for sender's preferences to be state dependent. He shows that allowing the sender to commit with positive probability strictly helps both players in Crawford and Sobel's (1982) uniform-quadratic example.

[4] For a recent survey of the literature on psychological games see Battigalli and Dufwenberg (forthcoming).

2 Model

We begin by describing a two-player communication game in Section 2.1. We then enrich the model by adding costly deception and define our notion of credibility in Section 2.2. In Section 2.3 we analyze the number of messages employed by a credible sender.

2.1 The Communication Game

Consider a two-player game, with Sender (S, she) and Receiver (R, he). Players' payoffs depend on a state of the world and on Receiver's action. The state of the world is drawn from a set $\Omega \times \Theta$. The set $\Omega = \{\omega_1, \dots, \omega_N\}$ is finite and represents the "payoff relevant" part of the state of the world. The set $\Theta = [0, 1]$ is used in order to incorporate lotteries into Sender's choice of messages as described below. The prior probability of the payoff relevant part of the state is denoted by $\pi \in \Delta(\Omega)$, where $\pi(\omega)$ is the probability of state ω.[5] For simplicity, we assume that $\pi(\omega) > 0$ for all $\omega \in \Omega$. Without loss of generality, we assume that the prior distribution over Θ is uniform. These two prior distributions are stochastically independent.

Sender chooses a finite set of messages $M \subset \mathbb{R}$ and a measurable message function $\sigma(\omega, \theta)$: $\Omega \times \Theta \to M$ that is monotone non-decreasing in $\theta \in \Theta$.[6] Given a message function σ, we denote the probability that message m is sent in state ω by $q^{\sigma}(m, \omega) = \int_{\{\theta : \sigma(\omega, \theta) = m\}} d\theta$, and the probability that message m is sent by the message function σ by $q^{\sigma}(m) = \sum_{\omega \in \Omega} q^{\sigma}(m, \omega)\, \pi(\omega)$.

Receiver's beliefs about the payoff relevant part of the state are determined according to Bayes rule, whenever possible. If Sender does not send any message, or sends a message that was not supposed to be sent by σ, then Bayes rule cannot be applied and we assume that Receiver believes that the "worst has happened". Namely, Receiver's posterior belief is the one that induces the worst possible indirect payoff for Sender, as defined below. We denote by $p_m^{\sigma} \equiv q^{\sigma}(\cdot | m) \in \Delta(\Omega)$ the posterior *distribution* over Ω that is induced by message m, given the message function σ.

Receiver chooses an action from a compact set $A \subset \mathbb{R}$. The payoff for Receiver is given by $u_R(a, \omega)$. The payoff for Sender is given by her material payoff $u_S(a, \omega)$, and if she deceives Receiver then she also incurs a cost of deception which is described below. The functions $u_R(a, \omega)$ and $u_S(a, \omega)$ are assumed to be continuous in a.

Both Receiver and Sender are expected utility maximizers. Upon observing a message m, Receiver takes the action $a \in A$ that maximizes his expected payoff given the posterior belief

[5]We denote by $\Delta(X)$ the set of probability distributions over a set X.

[6]The assumption that the M is a subset of $\mathbb{R}$ imposes a convenient order on the set of messages and entails no loss of generality. The assumption that $\sigma(\omega, \theta)$ is monotone in its second parameter also involves no loss of generality because payoffs do not depend on θ. For any non-monotone message function there exists a monotone message function that induces identical payoffs and beliefs.

induced by m. For any posterior belief $p \in \Delta(\Omega)$, Receiver's optimal action is given by

$$a^*(p) = \arg\max_{a \in A} \left\{ \sum_{\omega \in \Omega} p(\omega) \cdot u_R(a, \omega) \right\}$$

where $p(\omega)$ is the probability that the belief p assigns to the state ω. If Receiver has more than one best response, then we assume that he chooses the one that is best for Sender.[7, 8]

We define $\hat{u}_i(p)$ to be the *indirect (material) payoff* of player $i \in \{S, R\}$ under posterior beliefs $p \in \Delta(\Omega)$. That is, $\hat{u}_i(p)$ is the material payoff of player i when Receiver takes his optimal action under beliefs $p \in \Delta(\Omega)$, or:

$$\hat{u}_i(p) = \sum_{\omega \in \Omega} p(\omega) \cdot u_i(a^*(p), \omega). \tag{1}$$

2.2 Credibility

Sender's material payoff when she sends message m', the state of the world is (ω, θ), and she is believed to be sending her messages according to the message function σ is therefore given by $u_S(a^*(p^\sigma_{m'}), \omega)$. In state (ω, θ) Receiver expects message $m = \sigma(\omega, \theta)$ to be sent. If $m' \neq \sigma(\omega, \theta)$, then Sender is said to deceive Receiver because message m' induces the "wrong" posterior belief $p^\sigma_{m'}$ instead of p^σ_m.

We assume that deception is costly to Sender. Suppose that the state of the world is (ω, θ). The cost to Sender from sending message m' instead of message $m = \sigma(\omega, \theta)$, given message function σ, is

$$c\left(m' | m, \sigma\right) = \alpha \cdot d\left(p^\sigma_{m'}, p^\sigma_m\right),$$

where $d : \Delta(\Omega) \times \Delta(\Omega) \to \mathbb{R}_+$ is a distance function between beliefs over Ω, and $\alpha \geq 0$ is a parameter that scales the cost of deception.[9] That is, the cost of sending a message m' when the state of the world is (ω, θ) is proportional to the distance between the posterior belief p^σ_m that should have been induced by the message $m = \sigma(\omega, \theta)$ and the posterior belief that is actually induced by the message m', which is $p^\sigma_{m'}$.

Hence, the total payoff of Sender from sending message m' when the state of the world is (ω, θ), when she is believed to be using the message function σ, is the difference between her material payoff and her cost of deception. I.e.,

$$u_S(a^*(p^\sigma_{m'}), \omega) - c\left(m' | \sigma(\omega, \theta), \sigma\right).$$

[7] This tie-breaking assumption does not necessarily work in favor of Sender. We discuss this issue further in the context of Example 3 below.

[8] $a^*(p_m)$ exists because u_R is continuous in a, and A is compact.

[9] A distance function $d(x, y)$ satisfies four properties: it is non-negative, symmetric, $d(x, x) = 0$ for every x, and it satisfies the triangle inequality.

As mentioned above, what distinguishes our approach is that in our model the cost of deception is endogenous. It depends on the "true state" of the world and on Sender's chosen message function σ, as opposed to only the true state. (cf., Sobel, 2020).

A message function σ is *credible* if for any two messages m and m', where $m \neq m'$, Sender does not benefit from sending the message m' when, according to σ, she should have sent the message m. Formally,

Definition 1 (Credibility) *A message function σ is credible if for any state of the world $(\omega, \theta) \in \Omega \times \Theta$ and message $m = \sigma(\omega, \theta)$ deception is not profitable:*

$$u_S(a^*(p_m^\sigma), \omega) \geq u_S(a^*(p_{m'}^\sigma), \omega) - c\left(m' | m, \sigma\right) \tag{2}$$

for every message $m' \in M$.

Credibility imposes an incentive compatibility constraint on Sender. If the cost of deception $c(m'|\sigma(\omega, \theta), \sigma)$ is infinite, then Sender has full commitment power. That is, she would never deviate from any message function she chooses, and so this incentive compatibility constraint is never binding. Otherwise, Sender may benefit from deviating from certain messages. In this case, Sender has only partial commitment power because she can commit only to those message functions from which she would not want to deviate. Partial commitment limits Sender's ability to communicate. However, as famously shown by Crawford and Sobel (1982), nontrivial communication is possible even when the cost of deception is zero and Sender has no commitment power whatsoever.

Because Sender may want to deviate from non-credible message functions and this is anticipated by Receiver, a non-credible message function cannot be part of equilibrium. Therefore, henceforth, we restrict our attention to credible message functions.[10]

Sender's problem is to choose a message set M and a credible message function σ that maximizes her expected payoff. That is:

$$\max_{\langle M, \sigma \rangle} \sum_{\omega \in \Omega} \sum_{m \in M} u_S\left(a^*(p_m^\sigma), \omega\right) \cdot q^\sigma(m, \omega) \cdot \pi(\omega) \tag{SP}$$

$$\text{s.t. } \sigma \text{ is a credible message function.}$$

2.3 An Upper Bound on The Number of Messages

With full commitment, no loss of optimality is implied by restricting attention to message functions that employ no more than $|\Omega|$ messages. However, the optimal message function with full commitment may violate credibility. Moreover, there exist (sub-optimal) message

[10] Because a constant message function (e.g. "silence") is credible, a credible message function exists.

functions whose payoffs cannot be obtained with only $|\Omega|$ messages.[11] It is therefore not a priori clear what is the number of messages that are required to support the optimal *credible* message function. On the one hand, employing a small number of messages decreases the number credibility constraints. On the other hand, it may be the case that the way to achieve the optimal credible outcome is to employ a large number of messages such that the gain from deviating from one message to another is small.

The next proposition implies the following two results: No loss of generality is implied by restricting attention to credible message functions that send no more messages than the number of states $|\Omega|$ plus one. And, no loss of optimality is implied by restricting attention to credible message functions that send no more messages than the number of states $|\Omega|$.

Proposition 1 ***(i)*** *For any credible message function, there exists another credible message function that generates an identical ex-ante expected payoff to Sender and employs no more than $|\Omega| + 1$ messages.* ***(ii)*** *For any credible message function, there exists another credible message function that generates a weakly higher ex-ante expected payoff to Sender and employs no more than $|\Omega|$ messages.*

The proof of part (i) of the proposition starts with the observation that by Carathéodory's Theorem (Rockafellar, 1997), for any message function there exists another (possibly non-credible) message function that generates an identical ex-ante expected payoff with no more than $|\Omega| + 1$ messages. The challenge is to show that given a *credible* message function that employs more messages, it is possible to reduce the number of messages in such a way that preserves credibility. To prove this, we show that in the process of reducing the number of messages, it is never the case that a message that was not sent in state ω under the original message function is sent in ω under the message function with the smaller number of messages. The proof of part (ii) of the proposition relies on the observation that the expected payoff that is generated by a credible message function that employs $|\Omega| + 1$ messages can be written as an average of the expected payoffs generated by two *credible* message functions that each employs no more than $|\Omega|$ messages. The argument that ensures credibility is the same as in part (i).

Corollary 1 *If there exists an optimal solution to Sender's problem (SP), then there exists a message function that attains the maximal ex-ante expected payoff to Sender and employs no more than $|\Omega|$ messages. Otherwise, it is possible to approximate the upper bound on the ex-ante expected payoff to Sender with a message function that employs no more than $|\Omega|$ messages.*[12]

[11] Consider the following example. Suppose that $\Omega = \{0,1\}$ and $\pi_0 = \pi_1 = \frac{1}{2}$. The set of receiver's actions is given by $A = \{a_1, a_2, a_3\}$. Sender's and receiver's payoffs are $u_S(a_1,\omega) = u_S(a_3,\omega) = 1$ and $u_S(a_2,\omega) = 0$, and $u_R(a_1,\omega) = \frac{1}{3} - \omega$, $u_S(a_2,\omega) = 0$, and $u_R(a_1,\omega) = \omega - \frac{2}{3}$, respectively. It is possible for Sender to achieve the expected payoff $\frac{2}{3}$ with three messages, but not with two.

[12] Optimal credible message functions might not exist for some specifications of u_R and u_S. However, the ex-

3 Credible Bayesian Persuasion

In this section we incorporate our notion of costly deception into a simple model of Bayesian persuasion (Kamenica and Gentzkow, 2011, hereafter KG). For simplicity of exposition, we impose the following three assumptions. First, we assume that the payoff-relevant part of the state ω is a real-valued random variable. Next, we assume that Receiver's optimal action depends only on the expected state. That is, given a posterior belief p the optimal action $a^*(p)$ depends only on the mean of p, denoted $\mu_p \equiv \mathbb{E}_p[\omega]$. Finally, we assume that Sender's preferences over Receiver's actions do not depend on the state.[13]

To facilitate the comparison between our model and that of KG, we start by writing Sender's problem as a constrained maximization problem over distributions of posterior beliefs, rather than message functions.

Recall that, given a message function σ, every message m that is sent in σ induces a posterior belief p_m^σ over the payoff relevant part of the state ω. Accordingly, the message function σ induces a distribution over posterior beliefs. We denote such a distribution of posterior beliefs by $\tau \in \Delta(\Delta(\Omega))$, and the probability that τ induces a posterior $p \in \Delta(\Omega)$ by $\tau(p)$. Thus:

$$\tau(p) = \sum_{\{m: p_m^\sigma = p\}} \sum_{\omega \in \Omega} q^\sigma(m, \omega)\, \pi(\omega).$$

A distribution of posterior beliefs τ is said to be *Bayes plausible* if the expected posterior belief it induces is equal to the prior. As famously shown by (KG) and Aumann and Maschler (1995), a distribution over posterior beliefs τ can be induced by some message function σ if and only if τ is Bayes plausible. We can therefore rewrite Sender's problem (SP) as follows:

$$\max_{\tau} \quad \sum_{p \in Supp(\tau)} \hat{u}_S(p) \cdot \tau(p) \tag{SP1}$$

$$\text{s.t.} \quad \sum_{p \in Supp(\tau)} p \cdot \tau(p) = \pi \tag{Bayes Plausibility}$$

$$\hat{u}_S(p) \geq \hat{u}_S(p') - \alpha \cdot d(p, p'), \quad \forall p, p' \in \text{Supp}(\tau) \tag{Credibility}$$

where $\text{Supp}(\tau)$ denotes the support of τ. A distribution of posterior beliefs τ that is Bayes plausible and credible is said to be *feasible*. Note that the "standard" problem of Bayesian persuasion involves maximizing the same objective function, under the same Bayes plausibility constraint. The new component that is introduced in our costly deception framework is

ante expected payoff to Sender is bounded. Therefore, if a credible optimal message function does not exist, then there exists a sequence of credible message functions that employ no more than $|\Omega|$ messages and that generate an ex-ante expected payoff to Sender that converges to this upper bound.

[13] Kamenica and Gentzkow (2011) refer to this case as one in which the "sender's payoff depends only on the expected state." This holds, for example, if $u_R(a, \omega) = -(a - \omega)^2$ and $u_S(a, \omega) = a$. It is easy to verify that in this case $a^*(p) = \mu_p$ and Sender's payoff from inducing the belief p is therefore $\hat{u}_S(p) = \mu_p$.

the credibility constraint.

We now proceed to characterize the solution to Sender's problem. Given Sender's indirect payoff function $\hat{u}_S$, the convex hull of the graph of $\hat{u}_S$, denoted $\mathrm{co}(\hat{u}_S)$, consists of all the convex combinations of elements in the graph of $\hat{u}_S$. That is,

$$\mathrm{co}(\hat{u}_S) = \left\{ \begin{array}{ll} (p, y): & \exists p_1, \ldots, p_k, p_i \in \Delta(\Omega) \text{ for all } i, \text{ and } \exists \lambda_1, \ldots, \lambda_k \geq 0, \sum_{i=1}^k \lambda_i = 1 \\ & \text{such that } p = \sum_{i=1}^k \lambda_i p_i \text{ and } y = \sum_{i=1}^k \lambda_i \hat{u}_S(p_i) \end{array} \right\}.$$

Given $\alpha \geq 0$, we define the set $\mathrm{co}_\alpha(\hat{u}_S)$ similarly to $\mathrm{co}(\hat{u}_S)$, with one difference: it consists of all the convex combinations of elements in the graph of $\hat{u}_S$ that satisfy an additional set of pairwise restrictions that are parametrized by α:

$$\mathrm{co}_\alpha(\hat{u}_S) = \left\{ \begin{array}{ll} (p, y): & \exists p_1, \ldots, p_k, p_i \in \Delta(\Omega) \text{ for all } i, \text{ and } \exists \lambda_1, \ldots, \lambda_k \geq 0, \sum_{i=1}^k \lambda_i = 1 \\ & \text{such that } p = \sum_{i=1}^k \lambda_i p_i \text{ and } y = \sum_{i=1}^k \lambda_i \hat{u}_S(p_i), \text{ and} \\ & \frac{|y_j - y_i|}{d(p_j, p_i)} \leq \alpha \text{ for every } i, j \end{array} \right\}.$$

with the convention that $\frac{0}{0} = 0$, so that if $y = \hat{u}_S(p)$ then $(p, y) \in \mathrm{co}_\alpha(\hat{u}_S)$ for all $\alpha \geq 0$.

If $\hat{u}_S(p) = y$ then we say that p is the underlying posterior belief that induces y. The set $\mathrm{co}_\alpha(\hat{u}_S)$ contains all the pairs (p, y) for which the value y can be achieved by randomization over indirect payoffs that are in the graph of $\hat{u}_S$, provided that: (i) the weights of the randomization are such that the associated underlying posteriors average to p, and (ii) the randomization does not involve indirect payoffs whose difference, divided by the distance between their underlying posteriors, is "too large" (i.e. exceeds α), which would make deception attractive to Sender.

Given $\alpha \geq 0$, define the *value* of belief p as follows:

$$V(p, \alpha) \equiv \sup\left\{y : (p, y) \in \mathrm{co}_\alpha(\hat{u}_S)\right\}.$$

If $(\pi, y) \in \mathrm{co}_\alpha(\hat{u}_S)$ then, by definition, there exists a Bayes plausible distribution τ (namely, a collection $p_1, \ldots, p_k$, such that $p_i \in \Delta(\Omega)$ for all i, and probabilities $\lambda_1, \ldots, \lambda_k$ such that $\sum_{i=1}^k \lambda_i p_i = \pi$) that is credible and induces the expected payoff y. Furthermore, given π, if y can be induced by some Bayes plausible and credible distribution τ then $(\pi, y) \in \mathrm{co}_\alpha(\hat{u}_S)$. The next result follows immediately:

Proposition 2 *For every $\alpha \geq 0$, the highest value that Sender can achieve in the problem (SP1) is given by $V(\pi, \alpha)$.*[14]

It may be naively expected that lower deception costs are better for Sender because they

[14]When the set $\left\{y : (p, y) \in \mathrm{co}_\alpha(\hat{u}_S)\right\}$ does not constain its supremum, by "achieve" we mean that it is possible to approximate $V(\pi, \alpha)$ arbitrarily closely.

make it easier for Sender to deceive in a way that benefits her. But deception never happens in equilibrium in our model, and lower deception costs weaken Sender's ability to commit, which hurts Sender. In fact, higher deception costs only expands the domain of message functions that are deemed credible, from which Sender can pick her preferred one. Thus, higher deception costs are always *beneficial* for Sender (indeed, if $\alpha < \alpha'$ then $\mathrm{co}_{\alpha}(\hat{u}_S) \subseteq \mathrm{co}_{\alpha'}(\hat{u}_S)$ which implies that $V(\pi,\alpha) \leq V(\pi,\alpha')$ for any prior π). If deception costs are sufficiently high, then the credibility constraint becomes non-binding and Sender's problem becomes identical to that of the Bayesian persuader of KG.[15] The next corollary summarizes these observations:

Corollary 2 *For any prior π, Sender's value is weakly increasing in the deception cost parameter α. For high enough α, Sender's value becomes identical to that of the Bayesian persuader of KG.*

The structure of the set $\mathrm{co}_{\alpha}(\hat{u}_S)$ depends on the distance function d. To proceed, we assume that the distance between any two beliefs $p, p' \in \Delta(\Omega)$ is measured by the difference between the means induced by these distributions, that is:

$$d\left(p,p'\right) = \left|\mu_p - \mu_{p'}\right|. \tag{3}$$

Since Receiver's action and Sender's payoff depend only on the expected state, we slightly abuse notation and write $\hat{u}_S(\mu_p)$ instead of $\hat{u}_S(p)$. Thus, the credibility constraint in Sender's problem (SP1) can be rewritten as follows:

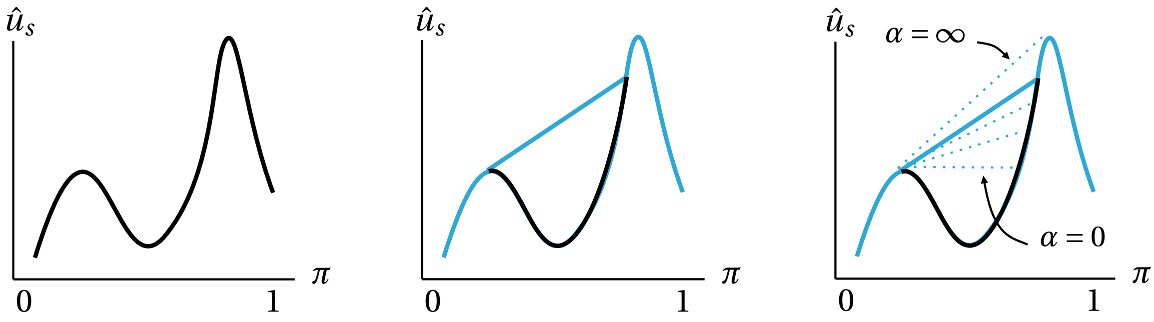
$$\left|\frac{\hat{u}_S(\mu_p) - \hat{u}_S(\mu_{p'})}{\mu_p - \mu_{p'}}\right| \leq \alpha. \tag{4}$$

It follows that for any two posterior beliefs that Sender induces in equilibrium, it must be the case that the gain from deviating from one posterior belief to the other, divided by the distance between the means of the two posteriors, does not exceed α.

Example 1. Suppose that the payoff relevant part of the state space is binary, with $\Omega = \{0,1\}$. In this case, a distribution p over Ω can be represented by the probability q that the state is $\omega = 1$, and the mean of p is given by $\mu_p = q$.

Condition (4) has a geometric interpretation. To see it, consider the indirect payoff function $\hat{u}_S$ that is depicted in Figure (1a). Suppose that the prior distribution is given by some $\pi \in [0,1]$. In Bayesian persuasion with full commitment ($\alpha = \infty$) Sender optimizes by "spliting" π into two probabilities, q and q', that are such that $\lambda \cdot q + (1-\lambda) \cdot q' = \pi$ for some $\lambda \in [0,1]$ (Bayes plausibility) so as to maximize the value of the objective function $\lambda \cdot \hat{u}_S\left(q\right) + (1-\lambda) \cdot \hat{u}_S\left(q'\right)$. The

[15] To see this, note that $\mathrm{co}_{\infty}(\hat{u}_S) = \mathrm{co}(\hat{u}_S)$ and thus $V(\pi,\infty) = \sup\left\{y : \left(p,y\right) \in \mathrm{co}(\hat{u}_S)\right\}$, which is exactly the value of Sender's problem in KG, for any prior π.



(a) $\hat{u}_S$ (b) $V(\pi,\alpha)$ for fixed α (c) $V(\pi,\alpha)$ with $\alpha \in [0,\infty]$

Figure 1: A Geometric Illustration of Credibility

credibility constraint (4) implies that the *slope* of the line that connects the payoffs associated with these two probabilities, $\hat{u}_S(q)$ and $\hat{u}_S(q')$, cannot exceed α.

Figure (1b) depicts Sender's value $V(\pi,\alpha)$ (in Blue) for a fixed deception cost α and different values of π. Note that the graph of this function comprises of parts that coincide with the graph of $\hat{u}_S$ and parts that are line segments between points on the graph of $\hat{u}_S$ with slope that does not exceed α.

Figure (1c) illustrates what happens to $V(\pi,\alpha)$ when α is varied between zero and infinity. The uppermost dotted line in the figure corresponds to the graph of V when deception costs are infinite, i.e., $\alpha = \infty$ (or are just high enough to be non binding). On the other extreme, the flat dotted line corresponds to the case where deception is costless. In that case Sender can only induce posterior beliefs that have identical indirect payoffs. This is the case that is analyzed by Lipnowski and Ravid (2020). The dotted lines in between correspond to different values of α; higher lines correspond to higher values of α. ■

3.1 Continuity and Discontinuity of Sender's Value Function

We now turn to discuss the continuity of the value function $V(\pi,\alpha)$. When V is discontinuous, Sender's expected payoff is highly sensitive to small changes in the prior beliefs and/or deception costs α.

Our first result shows that continuity of Sender's indirect payoff function implies continuity of her value function. The challenge in proving this result is to overcome the fact that the correspondence that maps the parameters (π,α) into the set of feasible distributions τ is not lower hemi-continuous (and therefore Berge's Maximum Theorem does not apply in our case). This implies that for a given distribution of posterior beliefs τ, a small change in α may imply that there is no feasible distribution of posterior beliefs in the neighborhood of τ. The next example 2 illustrates this difficulty.

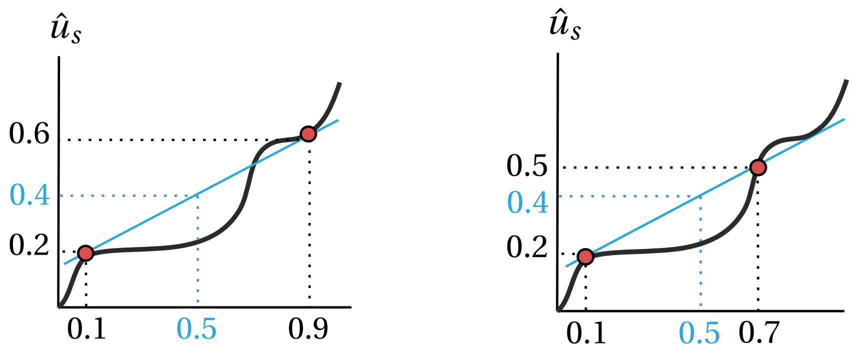


Figure 2: Continuity of the Distribution of Posteriors

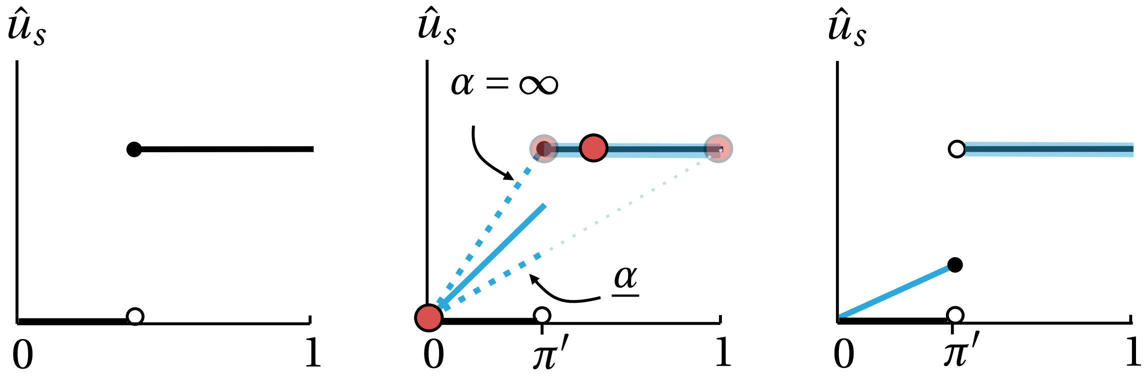
Example 2. Consider an example with two states $\Omega = \{0,1\}$ with equal prior probabilities $\pi = (\frac{1}{2}, \frac{1}{2})$, and where $\alpha = \frac{1}{2}$. Suppose that the function $\hat{u}_S$ is given by the bold curved line depicted in Figure (2a). The optimal credible distribution of posterior beliefs in this example is $\tau = (0.1, 0.9; \frac{1}{2}, \frac{1}{2})$ and it gives Sender an expected value that is equal to 0.4. To see that credibility is satisfied, notice that the slope of the line that connects the two points $(0.1, 0.2)$ and $(0.9, 0.6)$ (depicted in light blue) is $\frac{1}{2}$, which is smaller than or equal to α.

Suppose now that α is slightly decreased. Observe that it is impossible to find two posterior beliefs close to 0.1 and 0.9, respectively, that satisfy the credibility constraint (i.e., such that the line that connects the two points associated with these posteriors has a slope smaller than or equal to the new value of α, which is smaller than $\frac{1}{2}$). Thus, the correspondence that maps the parameters (π, α) into the set of feasible distributions is not lower hemi-continuous at $(\pi, \alpha) = ((\frac{1}{2}, \frac{1}{2}), \frac{1}{2})$. ■

To overcome this difficulty, we show that even if a feasible distribution τ is such that for some small change in (π, α) there is no feasible distribution that is close to τ, then there must exist another feasible distribution $\hat{\tau}$, that achieves the same expected value for Sender as τ, and $\hat{\tau}$ is such that for small changes in (π, α) there is a feasible distribution that is close to $\hat{\tau}$.

Example 2 (continued). As illustrated in Figure (2b), there exists a feasible distribution $\hat{\tau} = (0.1, 0.7; \frac{1}{3}, \frac{2}{3})$ that generates the same expected value for Sender of 0.4 as $\tau = (0.1, 0.9; \frac{1}{2}, \frac{1}{2})$. Note that for any parameters (π', α') that are close to $(\pi, \alpha) = ((\frac{1}{2}, \frac{1}{2}), \frac{1}{2})$, there exists a distribution over posteriors that is feasible with respect to (π', α') and is close to $\hat{\tau}$. For example, it is possible to pick a binary distribution of posterior beliefs that is supported on 0.1 and $0.7 - \varepsilon$ for some small $\varepsilon > 0$ that depends on α' and the curvature of the function $\hat{u}_S$. ■

We thus obtain the following result:



(a) (b) (c)

Figure 3

Proposition 3 *If $\hat{u}_S$ is a continuous function then $V(\pi, \alpha)$ is a continuous function in both π and α.*[16]

The opposite is not true in general: discontinuity of the indirect payoff function $\hat{u}_S$ does not *necessarily* imply that the function $V(\pi, \alpha)$ is also discontinuous.[17] However, in many cases, a discontinuity in $\hat{u}_S$ does imply a discontinuity of $V(\pi, \alpha)$ in both α and the prior π. Lipnowski, Ravid and Shishkin (2020) interpret the discontinuity in α as a "collapse of trust." We illustrate this discontinuity in α in the next example.

Example 3. Suppose that $\Omega = \{0, 1\}$. Consider an indirect sender's payoff function $\hat{u}_S$ such as the one depicted in Figure (3a). Figure (3b) depicts the expected payoff $V(\pi, \alpha)$ for a given value of α (in solid Blue). Notice that for this value of α, $V(\pi, \alpha)$ is discontinuous in the prior π. As α decreases, the function $V(\pi, \alpha)$ decreases with it for any $\pi \in [0, \pi']$. Once α drops below $\underline{\alpha}$, the function V coincides with $\hat{u}_S$, and so exhibits a discontinuity in α for every $\pi \in [0, \pi']$. ■

Example 3 shows that small changes in prior beliefs may imply large differences in the expected payoff of Sender. This type of discontinuity stands in contrast with the continuity of the value function in standard Bayesian persuasion (which is equivalent to the case where $\alpha = \infty$). This is because $V(\pi, \infty)$ is the lowest concave function that is above $\hat{u}_S$, and is therefore continuous.[18]

[16] A sufficient condition for the function $\hat{u}_S(p)$ to be continuous is that the action set A is a convex subset of $\mathbb{R}$ and $u_R(a, \omega)$ is strictly concave in a for every ω. In this case, for any posterior belief p the function $\mathbb{E}_p[u_R(a, \omega))]$ is strictly concave in a and so has a unique maximizer. Therefore, by the Theorem of the Maximum, $a^*(p)$ is continuous in p which implies that $\hat{u}_S(p)$ is continuous in p.

[17] For example, suppose that $\Omega = \{0, 1\}$ and the function $\hat{u}_S$ is such that $\hat{u}_S(q) = 1$ for $q \in [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ and $\hat{u}_S(q) = 0$ for $q \in (\frac{1}{3}, \frac{2}{3})$. I.e., the function $\hat{u}_S(\pi)$ is discontinuous. However, for any prior $q \in [0, 1]$, there exists a Bayes plausible distribution of posterior beliefs τ that is supported on the posterior beliefs 0 and 1 that is credible for any $\alpha \geq 0$. It therefore follows that $V(q, \alpha) = 1$ for every $q \in [0, 1]$ and $\alpha \geq 0$.

[18] Since the action space A is compact and the functions u_R and u_S are continuous, the function u_S is upper hemi-continuous, which implies that its concave closure V is continuous when $\alpha = \infty$.

Remark. The particular form of discontinuity in α described in Example 3 hinges on our assumption that, when indifferent, Receiver breaks ties in favor of Sender. The same type of discontinuity would also arise for any tie-breaking rule in which, when indifferent, Receiver picks an action that gives Sender a fixed fraction of the available surplus (for example, if Receiver picks the worst possible action for Sender, or randomizes between the best and worst actions for Sender with a fixed probability). However, notice that it is possible to restore continuity by using a more sophisticated tie-breaking rule in which, when indifferent, Receiver picks the best possible action for Sender, subject to the credibility constraint. To see this, suppose that in Example 3 above, when Receiver's posterior belief on state 1 is $q = \pi'$, Receiver mixes between the best and worst actions for Sender with probabilities $\pi'\alpha$ and $1-\pi'\alpha$, respectively. With this tie-breaking rule, the function $V(\pi,\alpha)$ would be continuous in α (but not in π). This is because for any prior $\pi \in [0,\pi']$, Sender would induce credible beliefs $q = 0$ and $q = \pi'$, with expected payoffs to Sender of 0 and $\alpha\pi'$, respectively, as depicted in Figure (3c).

We conclude this section with the following observation.

Proposition 4 *If $\hat{u}_S$ is Lipschitz continuous with constant K, then the credibility constraint* (4) *is never binding for any deception cost α larger than K, and it is then possible to implement the KG solution for Bayesian persuasion with full commitment.*

The proposition follows immediately from the definition of Lipschitz continuity (proof is omitted). Intuitively, credibility constrains the slope $\left|\frac{\hat{u}_S(\mu_p)-\hat{u}_S(\mu_{p'})}{\mu_p-\mu_{p'}}\right|$ for any two posterior beliefs p and p' in the support of the distribution τ. It therefore follows that if the slope of the function $\hat{u}_S$ is bounded below some constant K, then whenever the coefficient α is larger than K credibility is not binding.

3.2 The Effect of the Cost of Deception

As the cost of deception α decreases, the set of credible message functions for Sender shrinks. Sender can restore her credibility by either adopting a message function in which deception is more costly, or by adopting a message function in which the gain from deception is smaller. In this subsection we discuss these two strategies.

The next example shows how Sender can increase the cost of deception in response to a lower value of α by moving the means of the induced posterior beliefs farther apart.

Example 3 (continued). As α in Figure (3b) is lowered, the posterior beliefs that support the optimal distribution τ move farther apart. Intuitively, this movement increases the cost of deception and so restores the credibility of Sender's message function. This movement has the effect of "ungarbling" Sender's communicated information relative to the optimally

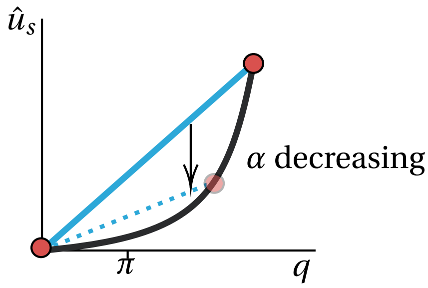


Figure 4: Convex $\hat{u}_S$

induced posteriors under full commitment. This ungarbling allows Receiver to make a more informed choice and so increases Receiver's ex-ante expected payoff. ■

This result is in the same spirit of what Lipnowski, Ravid and Shishkin (2020) refer to as "productive mistrust." Namely, a decrease in Sender's ability to commit implies that the equilibrium is more informative and consequently Receiver is made better off.

The other way in which Sender can respond to a decrease in the value of α is by decreasing the gain from deception. This can sometimes be achieved by additional garbling of prior beliefs, through moving the means of the induced posteriors closer together. Whether or not garbling or ungarbling is better for Sender depends on the specific context.

The next proposition describes a sufficient condition that ensures that Sender responds to a lower value of α by garbling her message to Receiver.

Proposition 5 *Suppose that the state space Ω is binary and Sender's indirect payoff function $\hat{u}_S$ is convex but not linear.*[19] *If $\alpha' > \alpha$, then Sender's optimal distribution of posterior beliefs under α is a garbling of the optimal distribution under α'. Consequently, lower deception costs are weakly harmful for both Sender and Receiver.*

Figure (4) depicts the case of a convex indirect payoff function $\hat{u}_S$ and illustrates that a lower α results in a distribution τ that is supported on posterior beliefs that are closer together.

Broadly speaking, a lower cost of deception implies it is more difficult for Sender to commit and so is accompanied by a higher level of mistrust. To appreciate the effect of mistrust it is useful to observe that Sender faces a tension between his incentive to reveal and conceal information to Receiver. Recall that Receiver *always* prefers all information to be revealed. Example 3 depicts a situation where Sender's and Receiver's interests are sufficiently opposed for Receiver to benefit from Sender's difficulty to commit. Proposition 5 depicts a situation where Sender's and Receiver's interests are sufficiently aligned for both Sender and Receiver

[19] If Sender's indirect payoff function is linear, then a message function that induces a single posterior belief that is equal to the prior is optimal.

to suffer from Sender's difficulty to commit. In the former case, Sender prefers to not disclose all the available information, and in order to preserve her credibility, she has to disclose more than she would want to if she was trusted by Receiver; in the latter case, Sender prefers to fully disclose all the available information, and in order to preserve her credibility, she has to disclose less than she would want to if she was trusted by Receiver. Notice however that in the two extreme cases of diametrically opposed and mutual Sender's and Receiver's interests with respect to the revelation of information, a change in the value of α makes no difference. In the case of opposing interests, silence on Sender's part is always credible. And, when Sender and Receiver have mutual interests, Sender would anyway not want to mislead Receiver.

4 Credible "Cheap Talk"

In the previous section we considered a model in which Sender's payoff is independent of the state of the world. In this section, we consider a model in which Sender's payoff does depend on the state. To that end we focus our attention a finite version of Crawford and Sobel's (1982) classic uniform-quadratic example.

Suppose that the set of states is given by $\Omega_N = \left\{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1\right\}$ for some large N, with a uniform prior distribution. The set of Receiver's actions is given by $A = \mathbb{R}$. Sender and Receiver's payoff functions are given by $u_S(a,\omega) = -(a-(\omega+b))^2$ and $u_R(a,\omega) = -(a-\omega)^2$, respectively, for some $b \geq 0$.

We retain the assumption that the distance between any two beliefs $p, p' \in \Delta(\Omega_N)$ is given by the difference between the means of these two beliefs, that is $d(p,p') = |\mu_p - \mu_{p'}|$. Hence, the cost of deception, namely the cost of inducing belief p' instead of p, is given by $\alpha \cdot |\mu_p - \mu_{p'}|$. When $\alpha = 0$, our model coincides with that of Crawford and Sobel (1982).

Sender chooses a set of messages $M = \left\{m_1, \ldots, m_J\right\}$ and a message function $\sigma : \Omega_N \to M$. For simplicity, we restrict our attention to pure strategy message functions with full support on the set of messages. This allows us to identify a message function with the partition it induces over the set of states Ω_N. We also identify each message m with the set of states where message m is sent, $m \equiv \{\omega : \sigma(\omega) = m\}$.

We denote the number of elements in the set m by $|m|$. The probability of sending message m is denoted $\rho(m) = \frac{|m|}{N+1}$. For simplicity, we assume that each message contains at least two states [20] The mean of message m is denoted $\mu_m \equiv \mathbb{E}[\,\omega \mid \omega \in m]$. We denote by $\underline{m}$ and $\overline{m}$ the smallest and largest states in m, respectively.

Receiver observes the message sent by Sender and chooses an action $a \in A$. Denote

[20] Because N can be chosen to be arbitrarily large, this does not impose a positive lower bound on the probability of each message. That is, it does not constrain the "fineness" of the message functions we consider. We make use of this assumption in the proof of Proposition 6 below.

Receiver's posterior belief over the state of the world, after observing the message m, by $p_m \in \Delta(\Omega)$. Given Sender's message strategy σ, the belief p_m is computed using Bayes rule as follows:

$$p_m[\omega] = \begin{cases} \frac{1}{\rho(m)} & \text{if } \omega \in m \\ 0 & \text{if } \omega \notin m \end{cases}$$

A simple calculation shows that if Sender employs the message function σ, then the action that maximizes Receiver's payoff, following message m, is given by the mean of m, i.e.:

$$a^*(m) = \mu_m.$$

Thus, the expected payoff to Sender from employing the messages function σ is given by:

$$-\sum_{m \in M} \rho(m) \operatorname{Var}[\omega | \omega \in m] \tag{5}$$

up to a constant.[21] The objective of Sender is, therefore, to find the credible message function σ that maximizes (5).

Given a message function σ, we order the messages according to the conditional means they induce, and denote the k^{th} message by m_k and its mean by μ_k. If two messages induce an identical expectation then they can be merged into one message without affecting credibility or the value of the objective function. Thus, no loss of generality is implied by assuming that $\mu_1 < \cdots < \mu_J$ (where the total number of messages, J, is less than N by Corollary 1).

Under a credible message function, type $\omega \in m_k$ of Sender prefers sending message m_k to sending any other message. In particular, she prefers sending m_k to sending m_{k+j} with $j \geq 1$:

$$-\left(\omega - \mu_k + b\right)^2 \geq -\left(\omega - \mu_{k+j} + b\right)^2 - \alpha\left(\mu_{k+j} - \mu_k\right). \tag{6}$$

The left-hand side of (6) is type ω's payoff from sending the message m_k, after which Receiver takes the action μ_k. The right-hand side is type ω's payoff from sending the message m_{k+j}, inducing Receiver's action μ_{k+j} but suffering deception cost of $\alpha \cdot \left(\mu_{k+j} - \mu_k\right)$.

Rewriting (6) yields:
$$\frac{\mu_k + \mu_{j+k}}{2} - \omega_k \geq b - \frac{\alpha}{2}$$

for all $\omega \in m_k$. Thus, a necessary and sufficient condition that ensures that any type $\omega \in m_k$ prefers reporting m_k to any other message m_{k+j} with $j \geq 1$ is the following incentive compatibility constraint:

$$\frac{\mu_k + \mu_{k+1}}{2} - \overline{m}_k \geq b - \frac{\alpha}{2}. \tag{ICup(k)}$$

The constraint ICup(k) is said to be *binding* if it is satisfied, but would have been violated

[21]Given a message function σ, Sender's expected payoff $-\sum_{m \in M} \rho(m) \mathbb{E}\left[(a^*(m) - \omega - b)^2 | \omega \in m\right]$ is equal to $-\sum_{m \in M} \rho(m) \mathbb{E}\left[(a^*(m_i) - \omega)^2 | \omega \in m\right]$ up to a constant that is independent of σ. This last expression is equal to the expected induced variance (5) and also (by definition) to Receiver's expected payoff.

if another message m, which is different from both m_k and m_{k+1}, had been sent in state $\overline{m}_{k+1}$.

An analogous argument shows that a necessary and sufficient condition that ensures that all types $\omega_k \in m_k$ prefer reporting m_k to any other message m_{k-j} with $j \geq 1$ is:

$$\frac{\mu_{k-1}+\mu_k}{2} - \underline{m}_k \leq b + \frac{\alpha}{2}. \qquad \text{ICdown(k)}$$

A partition that satisfies the incentive constraints ICup(k) and ICdown(k), for all $k \geq 1$, is said to be a *credible partition.* A credible partition that maximizes Sender's objective function (5) is said to be *optimal.*

Inspection of the IC constraints and objective function (5) reveals the following result:

Lemma 1 *If $\alpha \geq 2b$ then a partition of Ω_N into singletons is optimal for Sender for any N.*

Notice that a partition of Ω_N into singletons is equivalent to a message function that fully reveals the state ($i.e., \sigma(\omega) = \omega$). Lemma 1 implies that we may hereafter restrict our attention to the case where $\alpha < 2b$. Another consequence of the IC constraints is the following:

Lemma 2 *The number of messages in a credible partition is bounded from above by $1/(2b-\alpha)$.*

We proceed with the following definition.

Definition 2 *A message m_k is said to be convex if for every three states $\omega < \omega' < \omega''$, if $\omega, \omega'' \in m_k$, then also $\omega' \in m_k$. A partition of Ω_N into convex messages is said to be a convex partition.*

Crawford and Sobel (1982) famously showed that *any* equilibrium of the cheap talk model induces a convex partition. In our model this result no longer holds. In fact, *any* partition is credible for values of α that are sufficiently high. This implies the following two observations:

(1) In Crawford and Sobel (1982), if type ω is indifferent between two messages m, m' with $\mu_m < \mu_{m'}$, then every type $\omega' > \omega$ strictly prefers m' over m and every type $\omega'' < \omega$ strictly prefers m over m' (this is a consequence of the assumption that Sender's preferences satisfy the single crossing property). In contrast, in our setting, because the cost of switching to a different message is endogenous and depends on type's equilibrium message, it is possible to have two types $\omega < \omega'$ such that ω' prefers m over m' but ω prefers m' over m.

(2) In Crawford and Sobel (1982) the first element of the partition determines the entire partition structure. This is because the structure of the partition is determined by a set of types who are indifferent between pairs of contiguous elements in the partition. In contrast, in our case, even if we restrict our attention to only convex partition structures, then many more convex partitions are possible. Specifically, fixing the first element of the partition does not

pin down the next elements of the partition. Moreover, indifference conditions are not a necessary feature of the partition. Namely, it is possible to have convex partitions in which no type is indifferent between any pair of messages.[22]

We show that Sender's optimal message function σ that minimizes the weighted variance (5) induces a convex partition of the set Ω_N for large enough N. The main challenge is that, given a credible message function, it is difficult to find a credible "global" improvement for it. And, "local" improvements may violate credibility. Our approach is to perform a sequence of local improvements that converge to a convex partitional structure while correcting for violations of credibility along the way.

The next definition formalizes a notion of a partially convex message function. It is instrumental in describing the way in which a given message function σ is iteratively transformed through a sequence of steps, parametrized by k, into a fully convex message function that induces a convex partitional structure on a subset of low states.

Definition 3 *A partition of Ω_N into messages is said to be "tightly packed with k messages" on a set $\{0, \frac{1}{N}, \ldots, l\}$ if:*

1. *The union of the first k messages covers $\{0, \frac{1}{N}, \ldots, l\}$, i.e. $\cup_{j=1}^{k} m_j = \{0, \frac{1}{N}, \ldots, l\}$;*
2. *Each message m_j, $j \le k$, is convex; and*
3. *The incentive constraints $ICup(1), \ldots, ICup(k-1)$ are all binding.*

The next lemma characterizes the maximal number of messages that can be tightly packed into a set $\{0, \frac{1}{N}, \ldots, l\}$.

Lemma 3 *Given a length $l \in \Omega_N$, there exists a number $\hat{N}$ such that for all $N > \hat{N}$, the maximal number of messages that can be tightly packed into the set $\{0, \frac{1}{N}, \ldots, l\}$ is given by $I(l) \equiv \left\lceil \sqrt{\frac{1}{4} + \frac{l}{2b-\alpha}} - \frac{1}{2} \right\rceil$.*[23]

Inspection of the proof of Lemma 3 reveals that if two partitions are tightly packed on the set $\{0, \frac{1}{N}, \ldots, l\}$ and have the same number of elements on this set, then they coincide on this set. Therefore, there is a unique partition with $I(1)$ elements on the set $\{0, \ldots, 1\}$. As expected, if $\alpha = 0$ then $I(1)$ is also the number of intervals in the most informative equilibrium identified in the uniform quadratic example in Crawford and Sobel (1982).

The next proposition describes the optimal partition for Sender.

[22] We note that this last observation is not a consequence of the fact that we consider a discrete version of Crawford and Sobel's model, and would persist even if we let the set of states Ω be a continuum.

[23] The function $\lceil x \rceil$ denotes the smallest integer larger than or equal to x.

Proposition 6 *There exists a number $\hat{N}$ such that for all $N > \hat{N}$, the optimal partition of Ω_N consists of $I(1) = \left\lceil \sqrt{\frac{1}{4} + \frac{1}{2b-\alpha}} - \frac{1}{2} \right\rceil$ tightly packed messages on Ω_N.*

To prove Proposition 6 we provide an iterative convergent algorithm that improves upon any credible partition that does not partition the set Ω_N into $I(1)$ tightly packed messages. We describe the algorithm in the text and defer the detailed proof to the appendix.

Start with a credible partition that does not consist of $I(1)$ tightly packed messages on Ω_N. Let k be the highest index for which the messages $m_1, \ldots, m_{k-1}$ are tightly packed on the set $\{0, \ldots, l_{k-1}\}$ where $l_j \equiv \frac{1}{N}\left(|m_1| + \cdots + |m_j|\right)$ for any $j > 0$. Figure (5a) illustrates such a partition (notice that messages m_k, m_{k+1}, m_{k+2} are not convex). If no such collection of messages exists, then $k = 1$. And if all the messages are already tightly packed, but the number of messages is smaller than $I(1)$, then k is equal to the number of messages in that partition.

Algorithm Convexify and repack

Require: Messages $m_1, ..., m_{k-1}$ are convex and tightly packed on $\{0, \ldots, l_{k-1}\}$

Part I - Convexify message m_k to the Left

while message m_k is not convex and adjacent to message m_{k-1} **do**
  let ω be the smallest state in m_k for which $(\omega - \frac{1}{N}) \in m_j$ for some m_j with $j > k$
  "swap" states ω and $\omega - \frac{1}{N}$ between messages m_k and m_j as follows:
    reassign state ω from message m_k into message m_j, and
    reassign state $\omega - \frac{1}{N}$ from message m_j into message m_k;
end while

Part II - Repack

Repartition $\{0, \ldots, l_k\}$ into $I(l_k)$ tightly packed messages.

The algorithm described above "packs message m_k" and produces a new partition in which $I(l_k)$ messages are tightly packed on the set of states $\{0, \ldots, l_k\}$, all the ICup constraints are satisfied and the modified partition yields a higher value of the objective function (5) to Sender, compared to the original partition.

The algorithm consists of two parts. In Part I message m_k is "convexified to the left" through a series of swaps of messages across states until message m_k is convex and placed immediately to the right of message m_{k-1}. At the end of Part I of the algorithm, the partition (i.e. message function) takes the form depicted in Figure 5(b). Intuitively, convexification to the left improves the value of the objective function because it decreases the variance of some messages while not affecting the variance of others and not affecting the probabilities with which messages are sent. However, notice that after the change $ICup(k-1)$ may no longer hold. This is because the convexification to the left of m_k decreases the mean μ_k, making it

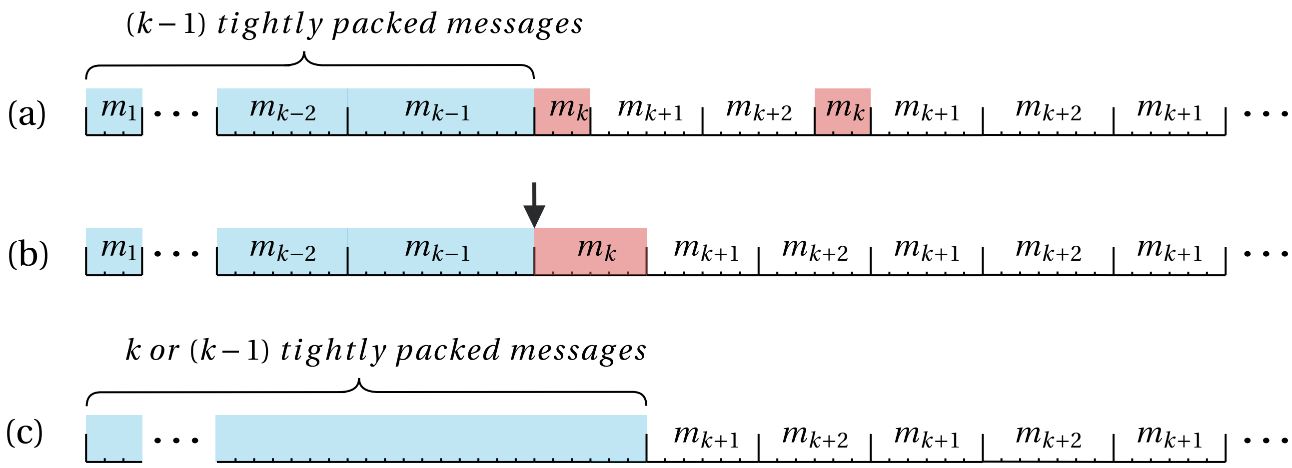


Figure 5: The Convexification of Message m_k

more attractive for higher types in m_{k-1} to deviate and report m_k. To restore incentive compatibility we proceed to the second part of the algorithm.

Part II of the Algorithm repartitions the set $\{0,\ldots,l_k\}$ into $I(l_k)$ tightly packed intervals. In the proof we show that this suffices to ensure that all the other $ICup$ constraints are also satisfied, and in particular $ICup(I(l_k))$. The result of this part is depicted in Figure 5(c). Note that it could be the case that $I(l_k) = k-1$, so that repartitioning may in fact *decrease* the total number of messages. Nevertheless, in the proof we show that the overall effect of convexifying m_k to the left and re-partitioning $\{0,\ldots,l_k\}$ improves the value of the objective function.

If the partition generated by the algorithm does not consist of $I(1)$ messages that are tightly packed on Ω_N, then we apply the algorithm again on that partition. Because in each iteration of the algorithm the cardinality of the set of states on which the message function is tightly packed strictly increases, the process converges to the partition with $I(1)$ tightly packed messages on Ω_N in a finite number of iterations.

In the proof we show that whenever all the $ICup$ constraints are binding, which is the case in any partition that consists of tightly packed messages, then all the ICdown constraints are satisfied as well. Thus, the obtained partition, in which $I(1)$ messages are tightly packed on Ω_N is credible.

We conclude this section with a characterization of the messages that are induced by the optimal partition.[24]

Corollary 3 *The optimal partition consists of $I(1)$ messages. As the number of states N tends to*

[24]The corollary is an immediate implication of the facts that $I(1)$ messages are tightly packed, and that the highest state in the $I(1)$'s message is 1, as $N \to \infty$.

infinity, message m_k, $k \in \{1,\ldots,I(1)\}$ *is sent in states* $\overline{m}_{k-1} \leq \omega \leq \overline{m}_k$ *where:*

$$\overline{m}_k = \frac{k}{I(1)} + 2\left(b - \frac{\alpha}{2}\right) k(k - I(1)).$$

Notably, the value of $\overline{m}_k$ that is described in the corollary is identical to the value characterized by Crawford and Sobel (1982), except that in the expression here, Sender's bias is offset by the cost parameter, so that instead of b in Crawford and Sobel's result, we have $b - \frac{\alpha}{2}$.

5 Conclusion

The assumption that Sender is able to commit to her strategy is crucial for Bayesian persuasion, because different messages induce different actions, and Sender must be genuinely trusted to not just send the message that induces the action that she prefers the most. We addressed the question of what happens if this assumption is not satisfied by introducing a model of information design with endogenous deception costs.

As explained, the introduction of the possibility of costly deception bridges the gap between the assumption of full commitment that is employed in the information design literature (Bergemann and Morris, 2019) and the assumption of no commitment at all (cheap talk) that is employed in the literature on communication games (Crawford and Sobel (1982) and the subsequent literature.

In many situations, an agent with superior information, but imperfect commitment power, shares this information strategically in order to influence the behavior of other agents. In some cases, this imperfect commitment is mitigated by the fact that deception is costly. The framework presented here provides a foundation for an exploration of these issues.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond. 2019. “Preferences for truth-telling.” *Econometrica*, 87(4): 1115–1153.

Aumann, Robert J, and Michael Maschler. 1995. *Repeated Games with Incomplete Information.* Cambridge, MA: MIT Press.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2007. “Guilt in games.” *American Economic Review*, 97(2): 170–176.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2009. “Dynamic psychological games.” *Journal of Economic Theory*, 144(1): 1–35.

Battigalli, Pierpaolo, and Martin Dufwenberg. forthcoming. “Belief-Dependent Motivations and Psychological Game Theory.” *Journal of Economic Literature.*

Bergemann, Dirk, and Stephen Morris. 2019. “Information Design: A Unified Perspective.” *Journal of Economic Literature*, 57(1): 44–95.

Caplin, Andrew, and John Leahy. 2004. “The supply of information by a concerned expert.” *The Economic Journal*, 114(497): 487–505.

Chen, Ying. 2011. “Perturbed communication games with honest senders and naive receivers.” *Journal of Economic Theory*, 146(2): 401–424.

Crawford, Vincent P, and Joel Sobel. 1982. “Strategic information transmission.” *Econometrica*, 50(6): 1431–1451.

Ely, Jeffrey, Alexander Frankel, and Emir Kamenica. 2015. “Suspense and surprise.” *Journal of Political Economy*, 123(1): 215–260.

Fischbacher, Urs, and Franziska Föllmi-Heusi. 2008. “Lies in disguise: An experimental study on cheating.” *Journal of the European Economic Association*, 11: 525–547.

Geanakoplos, John, David Pearce, and Ennio Stacchetti. 1989. “Psychological games and sequential rationality.” *Games and economic Behavior*, 1(1): 60–79.

Gneezy, Uri. 2005. “Deception: The role of consequences.” *American Economic Review*, 95(1): 384–394.

Guo, Yingni, and Eran Shmaya. forthcoming. “Costly Miscalibration.” *Theoretical Economics.*

Kamenica, Emir. 2019. “Bayesian Persuasion and Information Design.” *Annual Review of Economics*, 11(1): 249–272.

Kamenica, Emir, and Matthew Gentzkow. 2011. “Bayesian Persuasion.” *American Economic Review*, 101(6): 2590–2615.

Kartik, Navin. 2009. “Strategic communication with lying costs.” *Review of Economic Studies*, 76(4): 1359–1395.

Lipnowski, Elliot, and Doron Ravid. 2020. “Cheap Talk with Transparent Motives.” *Economet-*

rica, 88(4): 1631–1660.

Lipnowski, Elliot, Doron Ravid, and Denis Shishkin. 2020. “Persuasion via Weak Institutions.” *Working Paper.*

Min, Daehong. 2018. “Bayesian Persuasion under Partial Commitment.” *Working paper.*

Nguyen, Anh, and Yong Teck Tan. 2019. “Bayesian Persuasion with Costly Messages.” *Working Paper.*

Ottaviani, Marco, and Peter Norman Sørensen. 2006. “Reputational cheap talk.” *The Rand journal of economics*, 37(1): 155–175.

Perez-Richet, Eduardo, and Vasiliki Skreta. 2020. “Test Design under Falsification.” *Working paper.*

Rockafellar, Ralph Tyrell. 1997. *Convex Analysis.* Princeton University Press.

Sobel, Joel. 2020. “Lying and Deception in Games.” *Journal of Political Economy*, 128(3): 907–947.

Appendix: Proofs

Proof of Proposition 1

We begin with the proof of part (i) of the proposition. Suppose that a message function σ sends more than $|\Omega|+1$ messages with a positive probability each. Every message $m \in M$ that is sent by σ induces a posterior belief (distribution) p_m^σ over the states. This belief can be represented by a vector in $\mathbb{R}^{|\Omega|-1}$. Sender's indirect material payoff from inducing the posterior belief p_m^σ is $\hat{u}_S(p_m^\sigma)$ as in Equation (1). Thus, each message m that is sent by σ induces a vector $(p_m^\sigma, \hat{u}_S(p_m^\sigma)) \in \mathbb{R}^{|\Omega|}$.

Denote Sender's ex-ante expected payoff under σ by $U_S(\sigma)$. Then,

$$\sum_{m\in M} p^\sigma(m)\cdot(q_m^\sigma, \hat{u}_S(p_m^\sigma)) = (\pi, U_S(\sigma)) \in \mathbb{R}^{|\Omega|}$$

where $\sum_{m\in M} q^\sigma(m)\cdot p_m^\sigma = \pi \in \mathbb{R}^{|\Omega|-1}$ follows from Bayes plausibility: the mean of the induced posterior beliefs is equal to the prior belief, and $\sum_{m\in M} q^\sigma(m)\cdot \hat{u}_S(p_m^\sigma) = U_S(\sigma) \in \mathbb{R}$ by definition of $U_S(\sigma)$. Therefore, the vector $(\pi, U_S(\sigma)) \in \mathbb{R}^{|\Omega|}$ belongs to the convex hull that is generated by the set $\{(p_m^\sigma, \hat{u}_S(p_m^\sigma))\}_{m\in M}$.

By Carathéodory's Theorem (Rockafellar (1997), Theorem 17.1) it is possible to write the vector $\{(\pi, U_S(\sigma)\}$ as convex combination of no more than $|\Omega|+1$ elements in the set $\{(p_m^\sigma, \hat{u}_S(p_m^\sigma))\}_{m\in M}$.

Suppose that the messages that induce these $|\Omega|+1$ beliefs in the original message function σ are given by $m_1,\ldots,m_{|\Omega|+1}$. Consider a message function σ' that sends messages $m'_1,\ldots,m'_{|\Omega|+1}$ that induce the same posterior beliefs as those induced by $m_1,\ldots,m_{|\Omega|+1}$, with the probabilities determined by Carathéodory's Theorem. By construction, $p^{\sigma'}_{m'_j} = p^{\sigma}_{m_j}$ for $j \in \{1,\cdots,|\Omega|+1\}$. Note that the message function σ' generates the same ex-ante expected payoff to Sender as σ.

We now show that the message function σ' satisfies credibility. Observe that:

$$q^{\sigma}(m_j,\omega) = 0 \Rightarrow q^{\sigma'}(m'_j,\omega) = 0 \quad \forall j \in \{1,\cdots,|\Omega|+1\}, \forall \omega \in \Omega.$$

Because, otherwise, $q^{\sigma'}(m'_j,\omega) > 0 = q^{\sigma}(m_j,\omega)$ for some $j \in \{1,\cdots,|\Omega|+1\}$ and $\omega \in \Omega$. Then, $p^{\sigma'}_{m'_j}[\omega] > 0$ while $p^{\sigma}_{m_j}[\omega] = 0$. This is a contradiction to the fact that $p^{\sigma'}_{m'_j} = p^{\sigma}_{m_j}$ for $j \in \{1,\cdots,|\Omega|+1\}$. Thus, every belief that is induced by σ' in some state ω was also induced by σ in ω. Therefore, the credibility of σ implies the credibility of σ'.

We now turn to prove part (ii) of the proposition. Part (i) of the proposition implies that we may restrict our attention to message functions that employ no more than $|\Omega|+1$ messages.

Consider a message function σ that employs $|\Omega|+1$ messages, that induce posterior beliefs $p_1,\ldots,p_{|\Omega|+1}$ with probabilities $\lambda_1,\ldots,\lambda_{|\Omega|+1}$, respectively, such that $\sum_{i=1}^{|\Omega|+1}\lambda_i p_i = \pi$. Denote the set of these posterior beliefs by $P = \{p_1,\ldots,p_{|\Omega|+1}\}$ and denote the ex-ante expected payoff to Sender that is generated by σ by $\sum_{i=1}^{|\Omega|+1}\lambda_i \cdot \hat{u}_S(p_i) \equiv U$. We may assume that each λ_i is positive and that each p_i is different from π because otherwise it is possible to induce an ex-ante expected payoff that is at least U with no more than $|\Omega|$ messages.

We proceed with the following lemma.

Lemma A.1 *Suppose that $S = \{x_1,\ldots,x_{d+2}\}$ is a set of $d+2$ vectors in $\mathbb{R}^d$. For any vector $p \in \mathbb{R}^d$ in the convex hull generated by S, denoted* $\mathrm{co}(S)$*, there exist at least two distinct subsets $S', S'' \subset S$ with no more than $d+1$ elements each, such that $p \in \mathrm{co}(S') \cap \mathrm{co}(S'')$.*

Proof. For any vector $x \in \mathbb{R}^d$, denote the vector's i^{th} coordinate by $x_{[i]}$, and set $\bar{x} \equiv \binom{1}{x} \in \mathbb{R}^{d+1}$. Define the matrices $X = [x_1 \ x_2 \ \cdots \ x_{d+2}] \in \mathbb{R}^{d\times(d+2)}$ and $\overline{X} = [\bar{x}_1 \ \bar{x}_2 \ \cdots \ \bar{x}_{d+2}] \in \mathbb{R}^{(d+1)\times(d+2)}$. Since $p \in \mathrm{co}(S)$, there exists a vector $\lambda = (\lambda_{[1]},\ldots,\lambda_{[d+2]})^T \in \mathbb{R}^{d+2}$ such that $\sum_{i=1}^{d+2}\lambda_{[i]} = 1$ and $X\lambda = p$.

The vectors $\bar{x}_1,\bar{x}_2,\ldots,\bar{x}_{d+2}$ are linearly dependent. Hence, there is a vector $\alpha = (\alpha_{[1]},\ldots,\alpha_{[d+2]})^T \in \mathbb{R}^{d+2}$, with coordinates not all equal to zero, such that $\alpha \in \ker\left(\overline{X}\right)$. Since $\sum_{i=1}^{d+2}\alpha_{[i]} = 0$ then α has at least one positive coordinate and at least one negative coordinate.

Suppose without loss of generality that the coordinates in α are ordered such that $\frac{\lambda_{[1]}}{\alpha_{[1]}} \leq$

$\cdots \leq \frac{\lambda_{[k]}}{\alpha_{[k]}} < 0 < \frac{\lambda_{[k+1]}}{\alpha_{[k+1]}} \leq \cdots \leq \frac{\lambda_{[d+2]}}{\alpha_{[d+2]}}$. We can therefore decompose the vector p as follows:

$$p = \sum_{i=1}^{d+2} \lambda_{[i]} \bar{x}_i = \sum_{i=1}^{k} \lambda_{[i]} \bar{x}_i + \frac{\lambda_{[k+1]}}{\alpha_{[k+1]}} \sum_{i=k+1}^{d+2} \alpha_{[i]} \bar{x}_i + \sum_{i=k+2}^{d+2} \left(\frac{\lambda_{[i]}}{\alpha_{[i]}} - \frac{\lambda_{[k+1]}}{\alpha_{[k+1]}} \right) \alpha_{[i]} \bar{x}_i.$$

Substituting $\sum_{i=k+1}^{d+2} \alpha_{[i]} \bar{x}_i = -\sum_{i=1}^{k} \alpha_{[i]} \bar{x}_i$ and rearranging yields

$$p = \sum_{i=1}^{k} \left(\frac{\lambda_{[i]}}{\alpha_{[i]}} - \frac{\lambda_{[k+1]}}{\alpha_{[k+1]}} \right) \alpha_{[i]} \bar{x}_i + \sum_{i=k+2}^{d+2} \left(\frac{\lambda_{[i]}}{\alpha_{[i]}} - \frac{\lambda_{[k+1]}}{\alpha_{[k+1]}} \right) \alpha_{[i]} \bar{x}_i.$$

Therefore, the vector $\beta = (\beta_{[1]}, \ldots, \beta_{[d+2]})^T$ that is defined such that $\beta_{[i]} = \left(\frac{\lambda_{[i]}}{\alpha_{[i]}} - \frac{\lambda_{[k+1]}}{\alpha_{[k+1]}} \right) \alpha_{[i]}$ satisfies $\sum_{i=1}^{d+2} \beta_{[i]} = 1$ and $X\beta = p$ and all its coordinates are non-negative. A similar argument shows that the vector $\gamma = (\gamma_{[1]}, \ldots, \gamma_{[d+2]})^T$ that is defined such that $\gamma_{[i]} = \left(\frac{\lambda_{[i]}}{\alpha_{[i]}} - \frac{\lambda_{[k]}}{\alpha_{[k]}} \right) \alpha_{[i]}$ satisfies $\sum_{i=1}^{d+2} \gamma_{[i]} = 1$ and $X\gamma = p$ and all its coordinates are non-negative.

Let $S' = \{x_1, \ldots, x_k, x_{k+2}, \ldots, x_{d+2}\}$ and $S'' = \{x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_{d+2}\}$. We have therefore showed that $p \in \text{co}(S')$ and $p \in \text{co}(S'')$. Moreover, notice that $\lambda_{[i]} = \nu \beta_{[i]} + (1-\nu)\gamma_{[i]}$ where $\nu = \frac{1}{1 - \frac{\lambda_{k+1}}{\alpha_{k+1}} \frac{\alpha_k}{\lambda_k}}$. ■

By Lemma A.1, given any set of beliefs $P = \{p_1, \ldots, p_{|\Omega|+1}\}$ that are each different from π, which are induced with positive probabilities $\lambda_1, \ldots, \lambda_{|\Omega|+1}$, respectively, such that $\sum_{i=1}^{|\Omega|+1} \lambda_i p_i = \pi$, there exist at least two subsets of beliefs $P', P'' \subset P$ with no more than $|\Omega|$ elements each, with associated probabilities λ' and λ'', which also average the prior belief π. With slight abuse of notation we also use λ' and λ'' to denote the $|\Omega|+1$ dimensional vectors of probabilities $p_1, \ldots, p_{|\Omega|+1}$ where instead of the probability associated with the belief that is missing from the subset P' and P'', respectively, we write zero.

Inspection of the proof of Lemma A.1 reveals that the vector λ can be written as a convex combination of the vectors λ' and λ''. Therefore, the expected payoff $U = \sum_{i=1}^{|\Omega|+1} \lambda_i \cdot \hat{u}_S(p_i)$ can be written as a convex combination of the expected payoffs $U' = \sum_{i=1}^{|\Omega|+1} \lambda'_i \cdot \hat{u}_S(p_i)$ and $U'' = \sum_{i=1}^{|\Omega|+1} \lambda''_i \cdot \hat{u}_S(p_i)$ associated with the two vectors of probabilities λ' and λ''. It follows that either U' or U'' is larger than or equal to U.

Finally, the message functions σ' and σ'' that induce the posterior beliefs in P' and P'', respectively, satisfy credibility because of the same argument used in the proof of part (i) of the proposition. Namely:

$$q^{\sigma}(m_j, \omega) = 0 \Rightarrow q^{\sigma'}(m'_j, \omega) = 0, p^{\sigma''}(m'_j, \omega) = 0 \quad \forall j \in \{1, \cdots, |\Omega|+1\}.$$

Because, otherwise, $q^{\sigma'}(m'_j, \omega), q^{\sigma''}(m'_j, \omega) = 0 > 0 = q^{\sigma}(m_j, \omega)$. A contradiction. Thus, it is never the case that a message is sent under σ' in a state where it was not sent under σ. Therefore, the credibility of σ implies the credibility of σ' and σ''.

Proof of Proposition 3

Fix a belief p^* and a cost parameter $\alpha^* \geq 0$. We show that for any (p,α) close to (p^*,α^*) (according to the Euclidean metric), $V(p,\alpha)$ is close to $V(p^*,\alpha^*)$.

Denote the set of posterior beliefs induced by the optimal message function (under the belief p^* and the cost parameter α^*) by P and denote the induced distribution over P by τ.[25] For any two posterior beliefs $p,p' \in P$, denote the weighted mean of p and p' by

$$\mu_{p,p'} \equiv \frac{\tau(p)}{\tau(p)+\tau(p')} \cdot \mu_p + \frac{\tau(p')}{\tau(p)+\tau(p')} \cdot \mu_{p'}.$$

Define

$$g(\mu_p,\mu_{p'}) \equiv \frac{\hat{u}_S(\mu_{p'}) - \hat{u}_S(\mu_p)}{\mu_{p'} - \mu_p}.$$

The value of $g(\mu_p,\mu_{p'})$ can be interpreted as the slope of the line that connects the point $(\mu_p, \hat{u}_S(\mu_p))$ with the point $(\mu_{p'}, \hat{u}_S(\mu_{p'}))$ on the mean/payoff plane. In Figure (6a) this is the slope of the dashed line. Notice that, given three posterior beliefs p, p' and p'' that are such that $\mu_p < \mu_{p'} < \mu_{p''}$, if $g(\mu_p,\mu_{p'}) = g(\mu_{p'},\mu_{p''}) = \alpha^*$ then $g(\mu_p,\mu_{p''}) = \alpha^*$. In this case, we the three points $(\mu_p, \hat{u}_S(\mu_p))$, $(\mu_{p'}, \hat{u}_S(\mu_{p'}))$ and $(\mu_{p''}, \hat{u}_S(\mu_{p''}))$ are all on the same line in the mean/payoff plane.

Credibility of the optimal message function implies that $|g(\mu_p,\mu_{p'})| \leq \alpha^*$ for any pair of posterior beliefs $p,p' \in P$. If the inequality is strict for all such pairs (i.e. the credibility constraint is not binding), then clearly the same value of V can be achieved by employing the same distribution of posteriors τ over the set of posterior beliefs P for any α that is sufficiently close to α^*.

We therefore assume that there is at least one pair of posterior beliefs $p,p' \in P$ for which $g(\mu_p,\mu_{p'}) = \alpha^*$ (the case of $-g(\mu_p,\mu_{p'}) = \alpha^*$ is analogous and is omitted). The next two lemmas are useful for the analysis that follows:

Lemma A.2 *Let $p,p' \in P$ be such that $\mu_p < \mu_{p'}$. For any two posterior means μ_x, μ_y such that $\mu_p \leq \mu_x < \mu_{p,p'} < \mu_y \leq \mu_{p'}$ there exists a set of posterior beliefs $\hat{P} = P \setminus \{p,p'\} \cup \{x,y\}$, where x and y are posterior beliefs that induce the means μ_x and μ_y, respectively, and a Bayes plausible distribution $\hat{\tau}$ over $\hat{P}$ that is such that*

$$\hat{\tau}(x) = (\tau(p)+\tau(p')) \cdot \frac{\mu_y - \mu_{p,p'}}{\mu_y - \mu_x}$$

$$\hat{\tau}(y) = (\tau(p)+\tau(p')) \cdot \frac{\mu_{p,p'} - \mu_x}{\mu_y - \mu_x}$$

[25] If such posteriors do not exist, pick posterior beliefs that induce a value of V that is close to $V(p^*,\alpha^*)$.

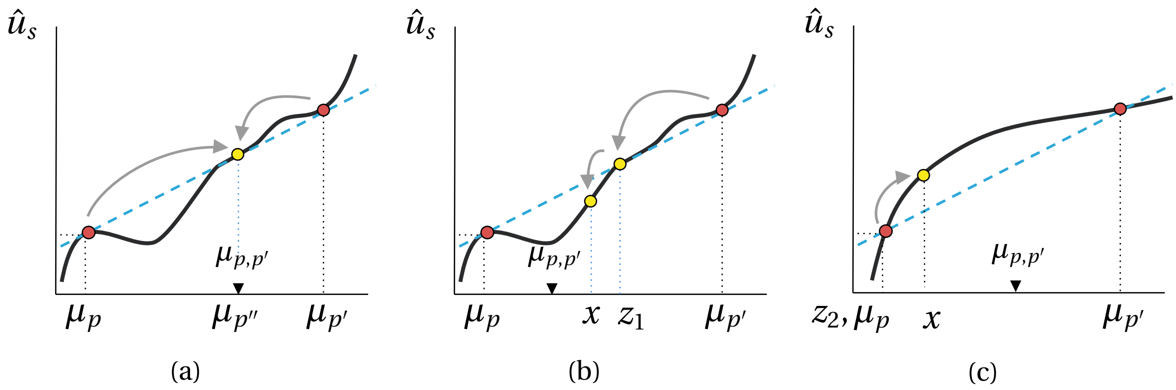


Figure 6

and $\hat{\tau} = \tau$ otherwise. We refer to the substitution of p, p' by x, y as the replacement of p and p' by x and y. *Furthermore, if $g(\mu_p, \mu_x) = g(\mu_x, \mu_y) = g(\mu_y, \mu_{p'})$, then the value of V induced by τ is the same as the value of V induced by $\hat{\tau}$.*

Lemma A.3 *Suppose that $p, p', p'' \in P$ are three posterior beliefs with means $\mu_p < \mu_{p'} < \mu_{p''}$ such that $g(\mu_p, \mu_{p'}) = g(\mu_{p'}, \mu_{p''}) = \alpha^*$. Then, it is possible to eliminate either p, or p'', or both, from P, and adjust the distribution over posteriors τ, in a way that preserves the value of V and preserves credibility.*

Fix a pair of posterior beliefs $p, p' \in P$ for which $g(\mu_p, \mu_{p'}) = \alpha^*$. By Lemma A.3, no loss of generality in implied by assuming that $g(\mu_p, \mu_y) < \alpha^*$ for all $y \in P \setminus \{p, p'\}$ (as otherwise at least one posterior belief can be eliminated from P). Credibility then implies that $\mu_y \notin (\mu_p, \mu_{p'})$ for all $y \in P \setminus \{p, p'\}$. [26]

We distinguish between three cases:

(i) Suppose that $g(\mu_p, \mu_{p,p'}) = \alpha^*$. In this case, the point $(\mu_{p,p'}, \hat{u}_S(\mu_{p,p'}))$ lies on the line that connects the points $(\mu_p, \hat{u}_S(\mu_p))$ and $(\mu_{p'}, \hat{u}_S(\mu_{p'}))$ in the mean/payoff plane, as illustrated in Figure (6a).

Modify the message function such that in any state in which the messages that induce p and p' were sent, the message function now sends only one message. Denote the posterior belief induced by this new message by p'' and notice that the mean of p'' is $\mu_{p''} = \mu_{p,p'}$. The fact that $g(\mu_p, \mu_{p,p'}) = g(\mu_{p,p'}, \mu_{p'}) = \alpha^*$ implies that the value of V is unaffected by the modification (see also the proof of Lemma A.3).

[26]To see this, suppose by way of contradiction that $\mu_y \in (\mu_p, \mu_{p'})$ and that $\hat{u}_S(\mu_{p'}) > \hat{u}_S\left(\mu_p\right)$ (the other case is handled similarly). Since $g\left(\mu_p, \mu_y\right) < \alpha^*$ then $\hat{u}_S(\mu_{p'}) - \hat{u}_S(\mu_y) > \hat{u}_S(\mu_{p'}) - \hat{u}_S\left(\mu_p\right) - (\mu_y - \mu_p) \cdot \alpha^*$. Using the fact that $g(\mu_p, \mu_{p'}) = \alpha^*$ we obtain $\hat{u}_S(\mu_{p'}) - \hat{u}_S(\mu_y) > (\mu_{p'} - \mu_y)\alpha^*$, and since $\mu_y \in (\mu_p, \mu_{p'})$ then $g\left(\mu_y, \mu_{p'}\right) > \alpha^*$, contradicting credibility of the message function.

After the modification, we have that $|g(\mu_{p''}, \mu_y)| < \alpha^*$ for all $y \in P \setminus \{p, p'\}$. Intuitively, this is because for any $\mu_y \notin (\mu_p, \mu_{p'})$, the slope of the line that connects the point $(\mu_y, \hat{u}_S(\mu_y))$ with the point $(\mu_{p''}, \hat{u}_S(\mu_{p''}))$ in the mean/payoff plane is between the slopes of: (A) the line that connects $(\mu_y, \hat{u}_S(\mu_y))$ with $(\mu_p, \hat{u}_S(\mu_p))$ and (B) the line that connects $(\mu_y, \hat{u}_S(\mu_y))$ with $(\mu_{p'}, \hat{u}_S(\mu_{p'}))$. Since, by credibility, both (A) and (B) are smaller than α^* in absolute value, the result follows.

Formally, fix any posterior belief $y \in P \setminus \{p, p'\}$. Since $g(\mu_p, \mu_{p''}) = \alpha^*$, we have that

$$g\left(\mu_y, \mu_{p''}\right) = \frac{u\left(\mu_{p''}\right) - u\left(\mu_y\right)}{\mu_{p''} - \mu_y} = \frac{u\left(\mu_p\right) + \alpha^*\left(\mu_{p''} - \mu_p\right) - u\left(\mu_y\right)}{\mu_{p''} - \mu_y}.$$

Differentiating g with respect to $\mu_{p''}$ yields:

$$\frac{\partial g\left(\mu_y, \mu_{p''}\right)}{\partial \mu_{p''}} = \left(\mu_p - \mu_y\right)\frac{\alpha^* - g\left(\mu_y, \mu_p\right)}{\left(\mu_{p''} - \mu_y\right)^2}. \tag{7}$$

Recall that $g\left(\mu_y, \mu_p\right) < \alpha^*$ for all $y \in P \setminus \{p, p'\}$. Thus, if $\mu_y < \mu_p$, the right-hand side of Equation (7) is positive and so $g\left(\mu_y, \mu_p\right) < g\left(\mu_y, \mu_{p''}\right) < g(\mu_y, \mu_{p'})$. And, if $\mu_y > \mu_{p'}$, the right-hand side of Equation (7) is negative and so $g\left(\mu_y, \mu_p\right) > g\left(\mu_y, \mu_{p''}\right) > g(\mu_y, \mu_{p'})$. It follows that $|g\left(\mu_y, \mu_{p''}\right)| < \max[|g\left(\mu_y, \mu_p\right)|, |g(\mu_y, \mu_{p'})|] \leq \alpha^*$ for all $\mu_y \not\in [\mu_p, \mu_{p'}]$.

Thus, the modified message function eliminates a pair of posterior beliefs for which the credibility constraint was binding, and replaced it with one posterior belief for which credibility is not binding with any other element in P.

(ii) Suppose that $g(\mu_p, \mu_{p,p'}) < \alpha^*$. In this case, the point $(\mu_{p,p'}, \hat{u}_S(\mu_{p,p'}))$ is *below* the line that connects the points $(\mu_p, \hat{u}(\mu_p))$ and $(\mu_{p'}, \hat{u}(\mu_{p'}))$ in the mean/payoff plane, as illustrated in Figure 6b).

Let $z_1 \in [\mu_p, \mu_{p'}]$ be the lowest mean that is greater than $\mu_{p,p'}$ for which $g(\mu_p, z_1) = \alpha^*$, i.e. $z_1 = \min[x | x > \mu_{p,p'}$ and $g(\mu_p, x) = \alpha^*]$. Note that z_1 necessarily exists, by the continuity of g and the intermediate value theorem (it could be the case that $z_1 = \mu_{p'}$).

Replace the posteriors p and p' by the posteriors p and p'', where p'' is a posterior with mean z_1, in the manner described in Lemma A.2 and illustrated in Figure (6b). This modification does not change the value of the function V because $g(\mu_p, \mu_{p'}) = g(\mu_p, z_1)$. Credibility of the original message function, and the fact that $z_1 \in (\mu_p, \mu_{p'})$, imply that $|g(\mu_y, \mu_{p''})| < \alpha^*$ for all $y \in P \setminus \{p\}$ (the analysis is identical to the one presented in case (i) above). Therefore the modified message function satisfies credibility.

Continuity of g implies that, for any $\varepsilon > 0$, there exists a $\hat{\delta} > 0$, such that if $0 < \delta < \hat{\delta}$ then there exists $x \in [z_1 - \varepsilon, z_1]$ such that $|g(\mu_y, x)| \leq \alpha^* - \delta$ for all $y \in P \setminus \{p'\}$. Thus, when α is close to α^*, we can modify the message function (by replacing the posterior beliefs p and p'' by p and p''', where p''' is a posterior belief with mean x, in the manner described in Lemma A.2 and illustrated in Figure 6b), such that credibility is satisfied and the value of V is only slightly affected.[27]

(iii) Suppose that $g(\mu_p, \mu_{p,p'}) > \alpha^*$. In this case, the point $(\mu_{p,p'}, \hat{u}_S(\mu_{p,p'}))$ is *above* the line that connects the points $(\mu_p, \hat{u}(\mu_p))$ and $(\mu_{p'}, \hat{u}(\mu_{p'}))$ in the mean/payoff plane, as illustrated in Figure 6c).

Let $z_2 \in [\mu_p, \mu_{p'}]$ be the highest value that is smaller than $\mu_{p,p'}$ for which $g(\mu_p, z_2) = \alpha^*$, i.e. $z_2 = \max[x | x < \mu_{p,p'} \text{ and } g(\mu_p, x) = \alpha^*]$. For brevity of notation we define $g(\mu_p, \mu_p) = \alpha^*$ and allow z_2 to be equal to μ_p, which is the case that is illustrated in Figure (6c). As in case (ii), z_2 necessarily exists by the continuity of g.

If $z_2 \neq \mu_p$, replace the posteriors p and p' by the posteriors p'' and p', where p'' is a posterior with mean z_2, in the manner described in Lemma A.2. As in case (ii) above, this modification preserves the value of V and the credibility of the message function.

Continuity of g implies that, for any $\varepsilon > 0$, there exists a $\hat{\delta} > 0$, such that if $0 < \delta < \hat{\delta}$ then there exists $x \in [z_2, z_2 + \varepsilon]$ such that $|g(\mu_y, x)| \leq \alpha^* - \delta$ for all $y \in P \setminus \{p'\}$. As in case (ii) above, when α is close to α^*, we can modify the message function (by replacing the posterior beliefs p' and p'' by p' and p''', where p''' is a posterior belief with mean x, in the manner described in Lemma A.2 and illustrated in Figure 6c), such that credibility is satisfied and the value of V is only slightly affected.

Thus, for any pair of posterior beliefs $p, p' \in P$ for which credibility is binding in the original message function, and for any small change in α^*, it is either the case that this pair can be eliminated without affecting the value of V (case i), or there exists a modification of the message function that restores credibility while only slightly affecting the value of V (cases ii and iii). Therefore, if (p^*, α) is close to (p^*, α^*) then the value $V(p^*, \alpha)$ is close to $V(p^*, \alpha^*)$.

To complete proof, suppose that the belief p is close to the belief p^*. Suppose also that the optimal message function under p^* induces a (credible) distribution τ over a set of posterior beliefs P. Then, there is a distribution $\hat{\tau}$ on the *same set* of posterior beliefs P, that assigns only slightly different weights to the elements of P compared to τ, that is Bayes plausible and credible. Therefore, when (p, α) is close to (p^*, α^*), the value $V(p, \alpha)$ is close to $V(p^*, \alpha^*)$.

[27] This is because $\hat{u}_S$ is a continuous function and because the the distribution over posterior beliefs, $\hat{\tau}$, that is described in the statement of Lemma A.2, is only slightly affected by the modification.

Proof of Lemma A.2

Observe that this replacement of posteriors is performed in a way that contracts the distribution of posterior means and preserves both the conditional mean $\mu_{p,p'}$ and the mean μ_{p^*}. This implies that $\hat{\tau}$ second-order-stochastically-dominates (SOSD) τ. This ensures that the distribution $\hat{\tau}$ is Bayes plausible.

Suppose now that $g\left(\mu_p, \mu_x\right) = g\left(\mu_x, \mu_y\right) = g\left(\mu_y, \mu_{p'}\right)$. Sender's value (V) from employing the modified message function is given by:

$$\sum_{q \in P \setminus \{p,p'\} \cup \{x,y\}} \hat{\tau}\left(q\right) \cdot \hat{u}_S\left(q\right) = \sum_{q \in P \setminus \{p,p'\}} \hat{\tau}\left(q\right) \hat{u}_S\left(q\right) + \hat{\tau}\left(x\right) \cdot \hat{u}_S\left(\mu_x\right) + \hat{\tau}\left(y\right) \cdot \hat{u}_S(\mu_y). \tag{8}$$

By construction, we have that $\hat{\tau}\left(x\right) = \frac{\mu_y - \mu_p}{\mu_y - \mu_x} \cdot \tau\left(p\right) - \frac{\mu_{p'} - \mu_y}{\mu_y - \mu_x} \cdot \tau\left(p'\right)$. Since $g\left(\mu_p, \mu_x\right) = g\left(\mu_x, \mu_y\right) = g\left(\mu_y, \mu_{p'}\right)$, then

$$\hat{\tau}\left(x\right) = \frac{\hat{u}_S(\mu_y) - \hat{u}_S(\mu_p)}{\hat{u}_S(\mu_y) - \hat{u}_S(\mu_x)} \cdot \tau\left(p\right) - \frac{\hat{u}_S(\mu_{p'}) - \hat{u}_S(\mu_y)}{\hat{u}_S(\mu_y) - \hat{u}_S(\mu_x)} \cdot \tau\left(p'\right).$$

By plugging this expression of $\hat{\tau}\left(x\right)$, and $\hat{\tau}\left(y\right) = \tau\left(p\right) + \tau\left(p'\right) - \hat{\tau}\left(x\right)$, into the right-hand side of Equation (8) we obtain:

$$\sum_{q \in P \setminus \{p,p'\}} \hat{\tau}\left(q\right) \hat{u}_S\left(q\right) + \tau\left(p\right) \cdot \hat{u}_S\left(\mu_p\right) + \tau\left(p'\right) \cdot \hat{u}_S(\mu_{p'}) = \sum_{q \in P} \hat{\tau}\left(q\right) \hat{u}_S\left(q\right),$$

which is Sender's value under the original message function.

Proof of Lemma A.3

Suppose that a credible message function induces the three posterior beliefs p, p', p'' as described in the statement of the lemma.

If $\mu_{p,p''} = \mu_{p'}$, modify the message function so that in any state in which the messages that induced p and p'' were sent, the modified message function would send the message that induced p' instead. Thus, the mean of the posterior belief induced by this message remains $\mu_{p,p''}$. The fact that $g(\mu_p, \mu_{p'}) = g(\mu_{p'}, \mu_{p''})$ implies that the value of V remains unchanged.[28]

If $\mu_{p,p''} < \mu_{p'}$, replace the posteriors p and p'' in P by p and p', in the manner described in Lemma A.2. If $\mu_{p,p''} > \mu_{p'}$, replace the posteriors p and p'' in P by p' and p'' in the manner described in Lemma A.2. These modifications do not change the value of the function V.

Finally, note that in all the cases described above, the modified message function does not induce a posterior belief that was not induced by the original message function. Thus, the

[28]To see this, note first that $\tau\left(p\right) \cdot \hat{u}_S(\mu_p) + \tau\left(p'\right) \cdot \hat{u}_S(\mu_{p'}) = \left(\tau\left(p\right) + \tau\left(p'\right)\right) \left(\frac{\tau(p)}{\tau(p)+\tau(p')} \hat{u}_S(\mu_p) + \frac{\tau(p')}{\tau(p)+\tau(p')} \hat{u}_S(\mu_{p'})\right)$. Next, since $\mu_{p'} = \mu_{p,p''}$ and $g(\mu_p, \mu_{p'}) = \alpha$ we have that $\hat{u}_S(\mu_p) = \hat{u}_S(\mu_{p,p''}) - (\mu_{p,p''} - \mu_p)\alpha$, and since $g(\mu_{p'}, \mu_{p''}) = \alpha$ we have $\hat{u}_S(\mu_{p''}) = \hat{u}_S(\mu_{p,p''}) + \left(\mu_{p''} - \mu_{p,p''}\right)\alpha$. By definition of $\mu_{p,p''}$ we then obtain $\tau\left(p\right) \cdot \hat{u}_S(\mu_p) + \tau\left(p'\right) \cdot \hat{u}_S(\mu_{p'}) = (\tau\left(p\right) + \tau\left(p'\right)) \cdot \hat{u}_S(\mu_{p,p''})$.

credibility constraints in Sender's problem (SP1) are only relaxed, and the fact that the original message function was credible implies that the modified one is also credible.

Proof of Proposition 5

We prove the proposition for the case in which $\hat{u}_S$ is increasing and convex. The proof for the case in which $\hat{u}_S$ is decreasing, or decreasing and then increasing is analogous.

Suppose that the state space is binary, i.e., $\Omega = \{l, h\}$ for some two numbers $l, h \in \mathbb{R}$ with $l < h$. A belief over Ω can be described by the probability $p \in [0,1]$ that the state is h. The prior belief is thus given by $\pi \in (0,1)$. The mean of belief p is $\mu_p = l + (h - l)\,p$. In what follows we normalize the parameters h and l to be 1 and 0, respectively, and therefore $\mu_p = p$.

According to Corollary 1 the optimal message function induces either one posterior belief that is equal to the prior π, or two credible posterior beliefs $p_L < \pi < p_H$, whichever generates a higher expected payoff to Sender. In the former case, the ex-ante expected payoff to Sender is $\hat{u}_S(\pi)$. In the latter case, the ex-ante expected payoff to Sender is $p_H \cdot \hat{u}_S(p_H) + p_L \cdot \hat{u}_S(p_L)$. Credibility requires that $\frac{\hat{u}_S(p_H) - \hat{u}_S(p_L)}{p_H - p_L} \leq \alpha$.

We distinguish between the following three cases:

(i) If $\hat{u}_S(1) - \hat{u}_S(0) \leq \alpha$, then the a message function that induces a distribution τ over the posterior beliefs $p_L^* = 0$ (realized with probability $1 - \pi$) and $p_H^* = 1$ (realized with probability probability π) is credible under α. Such a message function is optimal for Sender because it concavifies $\hat{u}_S$ on the interval $[0,1]$.

(ii) If $\frac{\hat{u}_S(\pi) - \hat{u}_S(0)}{\pi} < \alpha < \hat{u}_S(1) - \hat{u}_S(0)$, then the two optimally induced beliefs under α are $p_L^* = 0$ and p_H^* that is such that $\frac{\hat{u}_S(p_H^*) - \hat{u}_S(0)}{p_H^*} = \alpha$. To see this, note first that for any different pair of posterior beliefs $p_L < \pi < p_H$, decreasing p_L relaxes the credibility constraint and improves the expected payoff to Sender. Then, it is possible to increase p_H up to p_H^*, where the credibility constraint is binding, i.e., $\frac{\hat{u}_S(p_H^*) - \hat{u}_S(0)}{p_H^*} = \alpha$, which further increases the ex-ante expected payoff to Sender.

(iii) If $\alpha \leq \frac{\hat{u}_S(\pi) - \hat{u}_S(0)}{\pi}$, then the unique feasible policy induces just one posterior belief, which is equal to the prior π. This is because the convexity of $\hat{u}_S$ implies that $\frac{\hat{u}_S(p_H) - \hat{u}_S(p_L)}{p_H - p_L}$ is increasing in p_L and in p_H and therefore $\frac{\hat{u}_S(p_H) - \hat{u}_S(p_L)}{p_H - p_L} \geq \alpha$ for any $p_L \leq \pi$ and $p_H \geq \pi$. Thus, no message function can induce two (Bayes plausible) posterior beliefs in a credible way.

Notice that decreasing the value of α does not affect Sender's optimal distribution over posteriors so long as α remains in case (i) or (iii). As the value of α changes from case (i) to (ii), or as α decreases within case (ii), Sender's optimal distribution τ becomes more garbled. This is because the convexity of $\hat{u}$ implies that $\frac{\hat{u}_S(p_H)-\hat{u}_S(0)}{p_H}$ increases in p_H. Thus, a lower value of α implies a lower value of p_H^* (i.e. messages are less informative with respect to the state).

Proof of Lemma 1

A partition of Ω_N into singletons is equivalent to a message function that fully reveals the state, i.e., $\sigma(\omega) = \omega$. Notice that, in this case, $\mu_m = \overline{m} = \underline{m} = \omega$ for every message $m = \omega$. Inspection of the constraints ICup(k) and ICdown(k) reveals that they are all satisfied, which implies that $\sigma(\omega) = \omega$ is credible for every N. Optimality follows from the fact that $\text{Var}[\omega|\omega \in m] = 0$ for all m and, therefore, Sender's expected payoff, given in (5), attains its highest possible value.

Proof of Lemma 2

The credibility constraint ICup(k) and the fact that $\mu_k \leq \overline{m}_k$ implies that $\mu_{k+1} \geq \overline{m}_k + 2b - \alpha$. Since $\mu_{k+1} \leq \overline{m}_{k+1}$, it follows that $\overline{m}_{k+1} \geq \overline{m}_k + 2b - \alpha$. Therefore, it is impossible to fit more than $\frac{1}{2b-\alpha}$ messages into a credible partition.

Proof of Lemma 3

Suppose that m_k and m_{k+1} are two adjacent convex messages. Convexity implies that $\mu_k = \frac{\underline{m}_k+\overline{m}_k}{2}$ and $\mu_{k+1} = \frac{\underline{m}_{k+1}+\overline{m}_{k+1}}{2}$. The incentive constraint ICup(k) is then given by:

$$\frac{\underline{m}_k + \overline{m}_k}{2} + \frac{\underline{m}_{k+1} + \overline{m}_{k+1}}{2} - 2\overline{m}_k \geq 2b - \alpha.$$

Convexity implies also that $\overline{m}_k - \underline{m}_k = \frac{|m_k|-1}{N}$ and $\overline{m}_{k+1} - \underline{m}_{k+1} = \frac{|m_{k+1}|-1}{N}$. Since the messages are adjacent then $\underline{m}_{k+1} - \overline{m}_k = \frac{1}{N}$. We can therefore equivalently write ICup(k) as follows:

$$\frac{|m_{k+1}|}{N} - \frac{|m_k|}{N} + \frac{2}{N} \geq 4b - 2\alpha. \tag{9}$$

Equation (9) is a necessary and sufficient condition for $ICup(k)$ when messages are convex. If, in addition, $ICup(k)$ is binding (i.e. it would have been violated had a different message been sent in the state $\overline{m}_{k+1}$), then:

$$\frac{|m_{k+1}|}{N} - \frac{|m_k|}{N} + \frac{1}{N} < 4b - 2\alpha. \tag{10}$$

Denote $x = \frac{|m_1|}{N} = x > 0$. Then, the fact that the messages are tightly packed implies that $x + 4b - 2\alpha - \frac{2}{N} \leq \frac{|m_2|}{N} \leq x + 4b - 2\alpha - \frac{1}{N}$, $x + 8b - 4\alpha - \frac{4}{N} \leq \frac{|m_3|}{N} \leq x + 8b - 4\alpha - \frac{2}{N}, \ldots, x + (k-1)(4b - 2\alpha - \frac{2}{N}) \leq \frac{|m_k|}{N} \leq x + (k-1)(4b - 2\alpha - \frac{1}{N})$, and so on.

Thus, $\frac{|m_1|}{N}+\cdots+\frac{|m_k|}{N}$ is bounded between two sums of arithmetic series with k elements:

$$2k(k-1)\left(b-\frac{\alpha}{2}-\frac{1}{N}\right)+kx\leq\frac{|m_1|}{N}+\cdots+\frac{|m_k|}{N}\leq 2k(k-1)\left(b-\frac{\alpha}{2}-\frac{1}{2N}\right)+kx. \qquad (11)$$

Since x can be set arbitrarily small, then for sufficiently large N, the maximal number of messages that can be tightly packed into the set $\left\{0,\frac{1}{N},\ldots,l\right\}$ is given by

$$I(l)=\left\lceil\sqrt{\frac{1}{4}+\frac{l}{2b-c}}-\frac{1}{2}\right\rceil.$$

Finally, uniqueness of the partition on the set $\left\{0,\frac{1}{N},\ldots,l\right\}$ follows from the fact that the size of x determines the entire partition up to l. If two partitions have different values of x then the one with the shorter x falls short of covering the set $\left\{0,\frac{1}{N},\ldots,l\right\}$.

Proof of Proposition 6

Consider a credible partition that does not consist of $I(1)$ tightly packed messages on Ω_N. The algorithm described in the text "packs message m_k" and produces a new partition in which $I(l_k)$ messages are tightly packed on the set $\{0,\ldots,l_k\}$. We show that in *each iteration* of the algorithm the value of the objective function improves and all ICup constraints are preserved.

Our proof proceeds in two steps. In step 1, we show that performing Part I of the algorithm improves the value of the objective function. Furthermore, after Part I is performed all the ICup constraints, except for maybe one, are satisfied. If this one constraint is indeed violated, we show a modification of the partition after which: (i) *all* the ICup constraints are satisfied, and (ii) the objective function's value is higher than that of the original partition (before the execution of Part I of the algorithm).

In step 2, we show that partition produced in Step 1 is in fact sub-optimal relative to a partition in which messages are tightly "re-packed" in a maximal manner, and in which all ICup constraints are satisfied. Steps 1 and 2 can be repeated until the resulting partition is one that consists of $l(1)$ tightly packed messages on Ω_N.

To conclude the proof, we show that this final partition satisfies all the ICdown constraints, and it is therefore credible.

Step 1 (Fix all ICup constraints and improve the objective function's value)

Part I of the algorithm "convexifies" message m_k to the left. The outcome of this process is illustrated in Figure (7a). We refer to the partition before the convexification as the "original partition" and to the partition after the convexification as the "convexified partition". Since in the convexified partition the variance of each message m_j is weakly smaller than the variance

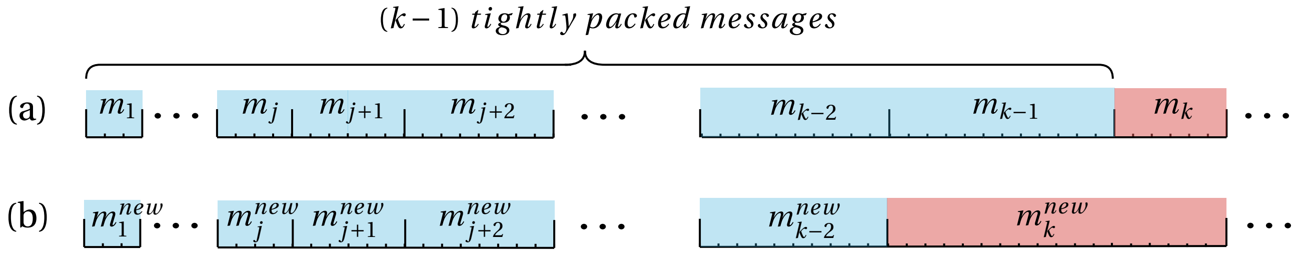


Figure 7: Step1, Case I

of m_j under the original partition, then the convexified partition attains a higher value for the objective function (5).

Notice that the convexification (as described in the algorithm in the text) involves sequential "swaps" of states between message m_k and messages m_j with $j > k$. Specifically, in each such swap a state ω is removed from message m_k and added to message m_j, whereas state $\omega - \frac{1}{N}$ is removed from m_j and added to m_k. Denote by ϕ_j the number of states swapped between messages m_j and m_k in the process of convexifying m_k. Define $\Phi_k = \sum_{j>k} \phi_j$ to be the total number of swaps. It follows that: (i) the mean μ_j in the convexified partition is larger than the mean μ_j in the original partition by $\frac{\phi_j}{N|m_j|}$ for all $j > k$, and (ii) the mean of μ_k in the convexified partition is smaller than the mean of μ_k in the original partition by $\frac{\zeta_k}{N}$, where $\zeta_k \equiv \frac{\Phi_k}{|m_k|}$.

In the convexified partition, the constraints $ICup(1), \dots, ICup(k-2)$ are satisfied because the convexification does not affect them. The constraints $ICup(k+1), \dots, ICup(J-1)$ are also satisfied. To see this, note that credibility of the original partition implies that $\mu_j \le \overline{m}_j \le \mu_{j+1} \le \overline{m}_{j+1}$ for all $j \le J-1$ and therefore the states $\overline{m}_{k+1}, \dots, \overline{m}_J$ are all higher than the state $\overline{m}_k$ (namely, the highest state in message m_k). Hence, the maximal state that belongs to each message m_j with $j > k$ is unchanged between the original and the convexified partition, i.e., the values of $\overline{m}_{k+1}, \dots, \overline{m}_J$ are unaffected by the convexification. Moreover, the the values of $\mu_{k+1}, \dots, \mu_J$ are all weakly larger in the convexified partition relative to the original one. Therefore, the fact that $ICup(k+1), \dots, ICup(J-1)$ are satisfied in the original partition implies that they are satisfied in the convexified partition as well.

In the convexified partition, the constraint $ICup(k)$ is satisfied with a slack. To see this, notice first that the convexification weakly increases μ_{k+1} relative to its value in the original partition. Next, note that although the convexification decreases μ_k by $\frac{\zeta_k}{N}$ relative to the orignal partition, it also decreases $\overline{m}_k$ by $\frac{L}{N}$, where L is the number of states associated with messages $m_{k+1}, \dots, m_J$ that are smaller than $\overline{m}_k$. Finally, observe that the number of swaps needed to convexify m_k (i.e., Φ_k) is smaller than L multiplied by the total number of states in m_k, that is

$$\Phi_k \le L \cdot |m_k|. \tag{12}$$

It follows that the sum $\mu_k + \mu_{k+1}$ decreases by no more than $\frac{\zeta_k}{N}$ while $\overline{m}_k$ decreases by at least $\frac{\zeta_k}{N}$. Thus, the fact that $ICup(k)$ is satisfied in the original partition implies that it is satisfied also in the convexified partition. In fact, observe that the convexification creates a slack of at least $\frac{\zeta_k}{2N}$ in the $ICup(k)$ constraint. We make use of this observation below.

If $ICup(k-1)$ is satisfied in the convexified partition, then all $ICup$ constraints are satisfied. In this case, jump directly to step 2 below. Otherwise, we distinguish between two cases.

Case I. Suppose that $|m_{k-1}| \le \lfloor 2\zeta_k \rfloor$. Merge message m_{k-1} and message m_k (which is now a convex message) into a new message called m_k^{new} with mean μ_k^{new}. For convenience of notation we rename all the other message from m_j to m_j^{new}. We refer to the resulting partition as the "merged partition". This partition, which is illustrated in Figure (7b), is composed of the messages $m_1^{new} \dots m_{k-2}^{new}, m_k^{new}, m_{k+1}^{new}, \dots m_J^{new}$. Notice that:

$$\mu_k^{new} = \mu_k - \frac{\zeta_k}{N} - \frac{|m_{k-1}|}{2N} \tag{13}$$

$$\mu_k^{new} = \mu_{k-1} + \frac{|m_k|}{2N} \tag{14}$$

$$\mu_j^{new} = \mu_j + \frac{\phi_j}{N|m_j|} \qquad \text{for all } j \ge k+1 \tag{15}$$

$$\mu_j^{new} = \mu_j \qquad \text{for all } j \le k-2 \tag{16}$$

$$\overline{m}_k^{new} = \overline{m}_k - \frac{L}{N} \tag{17}$$

where μ_j is the mean of message m_j in the original partition, for all j. To see why Equation (13) holds, notice that μ_k^{new} is equal to the original value of μ_k, minus $\frac{\zeta_k}{N}$ (due to the convexification of m_k), minus $\frac{|m_{k-1}|}{2N}$ (due to merging of m_k with m_{k-1}). Equation (14) holds because the mean of the merged message m_k^{new} is larger than that of the original m_{k-1} by $\frac{|m_k|}{2N}$. Equations (15), (16) and (17) are all direct implications of the convexification of m_k.

In the merged partition, all the $ICup$ constraints are satisfied:

1. $ICup((k-2)^{new})$ is satisfied because $\mu_k^{new} > \mu_{k-1}$, whereas $\mu_{k-2}^{new} = \mu_{k-2}$ and $\overline{m}_{k-2}^{new} = \overline{m}_{k-2}$. Therefore, the fact that $ICup(k-2)$ was satisfied in the original partition, i.e. $\frac{\mu_{k-2}+\mu_{k-1}}{2} - \overline{m}_{k-2} \ge \left(b - \frac{\alpha}{2}\right)$, implies that $\frac{\mu_{k-2}^{new}+\mu_k^{new}}{2} - \overline{m}_{k-2}^{new} \ge \left(b - \frac{\alpha}{2}\right)$.

2. $ICup(k^{new})$ is satisfied because, by Eqaution (13) and since $|m_{k-1}| \le 2\zeta_k$, we have that $\mu_k^{new} \ge \mu_k - \frac{2\zeta_k}{N}$. Thus, the facts that $ICup(k)$ was satisfied in the original partition, i.e. $\frac{\mu_k+\mu_{k+1}}{2} - \overline{m}_k \ge \left(b - \frac{\alpha}{2}\right)$, along with equations (12), (15), (17), imply that $\frac{\mu_k^{new}+\mu_{k+1}^{new}}{2} - \overline{m}_k^{new} \ge \left(b - \frac{\alpha}{2}\right)$.

3. All the other $ICup$ constraints are unaffected by the merge. The fact that they are satisfied in the convexified partition implies that they are satisfied in the merged partition.

We now show that the merged partition yields a higher value of the objective function (5) compared to the original partition. Algebraic manipulation shows that the objective function (5) is equal to the weighted sum of square means of the partition elements

$$\sum_{j=1}^{J} \rho\left(m_j\right)\left(\mu_j\right)^2 \tag{18}$$

up to a constant. We therefore have to show that:

$$\begin{aligned}
&\sum_{j\leq k-2} \rho\left(m_j\right)\cdot(\mu_j^{new})^2 + \rho\left(m_k^{new}\right)\cdot\left(\mu_k^{new}\right)^2 + \sum_{j\geq k+1} \rho\left(m_j\right)(\mu_j^{new})^2 \\
\geq &\sum_{j\leq k-2} \rho\left(m_j\right)\cdot\mu_j^2 + \rho\left(m_{k-1}\right)\cdot\mu_{k-1}^2 + \rho\left(m_k\right)\cdot\mu_k^2 + \sum_{j\geq k+1} \rho\left(m_j\right)\cdot\mu_j^2
\end{aligned}$$

where the left-hand side of the inequality is the value of (18) computed for the merged partition, and the right-hand side is the value of (18) computed for the original partition. Using Equations (15) and (16) above, and because $\rho(m_j) = \frac{|m_j|}{N+1}$, we rewrite the inequality as follows:

$$\frac{2}{N(N+1)} \sum_{j\geq k+1} \mu_j\phi_j + \sum_{j\geq k+1} \rho\left(m_j\right)\left(\frac{\phi_j}{|m_j|\,N}\right)^2 \geq \rho\left(m_{k-1}\right)\cdot\mu_{k-1}^2 + \rho\left(m_k\right)\cdot\mu_k^2 - \rho\left(m_k^{new}\right)\cdot\left(\mu_k^{new}\right)^2.$$

Notice that $\sum_{j\geq k+1} \rho\left(m_j\right)\left(\frac{\phi_i}{|m_j|N}\right)^2 \geq 0$ and $\mu_j > \mu_{k+1}$ for any $j > k+1$. It therefore suffices to show that:

$$\frac{2}{N(N+1)}\cdot\Phi_k\cdot\mu_{k+1} \geq \rho\left(m_{k-1}\right)\cdot\mu_{k-1}^2 + \rho\left(m_k\right)\cdot\mu_k^2 - \rho\left(m_k^{new}\right)\cdot\left(\mu_k^{new}\right)^2.$$

Plugging in $\rho\left(m_{k-1}\right) = \frac{|m_{k-1}|}{N+1}$, $\rho\left(m_k\right) = \frac{|m_k|}{N+1}$, and $\rho(m_k^{new}) = \frac{|m_k|}{N+1} + \frac{|m_{k-1}|}{N+1}$ and rearranging yields:

$$\frac{2}{N}\cdot\Phi_k\cdot\mu_{k+1} \geq -|m_{k-1}|\cdot(\mu_k^{new} - \mu_{k-1})(\mu_{k-1} + \mu_k^{new}) + |m_k|\cdot(\mu_k - \mu_k^{new})(\mu_k + \mu_k^{new}).$$

Using Equations (13) and (14), and since $\zeta_k = \frac{\Phi_k}{|m_k|}$, we rewrite the inequality as follows:

$$2\left(\mu_{k+1} - \mu_k\right)\Phi_k \geq \frac{1}{2}|m_k||m_{k-1}|\left(\frac{\Phi_k}{|m_k|\,N} + \frac{|m_{k-1}|}{2N} + \frac{|m_k|}{2N}\right) - \Phi_k\left(\frac{\Phi_k}{|m_k|\,N} + \frac{|m_{k-1}|}{2N}\right). \tag{19}$$

Finally, we use the fact that ICup(k) is satisfied in the original partition to find a lower bound on $\mu_{k+1} - \mu_k$. To do that, we write ICup(k) equivalently as follows:

$$\mu_{k+1} - \mu_k \geq 2\left(\left(\overline{m}_k^{new} - \mu_k^{new}\right) - \left(\mu_k - \mu_k^{new}\right) + \left(\overline{m}_k - \overline{m}_k^{new}\right)\right) + 2\left(b - \frac{\alpha}{2}\right).$$

The fact that m_k^{new} is a convex message with $|m_{k-1}| + |m_k|$ states implies that $\overline{m}_k^{new} - \mu_k^{new} =$

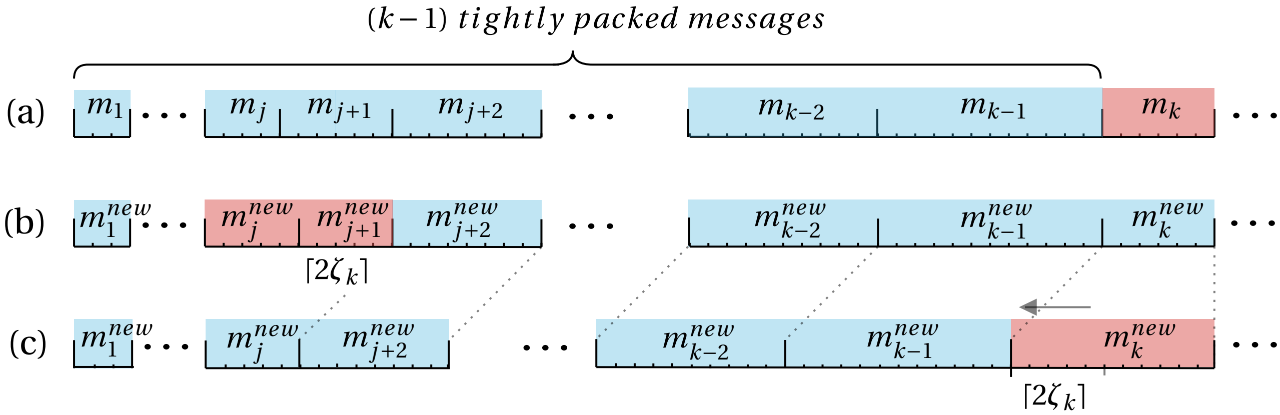


Figure 8: Step 1, Case II

$\frac{1}{2N}(|m_{k-1}|+|m_k|-1)$. Using Equations (12),(13) and (17) we then have that:

$$\mu_{k+1}-\mu_k \geq \frac{|m_k|-1}{N}+2\left(b-\frac{\alpha}{2}\right). \tag{20}$$

By plugging inequality (20) into inequality (19) and simplifying it follows that it suffices to show that:

$$|m_k|^2\,|m_{k-1}|^2+|m_k|^3\,|m_{k-1}|-8\Phi_k\,|m_k|^2+8\Phi_k|m_k|-4\Phi_k^2-16\Phi_k N|m_k|\left(b-\frac{\alpha}{2}\right)\leq 0. \tag{21}$$

The following lemma asserts that this inequality is indeed satisfied. [29]

Lemma A.4 *Inequality (21) is satisfied for all* $|m_{k-1}| \leq \lceil 2\zeta_k \rceil$.

Case II. Suppose that $|m_{k-1}| \geq \lceil 2\zeta_k \rceil$, and that $2\zeta_k$ is not an integer (as otherwise the analysis in case I above applies). Find the index $0 \leq j \leq k-1$ for which $\left|m_j\right| < \lceil 2\zeta_k \rceil \leq \left|m_{j+1}\right|$. For simplicity of notation assume that $m_0 = \emptyset$ and $|m_0| = 0$. Re-partition the union of the two messages $m_j \cup m_{j+1}$ into two new messages: m_{j+1}^{new} with number of states $|m_{j+1}^{new}| = \lceil 2\zeta_k \rceil$ and m_j^{new} with number of states $m_j^{new} = \left|m_j\right| + \left|m_{j+1}\right| - \lceil 2\zeta_k \rceil$. Rename all the other message from m_j to m_j^{new}, as illustrated in Figure (8b). This modified partition weakly *improves* the value of the objective function, compared to the original partition because: (i) the convexifying of m_k weakly decreased the variance of all messages, and (ii) the repartitioning of $m_j \cup m_{j+1}$ into m_j^{new} and m_{j+1}^{new} makes the two messages “more equal” in their number of states compared to m_j and m_{j+1} in the original partition, and so decreases the weighted variance further.

After repartitioning, the constraints $ICup(j^{new})$ and $ICup((k-1)^{new})$ are perhaps violated. To fix this, we eliminate message m_{j+1}^{new} whose length is exactly $\lceil 2\zeta_k \rceil$ as follows: we

[29] Notice that the lemma asserts that the inequality is satisfied for values of $|m_{k-1}|$ that are less than, or equal to, $\lceil 2\zeta_k \rceil$, while in Case I we make the weaker assumption that $|m_{k-1}| \leq \lfloor 2\zeta_k \rfloor$. We use this result in Case II below.

"shift" to the left messages $m_{j+2}^{new}, \ldots, m_{k-1}^{new}$ by $\lceil 2\zeta_k \rceil$ states, and add $\lceil 2\zeta_k \rceil$ states to message m_k^{new} from the left, as illustrated in Figure (8c)[30]

After this modification, all the $ICup$ constraints are satisfied:

1. The constraint $ICup((j-1)^{new})$ is satisfied because $\mu_j^{new} \geq \mu_j$, whereas $\mu_{j-1}^{new} = \mu_{j-1}$ and $\overline{m}_{j-1}^{new} = \overline{m}_{j-1}$. Thus, the fact that $ICup(j-1)$ is satisfied in the original partition implies that $ICup((j-1)^{new})$ is satisfied in the modified partition.

2. The constraint $ICup(j^{new})$ is satisfied. To see this note first that, by construction, $|m_j^{new}| < |m_{j+1}|$ and $|m_{j+2}| = |m_{j+2}^{new}|$. Next, notice that credibility of the original partition, and the fact that m_{j+1} and m_{j+2} are two convex and adjacent messages imply, by Equation (9), that $\frac{|m_{j+1}|}{N} \leq \frac{|m_{j+2}|}{N} - 4\left(b - \frac{\alpha}{2}\right) + \frac{2}{N}$. Therefore, $\frac{|m_j^{new}|}{N} \leq \frac{|m_{j+2}^{new}|}{N} - 4\left(b - \frac{\alpha}{2}\right) + \frac{2}{N}$, which guarantees by Equation (9) that $ICup(j^{new})$ is satisfied.

3. The constraint $ICup((k-1)^{new})$ is satisfied. To see this, note that $\mu_k^{new} = \mu_k - \frac{\zeta_k}{N} - \frac{\lceil 2\zeta_k \rceil}{2N}$ (the convexification of m_k to the left decreased μ_k by $\frac{\zeta_k}{N}$, and the addition of states from the left further decreased the mean by $\frac{\lceil 2\zeta_k \rceil}{2N}$). Furthermore, $\mu_{k-1}^{new} = \mu_{k-1} - \frac{\lceil 2\zeta_k \rceil}{N}$ and $\overline{m}_{k-1}^{new} = \overline{m}_{k-1} - \frac{\lceil 2\zeta_k \rceil}{N}$ due to the shift of messages to the left. Taken together, the last three observations imply that since $ICup(k-1)$ was satisfied in the original partition, then $ICup((k-1)^{new})$ is satisfied in the new partition.

4. The constraint $ICup(k^{new})$ is satisfied. This is because the convexification of m_k to the left implies that $\mu_{k+1}^{new} \geq \mu_{k+1}$. Shifting the messages to the left imply that $\mu_k^{new} = \mu_k - \frac{\zeta_k}{N} - \frac{\lceil 2\zeta_k \rceil}{2N}$ (as explained above) and $\overline{m}_k^{new} = \overline{m}_k - \frac{L}{N}$. Note also that $\frac{1}{2}\left(\frac{\zeta_k}{N} + \frac{\lceil 2\zeta_k \rceil}{2N}\right) \leq \frac{L}{N}$.[31] Taken together, these observations imply that since $ICup(k)$ was satisfied in the original partition, then $ICup(k^{new})$ is satisfied in the new partition.

5. All the other $ICup$ constraints are unaffected by the shift.

The modification improves the value of the objective function compared to the original partition. To see this, recall first that the partition illustrated in Figure (8a), which is the outcome of convexifying message m_k to the left (performed by Part I of the algorithm), improves the value of the objective function relative to the original partition. Next, as explained above, the partition depicted in Figure (8b) improves on the partition depicted in Figure (8a). Finally, inspection of Figure (8c) reveals that it consists of messages with the same number of states as the partition in depicted in Figure (8b), except for message m_k^{new} in Figure (8c), which can

[30]We say that a convex message m is shifted to the left by x states if $\underline{m}^{new} := \underline{m} - x$ and $\overline{m}^{new} := \overline{m} - x$ where m^{new} denotes message m after the shift.

[31]To see this, suppose that $\Phi_k = |m_k|L - x$ for some (integer) $x \geq 0$. Then $\frac{1}{2}\left(\frac{\zeta_k}{N} + \frac{\lceil 2\zeta_k \rceil}{2N}\right) = \frac{1}{2}\left(\frac{L}{N} - \frac{x}{N|m_k|} + \frac{1}{2N}\left\lceil 2L - \frac{2x}{|m_k|}\right\rceil\right) \leq \frac{L}{N} - \frac{1}{2}\frac{x}{N|m_k|}$.

be viewed as a merge between messages m_k^{new} and m_{j+1}^{new} in Figure 8(b). It is useful to perform this merge in two steps: first, shift message m_{j+1}^{new} to the right so that it lies between messages m_{k-1}^{new} and m_k^{new} in Figure (8b); and then, merge messages m_{j+1}^{new} and m_k^{new} as illustrated in Figure (8c). Because the number of states in message m_{j+1}^{new} is exactly $\lceil 2\zeta_k \rceil$, the argument used in Case I above (and Lemma A.4) can be applied here, where m_{j+1}^{new} takes the place of message m_{k-1} in the argument presented in Case I.

Step 2: Show that Part II of the algorithm improves the objective function's value further
Part I of the algorithm, followed by the modifications described above (according to Case I or Case II), produce a partition with convex messages on the set $\{0,\ldots,l_k\}$ that satisfies all the $ICup$ constraints and improves upon the value of the objective function compared to the original partition. The next lemma asserts that executing Part II of the algorithm on this partition preserves all the $ICup$ constraints and further improves the value of the objective function.

Lemma A.5 *Let P be a partition that satisfies all the $ICup$ constraints with $\hat{J}$ convex messages on the set of states $\{0,\ldots,l_{\hat{J}}\}$. Then, tightly packing the messages on the set $\{0,\ldots,l_{\hat{J}}\}$ in a maximal manner preserves all the $ICup$ constraints and improves the value of the objective function.*

Finally, to complete the proof of the proposition, notice that when $I(1)$ messages are maximally tightly packed on Ω_N then all the $ICup$ constraints are binding (by definition). In this case, all the $ICdown$ constraints are satisfied as well. To see this, fix j and notice that

$$\frac{\mu_{j-1}+\mu_j}{2} - \underline{m}_j < \frac{\mu_{j-1}+\mu_j}{2} - \overline{m}_{j-1} < b - \frac{\alpha}{2} + \frac{1}{2N}$$

where the first is by definition and the second inequality follows from the fact that the $ICup(j-1)$ constraint is binding. It follows that for large enough N, we have that $\frac{\mu_{j-1}+\mu_j}{2} - \underline{m}_j < b + \frac{\alpha}{2}$. This completes the proof of the proposition. ■

Proof of Lemma A.4

The left-hand-side of (21) is quadratic and convex in $|m_{k-1}|$. Therefore, to verify that (21) is satisfied for all $|m_{k-1}| \leq \lceil 2\zeta_k \rceil$ it suffices to check that it is satisfied for $|m_{k-1}| = 0$ and for $|m_{k-1}| = \lceil 2\zeta_k \rceil = \left\lceil \frac{2\Phi_k}{|m_k|} \right\rceil$.

Verifying that (21) is satisfied for $|m_{k-1}| = 0$ is straightforward. To verify that (21) is satisfied for $|m_{k-1}| = \left\lceil \frac{2\Phi_k}{|m_k|} \right\rceil$, suppose first that $\frac{2\Phi_k}{|m_k|}$ is an integer. In this case, substituting $|m_{k-1}| = \frac{2\Phi_k}{|m_k|}$ into inequality (21) yields:

$$2\Phi_k|m_k|\left(4-3|m_k|-8N\left(b-\frac{\alpha}{2}\right)\right)\leq 0$$

which is satisfied for all values of $|m_k|$ when $N > 1/(8(b-\frac{\alpha}{2}))$.

Suppose next that $\frac{2\Phi_k}{|m_k|}$ is not an integer. Notice that in this case $|m_k| \geq 3$. To verify that (21) is satisfied for $|m_{k-1}| = \left\lceil \frac{2\Phi_k}{|m_k|} \right\rceil$ it suffices to check that it is satisfied for $|m_{k-1}| = \frac{2\Phi_k}{|m_k|} + 1$. Substituting $\Phi_k = \frac{|m_{k-1}||m_k|-|m_k|}{2}$ into (21) yields:

$$|m_k|^2\left(4|m_k|-3|m_{k-1}|\left(|m_k|-2\right)-5-8N\left(|m_{k-1}|-1\right)\left(b-\frac{\alpha}{2}\right)\right)\leq 0$$

Recall that, by assumption, message $|m_{k-1}|$ contains at least two states, i.e., $|m_{k-1}| \geq 2$. Thus, the last inequality is satisfied for all $|m_k| \geq 3$ when $N > 1/(8(b-\frac{\alpha}{2}))$.

Proof of Lemma A.5

Suppose that messages 1 through $\hat{J}$ are not tightly packed. It follows that the $ICup(j)$ constraint is not binding for some message m_j, $j < \hat{J}$. In this case, it is possible to re-assign the smallest state in message m_{j+1} into message m_j in a way that satisfies all the $ICup$ constraints (because the $ICup(j)$ constraint is not binding and the change simultaneously increases both μ_j and μ_{j+1}). This reassignment improves the value of he objective function (5) because it moves the number of states in messages m_j and m_{j+1} closer together, which decreases their weighted variance. This implies that tightly packing the $\hat{J}$ messages on states $\{0,\ldots,l_{\hat{J}}\}$ satisfies all the $ICup$ constraints (by definition) and improves the value of the objective function.

If the messages 1 through $\hat{J}$ are tightly packed, but not maximally tightly packed, then maximally tightly packing messages into states $\{0,\ldots,l_{\hat{J}}\}$ satisfies all the $ICup$ constraints and improves the value of the objective function.

To see this, suppose that $P = (m_1^P,\ldots,m_k^P)$ and $Q = (m_1^Q,\ldots,m_{k+1}^Q)$ are two tightly packed partitions with k and $k+1$ elements, respectively, on the set $\{0,\ldots,\hat{\omega}\}$. Denote the value of the objective function (5) restricted to the set $\{0,\ldots,\hat{\omega}\}$ that is induced by these two partitions by $V(P) = \sum_{i=1}^{k}\rho(m_i^P)\cdot\mathrm{Var}(m_i^P)$ and $V(Q) = \sum_{i=1}^{k+1}\rho(m_i^Q)\cdot\mathrm{Var}(m_i^Q)$, respectively, where $\mathrm{Var}(m_i)$ denotes the variance of the (convex) message m_i.

Notice that since all the $ICup$ constraints are binding in both P and Q, then $|m_i^P| >$

$|m^Q_{i+1}| > |m^Q_1|$ for all $1 \le i \le k$. It follows that

$$\begin{aligned}
V(P) &= \sum_{i=1}^{k} \rho(m^Q_{i+1}) \cdot \mathrm{Var}(m^P_i) + \sum_{i=1}^{k} \Big(\rho(m^P_i) - \rho(m^Q_{i+1})\Big) \cdot \mathrm{Var}(m^P_i) \\
&\ge \sum_{i=1}^{k} \rho(m^Q_{i+1}) \cdot \mathrm{Var}(m^Q_{i+1}) + \sum_{i=1}^{k} \Big(\rho(m^P_i) - \rho(m^Q_{i+1})\Big) \cdot \mathrm{Var}(m^Q_1) \\
&= \sum_{i=2}^{k+1} \rho(m^Q_i) \cdot \mathrm{Var}(m^Q_i) + \left(\hat{\omega} - \sum_{i=2}^{k+1} \rho(m^Q_i)\right) \cdot \mathrm{Var}(m^Q_1) \\
&= V(Q)
\end{aligned}$$

where the inequality follows from the fact that the variance increases in the number of states in a convex message.

Finally, the fact that the $ICup(k)$ constraint is satisfied in partition P, and the fact that $\mu^Q_{k+1} > \mu^P_k$ imply that the $ICup(k+1)$ constraint is satisfied in partition Q. Hence, partition Q satisfies all the $ICup$ constraints.