Society for Economics Research in India
Working Paper Series

COMPETITIVE DISCLOSURE OF INFORMATION TO A RATIONALLY INATTENTIVE CONSUMER

Vasudha Jain
(Indian Institute of Technology, Kanpur)
Mark Whitmeyer
(Arizona State University, Arizona)

Working Paper No. 21
http://seri-india.org/research

2021

Competitive Disclosure of Information to a Rationally Inattentive Consumer

Vasudha Jain* Mark Whitmeyer†

Abstract

Firms strategically disclose product information in order to attract consumers, but recipients often find it costly to process all of it, especially when products have complex features. We study a model of competitive information disclosure by two firms vying for a consumer who may save on attention costs by systematically ignoring some information. We find that for a large class of parameters, it is an equilibrium for the firms to provide the consumer with her first-best level of information–i.e., as much as she would learn if she herself controlled information provision. Importantly, this is not true if the consumer could costlessly process information, or if there were only one firm. Our key contribution is to identify a novel channel through which an interaction of firm competition with consumer inattention encourages information disclosure: information on one firm substitutes for information on the other, which nullifies the profitability of a unilateral withholding of information. Our results have implications for a broad range of applications.

Keywords: Bayesian persuasion; Information design; Multiple senders; Competition; Rational Inattention; Search

JEL Classifications: D82; D83

*Corresponding author. Department of Economic Sciences, Indian Institute of Technology Kanpur. Email: vasudhaj@iitk.ac.in.

†Department of Economics, Arizona State University. Email: mark.whitmeyer@gmail.com.

The first draft of this paper was completed when both authors were at the University of Texas at Austin. Thanks to Vasiliki Skreta for extensive feedback, and to V. Bhaskar, Max Stinchcombe, Caroline Thomas and Tom Wiseman for helpful discussions. Thanks also to participants at numerous seminars and conferences for thoughtful comments and questions.

Constant attention wears the active mind,
Blots out our powers, and leaves a blank behind.

Charles Churchill
Epistle to William Hogarth

1 Introduction

Firms often seek to alter buyers' perceptions of the quality of their products by strategically controlling the availability of information. For instance, they can set policies on how long a consumer can try a product, or how customer reviews posted online are moderated. However, products often have complex features, and processing information about them can entail substantial investment of time and cognitive resources. While more precise information helps consumers make better decisions, it often comes at a higher implicit cost.

It is natural to expect that firms, when crafting their information disclosure strategies, account for the fact that making information available need not suffice to influence beliefs– consumers must find it in their interest to invest in processing that information. Further, if a firm competes with other firms, then any disclosure decision may influence not only its competitors' incentives to disclose information about their own products, but also how consumers would choose to allocate their limited attention to learning about each product.

In this paper, we explore the strategic forces that operate in this environment. Specifically, we study how much information about product quality would be disclosed by two firms that compete to sell to a consumer who is known to be *rationally inattentive*, in that she endogenously chooses how much information about each firm to process, at a cost increasing in the amount of that information. A firm's disclosure decision thus places an upper bound, but not a lower bound, on the consumer's information.

Two elements lie at the heart of our analysis. The first is *substitutability* between sources of information for the consumer: highly precise information on one firm diminishes the value of information on the other. To understand this, note that upon acquiring information about one firm, information about the other is valuable to the consumer only if it changes her belief about which one is better. Now suppose that information obtained about the first firm is highly precise–e.g., say that the consumer is 95% sure that its quality is high. Now, she would select the second firm only if she learns that its quality is even likelier to be high. However, such precise information about the second firm would come at a high cost and would have relatively low value, because it would move the probability of the selected product having high quality by at most 5 percentage points.

The second key element is an implication of the consumer's information processing costs: it might not be optimal for the consumer to become *certain* of either firm's quality, even when such precise information is available (and can be processed at a finite cost). As a result, if both firms were to provide full (i.e., perfectly accurate) information on the quality of their respective products, the consumer might rationally ignore some of each firm's information, *thereby leaving scope to learn more about each of them.* We refer to the amount of information (aggregated across firms) that the consumer optimally processes in this scenario as her *first-best* level of information.

Naturally, by withholding information, either firm might be able to prevent the consumer from achieving this first-best. That is, it might be able to force the consumer to operate under a higher degree of uncertainty about product quality than she would like. **The main question we ask** in this paper is whether there is an equilibrium in which firms choose *not* to withhold information in this manner.

Propositions 4.2 and 4.3 tell us that so long as information processing costs and prior uncertainty about quality are not too low, the answer is yes. That is, it *is* an equilibrium for the firms to provide the consumer with as much information as she would process if she had potential access to full information.

Importantly, we find that the consumer might be able to elicit more information from each firm (and might end up buying from a higher quality firm on average) relative to a scenario where information processing is costless, and also relative to a scenario where information processing is costly but there is only one firm. Therefore, *both ingredients in our model–competition and information processing costs–are critical for our results*, and the key contribution of this paper is to uncover, in a stylized model, a novel strategic force that encourages greater disclosure of information when these elements combine. Indeed, substitutability between information sources arises only when there is more than one firm, and it is only due to information processing costs that the consumer finds it optimal to "leave some information on the table".

Our model is useful to understand how firms interact not only with individual consumers, but also with specialists. Consider, for instance, the situation encountered by doctors and pharmaceutical companies. Patients rely on their doctors to make important medical decisions for them, such as the decision of which medication to take. Often, multiple drugs exist to treat the same condition, but nevertheless differ in subtle ways that can prove crucial for patients. Pharmaceutical companies conduct clinical trials to produce information on the safety and efficacy of their drugs, and make this information available to doctors through articles in medical journals, promotional pamphlets etc. Although they are prohibited from falsifying facts, they may strategically decide how much information to reveal and in what form. For example, some important but subtle details–such as whether adverse side effects had led many clinical trial subjects of a certain demographic group to drop out midway–may be omitted or buried in footnotes. Such situations are ubiquitous in present times, when doctors find themselves inundated with information on COVID-19 treatments.

A well intentioned doctor has her task clearly cut out–she should study all published material made available to her, and let that information guide her prescription decisions. However, absorbing all details involves substantial time and effort, and doctors typically find it difficult to keep up. Tellingly, Alper et al. (2004) find that it would take a doctor *six hundred hours* to skim all research relevant to general practice that is published in just one month. Consequently, they are likely to, e.g., pay attention only to some published summary statistics, and skip the kinds of subtle details referenced earlier.

Pharmaceutical companies, when choosing their disclosure strategies, take into consideration the lack of attention on the part of the recipients: they may design pamphlets in a way that the most favorable pieces of evidence stand out, or other strategies of that ilk. As Goldacre (2014) explains, "They (doctors) need good quality information, but they need it, crucially, under their noses. The problem of the modern world is not information poverty, but information overload...So doctors will not be going through every trial, about every treatment relevant to their field...They will take shortcuts, and *these shortcuts can be exploited* [emphasis added]".

In our baseline model, there are two senders (e.g. firms), each with a stochastic binary type (quality) drawn independently.[1] Prior to learning anything, each of them decides how much information to disclose about his own type.[2] Rather than imposing a specific (and inevitably restrictive) information generation technology, we follow the influential work of Kamenica and Gentzkow (2011) in allowing *flexible* choice of disclosure rules. This simply means that each sender may choose any Blackwell experiment, which corresponds to a distribution of posterior beliefs whose average is the prior belief. For example, the distribution of posterior beliefs corresponding to full information has binary support (since types are binary), and that corresponding to no information has degenerate support.

[1] Our focus is on strategic provision of information on quality, and so we exogenously fix prices at 0 .

[2] Note in particular that a sender has no control over information disclosed about his competitor.

There is a receiver (e.g. consumer or doctor), who wishes to choose the sender with a higher type. She visits the senders sequentially. When she visits the first one, she observes how much information has been made available by him, but may choose to acquire less precise, or coarser, information than that. Formally, she is free to choose any mean preserving contraction–or *garbling*–of the sender's experiment, and a draw from that garbling determines her posterior belief about that sender.

The reason she might want to undertake this garbling is that obtaining more precise information is costly: her attention costs are lower for a less informative garbling. The receiver has to balance this reduction in attention costs against the worsening quality of information on a decision-relevant variable. Think back to the doctor example, and the shortcuts she might take: she might read just the first few pages of an article, only the nontechnical parts, only the technical sections, or even just the title. All of these correspond to different levels of information, and all of these impose on the receiver different costs–a grueling slog through a complicated model takes more out of the receiver than does a quick skim of the conversational portions.

With the first posterior in hand, the receiver decides whether to visit the second sender. Importantly, we do not impose that she must visit the second sender in order to choose him. If she does decide to visit him, the protocol is identical to that for the first sender: she chooses a garbling of his chosen experiment subject to an information cost. Finally, she chooses the sender favored by her posterior beliefs.[3] Each sender wants to maximize the probability of being chosen.

As we show, the receiver's learning strategy has an intuitive feature that drives our analysis: it can happen that she will have "seen enough" at the first sender and need not visit the second sender–her belief about the first sender may be so high that she chooses him without ever visiting the second, and it may be so low that she chooses the second, sight unseen. Returning to the example: if a doctor is fairly certain that drug A is of low quality, then she would be willing to prescribe drug B without learning anything about it, and vice-versa.

In this setting, a pertinent benchmark is what the receiver would do if she had potential access to full information on each sender, so that attention costs were the only factor potentially limiting her learning.[4] We show that in this case–the *first-best* scenario for her–she would always learn *something* from at least one sender, but never learn any sender's type

[3]If she does not visit a sender, her posterior is equal to the prior.

[4]This is equivalent to an environment where the receiver could herself control how much information is provided by senders.

with *certainty*. Furthermore–and this is crucial–fixing any prior, for a high enough attention cost parameter, it is optimal for her to learn from exactly *one* sender.

The main question we address in this paper is whether in our game with strategic senders, there is an equilibrium in which the receiver ends up with this first-best outcome. That is, we ask whether senders might voluntarily provide as much information as the receiver would acquire in her first-best scenario.

We find that the answer is yes under general conditions. In particular, for any prior, there is such an equilibrium so long as an attention cost parameter is above a threshold.[5] This is a departure from closely related models in the literature, where either there is only one sender and the receiver faces attention costs, or there are two senders but no attention costs: in those environments, the consumer is unable to obtain her first-best in equilibrium.

Our analysis produces a sharp economic insight into why a combination of these ingredients–competition and attention costs–gives us more information disclosure. Recall our observation that for high enough costs it is optimal for the receiver to learn from exactly one sender. Now suppose that the sender from whom she plans to learn, unilaterally deviates and restricts her learning. Then since the other sender continues to provide full information, the receiver could just switch to learning from him instead. Her *ex ante* payoffs, and the probability of choosing correctly between the senders, would remain unaffected. Since a sender's payoffs ultimately depend only on this probability, the deviation ends up being unprofitable for him.

The result is driven by the fact that for the receiver, the two sources of information are partial substitutes, and due to attention costs she never learns *fully* from either source. Then, in the event of a unilateral provision of less information by one sender, she has the option of paying more attention to the non-deviating sender. In a large range of circumstances, she is able to do so in way that maintains the likelihood of choosing correctly between the senders. As discussed ahead, this insight is novel in the literature.

Our results are robust to a variety of modifications of the model. While we concentrate our attention on experiments that are unobserved by the receiver until her visit (since that fits our leading example best), our main results go through if the experiments are chosen publicly by the senders. As we discover, if anything this strengthens our main results: publicly chosen experiments make it easier for the receiver to see, and hence react to, a sender's deviation. Similarly, for ease of exposition, for most of the paper we focus on the symmetric case in which both senders are identically likely to be the high type. Again, we show that this restriction is

[5]Equivalently, fixing attention costs above a threshold, this is true over an interior interval of prior means. The interval expands as attention costs grow, and approaches the full range as they explode.

not needed for inattention to beget the first-best level of information, and that a result of a similar ilk holds for heterogeneous means (that are not too far apart). Finally, we modify the receiver's information cost function, itself, to allow it to depend on the distribution chosen by the sender. Like the public experiments modification, this only supplements the incentives that drive our main results, which thus continue to hold.

1.1 Related Literature

To the best of our knowledge, this paper is the first to look at competitive information design with information processing costs faced by the receiver. This relates thematically to several strands of the literature.

Rational inattention. Since in our model, the receiver's decision to garble a sender's experiment is the result of an optimization problem that accounts for attention costs, she is *rationally* inattentive as in the economics literature pioneered by Sims (2003). Mackowiak et al. (2020) provide an excellent review of this literature and its wide range of applications, and cite experimental and empirical work suggesting that agents behave in ways consistent with the predictions of these models.

Information disclosure by a single sender to a rationally inattentive agent. The specific framework of rational inattention that we adopt follows Lipnowski et al. (2020a) and Lipnowski et al. (2020b). These papers consider the problem of a principal whose preferences over actions are perfectly aligned with those of an agent who privately bears a cost of paying attention. The former paper establishes conditions under which the principal would want to restrict the agent's information with a view to manipulating her attention. The latter imposes more structure on the problem and characterizes the principal's optimal disclosure rule.

Wei (2020) extends this analysis to an environment with preference misalignment between the principal and the agent. He considers a binary types, binary action model with a single principal who has state independent preferences, and an exogenous threshold of acceptance for the agent. It is shown that the principal necessarily finds it in his interest to restrict the agent's learning. In the present paper, we show how competitive forces change this.

Bloedel and Segal (2020) take a different approach to a problem similar to Wei's. In their framework, after observing the sender's experiment, but before seeing its realization, the receiver can choose a mapping from signal realizations to distributions over "perceptions", incurring an entropy reduction cost. Then, the receiver observes the realized perception, and not the actual signal realization. As Lipnowski et al. (2020a) explain, this is conceptually different from our paper (and theirs), since the receiver in our model pays a cost to reduce uncertainty about the *state*, and not the sender's message. Both Matyskova (2018) and Montes (2020) study a persuasion model where the receiver, after observing the realization from the sender's signal, can acquire additional information on the state at a cost proportional to the reduction in entropy.

Competitive information disclosure without rational inattention. Our work is also closely related to papers on competitive information design without any attention costs. With two senders, this has been studied by Boleslavsky and Cotton (2015), who identify the unique equilibrium. Crucially, providing full information is *not* an equilibrium with zero attention costs, and we show that this continues to hold for positive but small attention costs (although, as we show, the reason for this non-existence is very different).

Au and Whitmeyer (2021) extend the competitive persuasion scenario to a sequential (directed) search setting, and allows for fixed search (visit) costs (but assumes costless information processing, so that there would be no incentive to learn less than is possible). He finds that although search costs may also encourage information provision by the sellers, the mechanism is different to that of this manuscript. There, the information policies are publicly posted by the sellers *ex ante*, and it is the resulting competition between sellers to be visited first that encourages them to be informative (analogous to how publicly posted prices encourage competition in models of price competition when the consumer's search is costly–see, e.g., Armstrong (2017)). In contrast, here, the information processing cost enables the searcher to make a credible threat that sustains the full information equilibrium regardless of whether information policies are posted publicly or not–if a seller deviates, the searcher will learn nothing from him, rendering that deviation unprofitable.

Some other papers in the competitive information design literature that bear mentioning are Au and Kawai (2020), Boleslavsky and Cotton (2018) and Albrecht (2017). The result that competition encourages information disclosure is familiar from Au and Kawai (2020),[6] but introducing information processing costs suggests a completely different channel for why this might be true: the fact that multiple information sources serve as substitutes matters only in the presence of information processing costs. Moreover, in the absence of costs, providing *full* information is not an equilibrium with two (or any finite number of) senders, as it is in our model.

Board and Lu (2018) also look at sellers competing through information to entice buyers.

[6]The same is also true of some papers in the *cheap talk* literature, e.g. Battaglini (2002), which have a very different flavor.

However, there, search is random, the number of sellers is uncountable, and the decision by a seller of how much and what sort of information to disclose is made upon the buyer's visit. Thus, the problem is different from the scenario analyzed here (or in Au and Whitmeyer (2021)), where the searcher must choose whom to visit as part of a stopping problem. Moreover, in Board and Lu (2018) the value of each seller's goods to the buyer are perfectly correlated whereas here they are independent. Accordingly, one of the key inputs of their model–how much a seller can observe about the consumer's belief–is absent here.

Consumer search. There is a sizable literature on consumer search, and papers such as Diamond (1971) and Burdett and Judd (1983) study price equilibria in that context. In this literature, consumers face fixed costs of finding a firm, but once a firm is visited, it is costless to learn about its product. Matějka and McKay (2012) incorporate rationally inattentive consumers into such a framework, but restrict attention to price competition rather than information disclosure by firms.

Costs borne by sender. Another group of papers looks at what happens if there is no competition, but costs are on the sender's side instead of the receiver's. Gentzkow and Kamenica (2014) look at optimal persuasion mechanisms when the sender pays higher costs (proportional to entropy reduction) of designing more informative experiments. Likewise, Le Treust and Tomala (2019) consider constraints on the sender's information transmission channel. Interestingly, they find that the sender's payoff from the optimal solution is the concave closure of his payoff function, net of entropy reduction costs. Thus, these costs arise endogenously in their model.

The rest of this paper is organized as follows. Section 2 presents our baseline model. Section 3 presents results for the benchmark with a single sender. Section 4 presents the equilibrium analysis with two senders and spells out how the level of attention costs matters. Section 5 illustrates the robustness of our results to the various modifications mentioned in the introduction, and Section 6 concludes. The Appendix contains proofs that are not presented in the main text.

2 Baseline Model

There are two senders indexed by $i \in \{1, 2\}$, and a receiver. Sender i has type $\omega_i \in \Omega_i \coloneqq \{0, 1\}$, with the types being drawn independently. The common prior belief is that $\Pr(\omega_i = 1) = \mu \in (0, 1)$ for $i \in \{1, 2\}$.

The receiver has to select one of the two senders, and she has no outside option.[7] Her payoff is equal to the type of the selected sender, minus *attention costs* that we elaborate on below. Sender i's payoff is 1 if he is selected, and 0 if not. All players maximize expected payoffs.

Timing. The game proceeds in the following 3 stages.

Stage 0: Each (*ex ante* uninformed) sender simultaneously commits to a Blackwell experiment that generates information about his own type. Such an experiment is a mapping from $\{0,1\}$ to the set of Borel probability measures over a compact metric space of signal realizations. Each signal realization, then, is associated with a posterior belief distribution on $\{0,1\}$, and an experiment induces a distribution over posterior beliefs. Hereafter, we identify a posterior belief with the belief on $\omega_i = 1$.

From the work of Kamenica and Gentzkow (2011), we know that the set of Blackwell experiments is isomorphic to the set of distributions of posterior beliefs whose average is the prior. Thus, at this stage 0, sender i commits to a distribution $p_i \in \Delta[0,1]$, with $\int_{[0,1]} x \, dp_i(x) = \mu$.

Stage 1: The receiver, who at this point *does not* observe the chosen distributions,[8] decides whether to visit any sender and, if so, which one.

Say she visits sender 1 first. Upon visiting she observes 1's distribution p_1, and is free to choose any $q_1 \in \Delta[0,1]$ that is a mean preserving contraction (or garbling) of p_1.[9] She takes a draw from q_1, which determines her posterior belief about sender 1.

Associated with any such q_1 is an **attention cost** that is proportional to the variance of the receiver's posterior:

$$C(q_1) = \int_{[0,1]} k(x-\mu)^2 dq_1(x), \tag{1}$$

where $k > 0$. Note that costs depend on q_1 and not directly on p_1. We defer a discussion of these costs to Section 2.1.1.

Stage 2: The receiver then decides whether to visit sender 2. If she does, she observes p_2 and chooses a garbling q_2, once again incurring an attention cost $C(q_2)$. She takes a draw from q_2, which determines her posterior belief about this sender. Finally, she selects the sender for whom her posterior belief is higher.[10] She need not have visited a sender or learned

[7] Our results hold with an outside option, as long as its expected quality is not too high.

[8] See section 2.1.2 for a discussion of this assumption.

[9] q is a garbling of p if the random variable associated with q second order stochastically dominates–and has the same mean as–the random variable associated with p. It is a strict garbling if additionally $q \neq p$. Trivially, $q_1 = p_1$ or $q_1 = \delta_\mu$ is always an option.

[10] As long as she learns something from at least one sender, the posteriors would never be equal. (See

anything from him in order to select him.

Figure 1 illustrates the sequence of moves in the game.

Notice that the receiver's optimal garbling at stage 2 potentially depends on the belief she draws at stage 1. She may be more or less inclined to learn about the second sender, depending on how much uncertainty has already been resolved about the first one. Indeed, as we shall see, if the stage 1 belief is close enough to 0 or 1, she chooses not to learn at all at stage 2, and this fact plays a crucial role in our analysis.

The distribution offered by the sender visited first dictates how much can be learned at stage 1. Then in light of the preceding observation, if both senders offer different distributions, the choice of whom to visit first (if anyone) matters for payoffs.

Before proceeding, we point out the following **characterization of the set of garblings of a binary distribution**, which we shall extensively use:

q is a garbling of a distribution with support $\{\nu_1, \nu_2\} \iff supp(q) \subseteq [\min\{\nu_1, \nu_2\}, \max\{\nu_1, \nu_2\}]$.

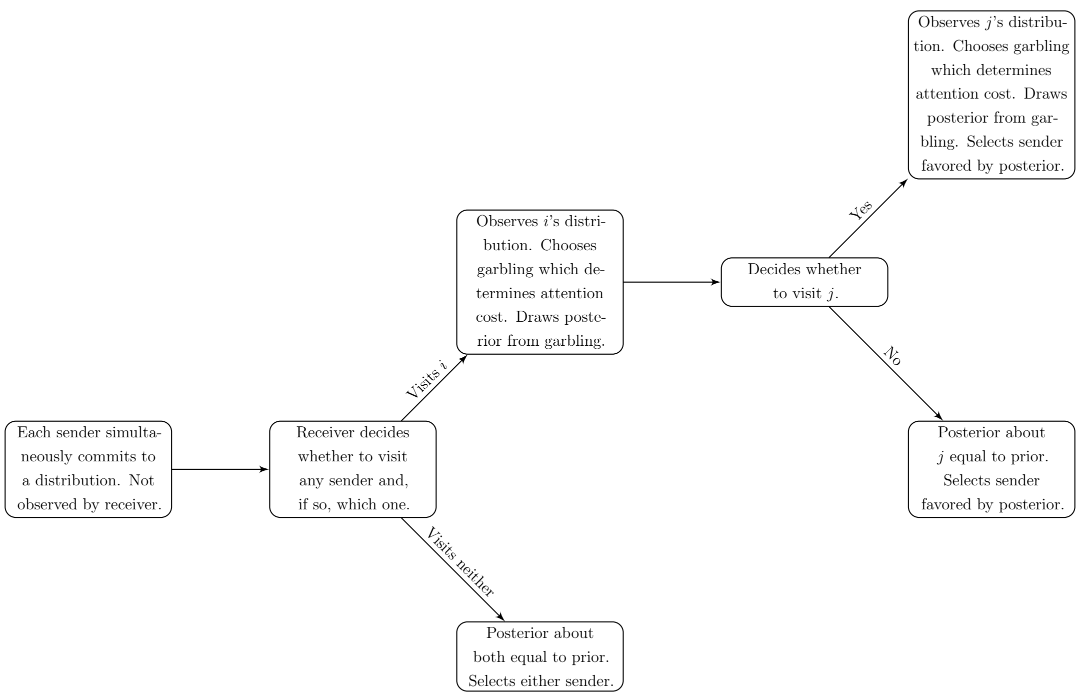

Figure 1: Sequence of moves

footnote 17.) They would be equal only if she does not learn from either sender–in that case she may choose between the senders in any way.

Strategies and solution concept. A pure strategy for sender i is a choice of a distribution $p_i \in \Delta[0,1]$ whose average is μ. A pure strategy for the receiver consists of i) a choice of which sender to visit first, if any; ii) a choice of garbling for any distribution offered by the sender she visits first; iii) a choice of whether to visit the second sender for each belief drawn in the previous stage; iv) a choice of garbling for the second sender, for each distribution offered by him and each posterior belief drawn in the previous stage; v) a choice of which sender to select at the end for each pair of posterior beliefs.

Throughout, we restrict attention to pure strategies, with one exception, *viz.*, we allow the receiver to randomize over the order of visits.

At stage 0, the receiver does not observe the senders' distributions, but forms beliefs about them. These may be updated in the course of the game: firstly, upon visiting a sender, she becomes certain of the distribution chosen by him; secondly, upon visiting the first sender she can in principle update her belief about the second sender's distribution too. Further, upon visiting a sender and observing his choice of experiment, she may in principle update her belief about either sender's type.

In equilibrium, we require the receiver's beliefs to be given by Bayes' rule where possible, and for every player's behavior to be sequentially rational. Further, we impose a condition similar in spirit to the "no-signaling-what-you-don't-know" condition (defined by Fudenberg and Tirole (1991) for multi-period games with observed actions): upon visiting a sender and learning that he has deviated, the receiver does not update her belief about the other sender's experiment or about either sender's type.[11]

2.1 Discussion of modeling choices

2.1.1 Attention costs

Attention costs, in our framework, are costs incurred to process information on a sender's type. Through his choice of a Blackwell experiment, a sender can control how much information on his type is available–in other words, he can put a cap on what can be learned. But a recipient may choose to ignore some of that information and take a draw from a less informative experiment, thereby reducing attention costs. For instance, a pharmaceutical company can decide how much research on its drug to publish, but a doctor might choose to read a subset

[11] Our requirements are akin to subgame perfection, but formally we cannot use subgame perfect equilibrium as our solution concept because the continuation game following the receiver's arrival at the first sender is not a proper subgame. Therefore, we use perfect Bayesian equilibrium with additional restrictions on off-path beliefs.

of that. Her costs would depend on how much of the research she chooses to read, not on how much was published. In particular, both the act of understanding how much information is available, and the act of garbling, are *per se* costless.

To capture this notion, we posit an attention cost function that takes as input only the distribution from which the receiver draws her posterior belief.[12] Accordingly, the cost specified by equation 1 is simply proportional to the variance of the posterior belief under the receiver's chosen garbling.

This cost function is *posterior separable* as in Caplin et al. (2019). That is, associated with each posterior x is a cost $k(x-\mu)^2$, so that more precise beliefs–those that are further away from the prior–cost more, and this is integrated to determine the cost of a distribution of posteriors.

Since $k(x-\mu)^2$ is strictly convex, by Jensen's inequality we have

$$q \text{ is a garbling of } p \implies C(q) \leq C(p),$$

with the inequality strict for strict garblings.[13] For instance, $C(q)$ is minimized for the uninformative distribution δ_μ, and maximized for the fully informative one with support $\{0,1\}$.

Clearly then, the receiver faces a trade-off in her choice of garblings q_1 and q_2–a garbling costs less, but also corresponds to a less informative experiment and is less valuable for her decision problem (Blackwell 1951, 1953). Returning to our example, the more extensive or detailed the research a doctor chooses to read, the costlier it is to draw an inference from it; but also, the more confidence she can place in that inference.

2.1.2 Privately chosen experiments

Recall that at stage 0, the receiver does not observe the experiments chosen by the senders. It is important to bear in mind that despite this assumption, the senders' distributions are correctly anticipated by the receiver *in equilibrium*. In fact, the only substantive implication of this assumption is that a sender's *deviation* is not observed until he is visited. While this assumption plays a role in our analysis of the baseline model, we show in section 5.1 that our main results go through even if experiments are public at stage 0.

[12]Of course, the distribution offered by a sender indirectly matters by restricting the set of distributions available to the receiver.

[13]Of course, this property holds for any strictly convex function instead of $k(x-\mu)^2$, but we work with this specification for tractability, as in Lipnowski et al. (2020b).

2.1.3 Independent learning

In our model, the senders choose independent experiments and the receiver learns about their types via a pair of independent experiments that she chooses sequentially.

However, notice that the receiver only cares about information on which sender is better, i.e. on the *difference* in realized types. One can conceive of an alternative model where she is in fact allowed to learn directly about the difference in quality–i.e., to introduce correlation between the two experiments she learns from.[14] An option to do so might be valuable for the receiver by reducing the expected cost of acquiring information of a certain gross value.

However, we disallow this option here to keep the model consistent with applications of interest: in all of the examples that motivate this paper, there is no practical way for the receiver to introduce such correlation–she must visit one sender at a time, learning about him independently of the other available alternative, and incurring attention costs separately.

3 Benchmark: Single Sender

We begin by taking a brief look at what happens if there is only one sender. The receiver chooses a garbling of that sender's distribution and accepts his product if the belief drawn from it is above a threshold $\lambda \in (0, 1)$. Payoffs clearly depend only on the distribution finally chosen by the receiver. In any (subgame perfect) equilibrium, the sender offers a distribution to maximize his expected payoff, correctly anticipating the receiver's optimal garbling behavior.

Following the arguments in Wei (2020), the setup with a single sender permits two simplifications that are not valid in our two-sender model. One: to obtain the set of equilibrium outcomes, it is without loss of generality to restrict the sender to incentive compatible distributions–those that the receiver would not want to garble. This leads to the second simplification, which is that it is without loss to restrict him to binary and degenerate distributions. (Since the receiver has only two actions, she never wants to pay to generate more than two beliefs.)

We refer to a distribution offered by the sender in a sender-preferred equilibrium as *sender-optimal.*

Fixing a prior belief μ on the sender's type, if $\lambda < \mu$, it is immediate that any sender-optimal distribution is such that nothing is learned and he is accepted with certainty. For

[14] This is found, for instance, in Matějka and McKay (2015), where an agent has to choose between N alternatives that have stochastic values, and where full information is available but attention is costly.

example, he can simply design an uninformative experiment.

Now say $\lambda > \mu$. If $k = 0$, then we have a standard Bayesian persuasion problem, and we know from prior work that the sender-optimal distribution has support $\{0, \lambda\}$. But if $k > 0$, this is no longer optimal, because the garbling chosen by the receiver in response to that would be δ_μ, and the sender would not be accepted. This is easy to see intuitively–at a belief λ, the receiver is indifferent between accepting and rejecting. When offered $\{0, \lambda\}$, her gross payoff from choosing any garbling is the same as the payoff from rejecting with certainty. But then there is no reason for her to pay to learn anything. To make it worth her while to do so, the sender would have to allow her to generate beliefs above λ.

The following proposition, adapted to our notation from Wei (2020), summarizes the results for this benchmark case.

Proposition 3.1. *Suppose there is a single sender, and the receiver has a threshold of acceptance $\lambda > \mu$. Then,*

1. *A sender-preferred equilibrium exists.*
2. *In a sender-preferred equilibrium, the receiver's garbling on path is strictly less informative (in the Blackwell sense) than her optimal garbling in response to full information.*

Proof. See Lemma 1 and Proposition 2 in Wei (2020). ■

Importantly, as will be shown ahead in this paper, the receiver has a unique optimal garbling in response to any binary distribution, which implies that this result holds for *any* equilibrium and we can omit the "sender-preferred" qualifier.

Corollary 1. *If there is a single sender, and the receiver has a threshold of acceptance $\lambda \in (0, 1)$, then full information is not offered by the sender in equilibrium.*

In response to full information, say the receiver's optimal garbling has support $\{\nu_1, \nu_2\}$ where $\nu_1 < \lambda < \nu_2$. Then these results state that in equilibrium, the distribution the receiver ends up with would be a strict garbling of this. Stated differently, in equilibrium the sender does restrict the receiver's learning, and *does not allow her to choose her first-best level of learning.*

The intuition roughly is that although the sender cannot implement his optimal no-garbling solution $\{0, \lambda\}$, he need not go all the way to providing full information. He can profitably restrict learning so that the higher belief in the support of the receiver's garbling is below ν_2, and the probability of its realization is higher.

As we see next, introducing an additional sender yields an interesting comparison to this.

4 Equilibrium Analysis with Two Senders

We now analyze the game described in Section 2, for an arbitrary $k > 0$ and $\mu \in (0, 1)$.

To start off, recall our observation that the receiver's order of visits matters when the two distributions on offer are different. In equilibrium she must correctly anticipate the distributions chosen, and the order of visits must be a best response to those. However, since she does not observe the chosen distributions at stage 0, any deviation by a sender goes undetected until and unless he is visited. This has the following implication, which we note for further reference. *Any deviation by a sender cannot affect either the receiver's decision to visit a sender, or the order of her visits.*

Next, note that if both senders offer the same distribution, then the receiver is indifferent between the two orders of visit (if she visits anyone). The analysis below will make it clear that the tie breaking rule in this case does not matter for our results, and we do not assume anything about it.

We now turn to the question of equilibrium existence. Suppose that each of the two senders offers no information, i.e. the distribution δ_μ. Then upon visiting either sender, the receiver is also restricted to choosing δ_μ. But then she expects to gain nothing by visiting a sender, and not visiting either of them is a best response. She may simply select sender 1 with any probability $p \in [0, 1]$, and sender 2 with probability $1 - p$. Clearly, if this best response is played, a deviation by a sender goes undetected, and does not make any difference to the outcome. Thus we have the following.

Proposition 4.1 (Equilibrium existence). *An equilibrium exists for all $\mu \in (0, 1)$, and for all $k > 0$. In particular, there is always an equilibrium in which each sender offers an uninformative distribution.*

Naturally, we are interested in finding other, more interesting equilibria. Of particular interest are equilibria that give the receiver her *first-best* payoff. Let us clarify what exactly we mean by this.

The receiver's first-best payoff is essentially the best she can achieve, across all profiles of sender behavior (but still subject to attention costs). Equivalently, it is the payoff she would get if she herself could choose the senders' distributions. Now, since every distribution with expectation equal to the prior is a garbling of the fully informative distribution, she has greatest latitude when both senders offer the fully informative distribution. Thus, her first-best payoff is attained when both senders offer full information.

However, she may attain the same payoff even when senders choose other less informative distributions. Suppose, for illustration, that when offered full information, the following is a best response for the receiver: visit Sender 1 first, choose the garbling $\{\mu - \epsilon, \mu + \epsilon\}$ for him; then visit Sender 2 and choose the uninformative garbling for him. Then even if, e.g., Sender 1 offers the distribution $\{\mu - 2\epsilon, \mu + 2\epsilon\}$ and Sender 2 offers no information, she gets to secure her first-best payoff. At the heart of this is the fact that due to attention costs, she might not really use full information even when allowed to.

The next observation is easy to make.

Remark. Suppose there is an equilibrium in which Sender i offers p_i. Then the receiver achieves her first-best payoff in this equilibrium if and only if her best response on path is also a best response on path to full information from both senders.

This observation is key to the *only if* direction of the following proposition. The *if* direction is obvious.

Proposition 4.2 (First-best). *For given k and μ, there is an equilibrium that gives the receiver her first-best payoff if and only if there is an equilibrium in which both senders offer full information.*

An implication of this is that in establishing conditions for the existence of a full information equilibrium, we establish conditions for an equilibrium where the receiver gets her first-best. Hence, we focus our attention on full information, and the next proposition presents our main result.

Proposition 4.3 (Full information equilibrium). *There is an equilibrium in which both senders offer full information if and only if $k > \frac{1}{2}$ and $\mu \in \left[\frac{1}{4k}, 1 - \frac{1}{4k}\right]$.*

It is worth highlighting that the parameter in the attention cost function is crucial. If k is above $\frac{1}{2}$, we obtain an interval of priors over which full information is an equilibrium, and this interval expands as k grows. In the limit, as $k \to \infty$, the interval converges to $(0, 1)$, the full range of priors. Thus, by having higher attention costs, the receiver might elicit better information from competing senders.

The following corollary states the same result differently.

Corollary 2. *For all $\mu \in (0, 1)$, there is an equilibrium in which both senders offer full information if and only if $k > \max\left\{\frac{1}{4\mu}, \frac{1}{4(1-\mu)}\right\}$ (weak inequality if $\mu \neq \frac{1}{2}$).*

Stated this way, one might conjecture that the result is trivially obtained because for high enough values of k, the receiver finds it optimal to not learn anything at all even when offered full information. As it turns out, this is not the case, and for any finite k she does undertake some learning from at least one sender when offered full information.

Instead, we obtain the full information equilibrium because for high enough values of k, the receiver finds it optimal to learn only about the quality of one sender, and completely ignore information on the other. The analysis ahead will clarify how this fact plays a crucial role.

Section 4.2 provides a proof of Proposition 4.3 (and presents additional results), but before we move on to that, it is instructive to examine another benchmark, where $k = 0$.

4.1 Benchmark: No Attention Costs ($k = 0$)

When $k = 0$, it is costless for the receiver to learn. This makes a substantive difference to the analysis, because she never has a strict incentive to garble either sender's distribution–there is no reason to leave any information on the table.[15] Notice that in this case the receiver's first-best is achieved only when she learns at least one sender's type fully.

For simplicity, here we assume the following tie-breaking rules i) if the stage 1 draw is 0 (or 1), she rejects (or accepts) that sender without visiting the other one, and ii) if the draws from both stages are the same, she selects the sender visited last.

Proposition 4.4 (No attention costs)**.** *Suppose $k = 0$. Then the following are true.*

1. *For all $\mu \in (0, 1)$, there is an equilibrium in which both senders choose an uninformative distribution.*

2. *For all $\mu \in (0, 1)$, there is no equilibrium in which both senders offer full information.*

The reason an uninformative equilibrium exists is identical to that for $k > 0$–it is a best response for the receiver to not visit either sender, but then a deviation is not detected and makes no difference to the outcome. The reasoning behind non-existence of a full information equilibrium, on the other hand, is very different for $k = 0$ and for small, positive k.

For $k = 0$, in response to full information from both senders, the receiver visits either one of them, learns his type perfectly, and immediately takes a decision. A sender's deviation cannot make a difference if he is not visited first. But if he is, a deviation to support $\{\epsilon, 1\}$ is profitable, where ϵ is arbitrarily close to zero. This is because if the receiver's draw from this

[15] This setup has been studied in Boleslavsky and Cotton (2018), Au and Whitmeyer (2021) and other papers, with the difference that they assume that the receiver observes the chosen distributions at stage 0.

distribution is ϵ, she continues to learn from the second sender, and rejects him if the draw then is 0.

On the other hand, if k is any positive quantity, then for a small enough stage 1 draw ϵ, the receiver is sure enough of the quality of the first sender that she does not find it worth her while to learn about the other one.

The following result establishes existence of other (less than fully) informative equilibria when attention costs are absent.

Claim 4.1. 1. *Let $k = 0$ and $\mu \leq \frac{1}{2}$. There is an equilibrium in which each player chooses the uniform distribution on $[0, 2\mu]$.*

2. *Let $k = 0$ and $\mu > \frac{1}{2}$. There is an equilibrium in which each sender chooses a CDF with a continuous portion $F(x) = \frac{x}{2\mu}$ on $[0, 2(1-\mu)]$ and a point mass of size $2 - \frac{1}{\mu}$ on 1. In such an equilibrium, the receiver's decision about whom to visit first must be fair (each sender is visited first with probability $\frac{1}{2}$).*

4.2 Positive Attention Costs ($k > 0$)

As previously discussed, for positive attention costs our main result pertains to the full information equilibrium, which is stated in Proposition 4.3 above. We begin by showing why it is true, and present the key arguments for $k = 1$. The structure of the proof is the same for a generic $k > 0$, and the details are relegated to the Appendix.[16] In essence, the argument will be that full information is an equilibrium when the receiver wishes to learn only from one sender.

4.2.1 Overview of proof of main result for $k = 1$

Recall that for $k = 1$, our main result (Proposition 4.3) states that full information is an equilibrium if and only if $\mu \in \left[\frac{1}{4}, \frac{3}{4}\right]$.

Start by considering any $\mu \in (0, 1)$, and suppose that each sender offers the fully informative distribution with support $\{0, 1\}$. To analyze the receiver's best response, we proceed in two steps–first, we determine her stage 2 best response for each belief drawn at stage 1; second, we use that to solve for the optimal stage 1 behavior. We make use of the technique of concavification for this.

[16]Focusing here on $k = 1$ eases exposition by allowing us to avoid several mechanical sub-cases that add little insight.

The receiver's stage 2 best response: First let us find the optimal stage 2 garbling, if the receiver visits the sender at that stage.

Say the draw from stage 1 is $x \in [0,1]$. Then, the receiver selects the second sender if and only if the stage 2 draw y turns out to be higher than x.[17] Her payoff from a stage 2 belief y is then $max\{x,y\}$, minus the attention cost associated with y. Now, since any distribution (with expectation equal to the prior) is a garbling of the fully informative one, her stage 2 optimization problem is simply given by

$$\max_{q \in \Delta[0,1]} \int_{[0,1]} max\{x,y\} - (y-\mu)^2 dq(y) \;\; s.t. \int_{[0,1]} y \; dq(y) = \mu.$$

Let $U_2(y;x) \coloneqq max\{x,y\} - (y-\mu)^2$ for $x, y \in [0,1]$. This is piecewise concave in y, and is plotted for a representative value of x in Figure 2. We know from Kamenica and Gentzkow (2011) that for any given x, the receiver's optimal q is determined using the concavification of $U_2(y;x)$ over $[0,1]$.[18] The concavification is the red curve in Figure 2. It is evident that depending on where μ lies, the optimal distribution of beliefs is either degenerate on μ, or is binary.

Lemma 4.1 (Stage 2 optimal garbling)**.** *Suppose that the receiver's stage 1 draw is $x \in [0,1]$ and she visits the sender at stage 2. The receiver's stage 2 optimal garbling is either degenerate or binary, and its support is as follows.*

1. *If $\mu < \frac{1}{2}$,*
$$\begin{cases} \left\{x-\frac{1}{4}, x+\frac{1}{4}\right\} & if \; \frac{1}{4} \le x < \mu + \frac{1}{4} \\ \{0, \sqrt{x}\} & if \; \mu^2 < x < \frac{1}{4} \\ \{\mu\} & if x \le \mu^2 \; or \; x \ge \mu + \frac{1}{4} \end{cases}$$

2. *If $\mu = \frac{1}{2}$,*
$$\begin{cases} \left\{x-\frac{1}{4}, x+\frac{1}{4}\right\} & if \; \frac{1}{4} < x < \frac{3}{4} \\ \{\mu\} & if \; x \le \frac{1}{4} \; or \; x \ge \frac{3}{4} \end{cases}$$

[17] It does not matter what we assume about the tie breaking rule when $y = x$. For any distribution offered, x would not belong to the support of the garbling chosen at stage 2. The reasoning is similar to the argument for why the standard Bayesian persuasion solution is not incentive compatible in the single-sender case (see Section 3).

[18] Given x, the concavification of $U_2(y;x)$ is the smallest concave function that lies weakly above $U_2(y;x)$ for all $y \in [0,1]$.

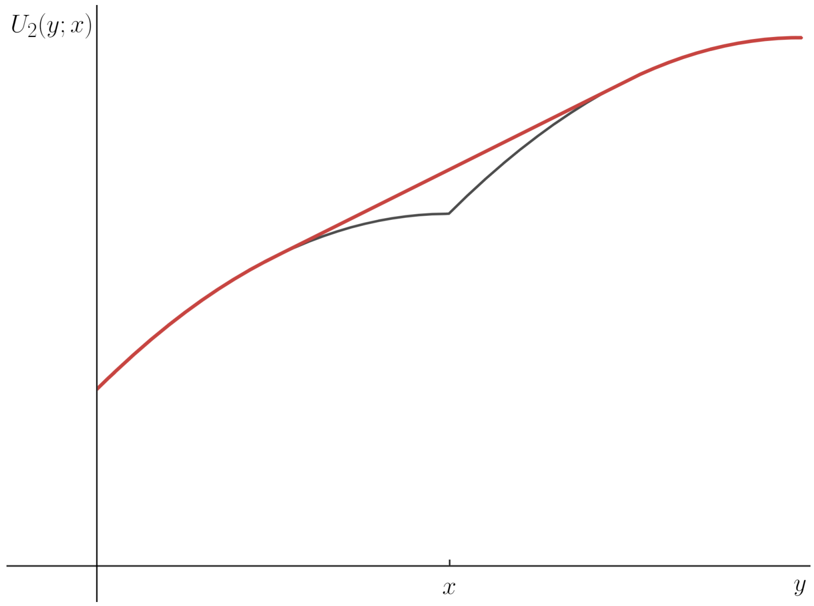

Figure 2: The receiver's stage 2 payoff function and its concavification (red).

3. If $\mu > \frac{1}{2}$,

$$\begin{cases} \left\{x - \frac{1}{4}, x + \frac{1}{4}\right\} & if\ \mu - \frac{1}{4} < x \leq \frac{3}{4} \\ \left\{1 - \sqrt{1-x}, 1\right\} & if\ \frac{3}{4} < x < 1 - (1-\mu)^2 \\ \{\mu\} & x \leq \mu - \frac{1}{4}\ or\ x \geq 1 - (1-\mu)^2 \end{cases}$$

The interesting thing to note here is that regardless of the prior, if the first stage draw is either very high or very low, then the receiver chooses not to learn anything from the second sender. This is intuitive–for a high enough belief that the first sender's quality is good, she deems it very unlikely that the second sender is better, and does not invest in learning about him. Instead, she simply accepts the first sender. Conversely, if the first stage draw is very low, she simply accepts the second sender.

Furthermore, the thresholds beyond which there is no learning at stage 2 depend on the prior. The prior is the expected quality of the second sender, so the higher it is, the larger (smaller) the range of stage 1 beliefs over which the second sender is accepted (rejected) without stage 2 learning.

For intermediate values of the stage 1 draw, the receiver does learn and chooses a binary distribution, selecting the second (first) sender at the higher (lower) realization. Since she has only two actions to choose from at this point, she would never choose a distribution with support on more than 2 beliefs: if she did, then she would be selecting the same sender at more than one belief, and could reduce her attention cost by collapsing those two into one.

For any stage 1 draw, if the stage 2 optimal garbling involves any learning, the receiver strictly gains from visiting the second sender. If it does not involve any learning, the receiver is indifferent between making the second visit and not, and she may resolve this in any manner.

The receiver's stage 1 best response: Using the above result, it is straightforward to obtain the receiver's first stage continuation payoffs for an arbitrary x, and determine her first stage optimal garbling from its concavification over $[0, 1]$. This leads to the following.

Lemma 4.2 (Stage 1 optimal garbling)**.** *Any distribution with expectation μ and support drawn from the following sets is optimal for the receiver at stage 1.*

1. $\left\{\mu - \frac{1}{4}\right\} \cup \left[\frac{1}{4}, \mu + \frac{1}{4}\right]$ *if* $\mu \in \left[\frac{1}{4}, \frac{1}{2}\right]$.
2. $\left[\mu - \frac{1}{4}, \frac{3}{4}\right] \cup \left\{\mu + \frac{1}{4}\right\}$ *if* $\mu \in \left[\frac{1}{2}, \frac{3}{4}\right]$.
3. $\{0, y_1(\mu)\}$ *if* $\mu < \frac{1}{4}$, *where* $y_1(\mu) \in \left(\mu, \frac{1}{4}\right)$.
4. $\{y_2(\mu), 1\}$ *if* $\mu > \frac{3}{4}$, *where* $y_2(\mu) \in \left(\frac{3}{4}, \mu\right)$.

The exact expressions for $y_1(\mu)$ and $y_2(\mu)$ are not important. The main thing to note here is that the stage 1 solution always involves some learning, and is unique if and only if $\mu \notin \left[\frac{1}{4}, \frac{3}{4}\right]$. Notably, in spite of the fact that there are only two senders and binary types in this model, the receiver may choose a distribution with support on more than two beliefs at stage 1. The reason is that each stage 1 belief is optimally followed by a different degree of learning at stage 2.[19]

The multiplicity of best responses for $\mu \in \left[\frac{1}{4}, \frac{3}{4}\right]$ captures the notion of substitutability between information sources: some of these responses involve learning more about the second sender, while others involve learning more about the first one, and the receiver is indifferent across these alternatives.

Note also that since the stage 1 optimal distribution always involves learning, a visit is necessarily made at this stage. The receiver does not care which sender is visited first, and she may randomize her choice in any way.

Full information equilibrium for $\boldsymbol{\mu} \in \left[\frac{1}{4}, \frac{3}{4}\right]$: When $\mu \in \left[\frac{1}{4}, \frac{3}{4}\right]$, we need to make a selection among the receiver's stage 1 optimal responses for the purpose of proving the

[19]This observation is familiar from models of dynamic rational inattention (e.g. Hébert and Woodford (2019), Zhong (2019)) where continuation payoffs depend on posterior beliefs.

existence of our equilibrium. Notice that the *most informative* (in the Blackwell sense) of the optimal distributions has support $\left\{\mu - \frac{1}{4}, \mu + \frac{1}{4}\right\}$, and by Lemma 4.1, this is the only one among them that is necessarily followed by no learning at stage 2. We assume (in the construction of our equilibrium) that the receiver breaks her indifference in favor of this distribution. That is, when indifferent, she'd rather not put off learning until the next stage.

To summarize: if $\mu \in \left[\frac{1}{4}, \frac{3}{4}\right]$, the receiver's on-path best response to full information is the following. Visit sender 1 with probability $q \in [0, 1]$, and Sender 2 with probability $1 - q$. Choose the garbling with support $\left\{\mu - \frac{1}{4}, \mu + \frac{1}{4}\right\}$ for the sender visited. If the belief drawn is $\mu - \frac{1}{4}$, select the other sender without learning anything from him.[20] If the belief drawn is $\mu + \frac{1}{4}$, select the visited sender without learning anything from the other one.[21]

Having specified on-path behavior, we examine what happens in the event of a unilateral sender deviation. We have already seen that by deviating, a sender does not affect the probability of being the one to be visited first. Moreover, if he is *not* the one to be visited first, his payoffs are not affected, since the receiver does not plan to learn anything from him. So, we only need to consider what happens if he deviates and is visited first.

In this case, the receiver's behavior would be altered if $\left\{\mu - \frac{1}{4}, \mu + \frac{1}{4}\right\}$ is *not* a garbling of the distribution he deviates to. We specify a sequentially rational protocol for the receiver in this event, and show that the sender's deviation is unprofitable.

Note that regardless of what the sender's deviation is, the receiver is permitted to learn nothing, i.e. choose support $\{\mu\}$. If she does so, then by Lemma 4.1, her optimal continuation behavior would be to visit the other sender (who has not deviated and continues to offer full information) and choose support $\left\{\mu - \frac{1}{4}, \mu + \frac{1}{4}\right\}$ for him.

The key thing to note is that by Lemma 4.2, this protocol–not learning at the first visit and learning according to $\left\{\mu - \frac{1}{4}, \mu + \frac{1}{4}\right\}$ at the second visit–is also one of the receiver's best responses to full information, which allows us to conclude that this is sequentially rational once she learns of the sender's deviation. We specify this protocol as the receiver's off path behavior.

The next Lemma implies that for the on- and off-path behavior specified for the receiver, the deviating sender does not gain (by deviating).

Lemma 4.3. *For all $\mu \in \left[\frac{1}{4}, \frac{3}{4}\right]$, conditional on being visited first, a sender's expected payoff is the same for any of the receiver responses specified in Lemma 4.2.*

Thus, providing full information is indeed an equilibrium.

[20] Either by not visiting him at all, or by visiting but not learning.

[21] ibid.

Let's take a closer look at the intuition behind this. We see that in the best case scenario for the receiver, i.e. when both senders allow her perfect information, attention costs lead her to learn from only one sender. It does not matter which sender that is–if it is a best response to learn only from the sender visited first, then clearly it is also a best response to learn only from the one visited second.

Now, we say that on path she chooses to learn from the first sender she visits. If, however, that sender deviates and restricts her learning, she is able to compensate for it by learning more from the other sender. The *ex ante* probability that she makes the correct choice thereby remains unaffected, and the deviating sender is unable to gain.

Comparison with single sender benchmark. Suppose that only one sender (say sender 1) provided information in this game. Then the receiver would be restricted to learning nothing about sender 2, and in her first-best she would choose the distribution $\{\mu - \frac{1}{4}, \mu + \frac{1}{4}\}$ for sender 1. Now, what happens in the game with a strategic sender 1? If ties are broken in favor of sender 1, then clearly he would offer no information in equilibrium. If ties are broken in favor of sender 2, then from Proposition 3.1 we know that sender 1 would, in equilibrium, offer a strict garbling of the receiver's first-best distribution. As usual this extends to other values of k, and we have the following.

Proposition 4.5 (Welfare gains from competition)**.** *When a full information equilibrium exists, the receiver's payoff from it is strictly higher than in a game where there is only one sender.*

Non existence of full information equilibrium for $\boldsymbol{\mu} \notin \left[\frac{1}{4}, \frac{3}{4}\right]$: When $\mu \notin \left[\frac{1}{4}, \frac{3}{4}\right]$, analogous reasoning does not apply, since the receiver's best response is unique and involves learning from both senders on path. In this case, there exists a deviation where a sender profitably restricts the receiver's learning in case he is visited second, without affecting what happens if he is visited first.

In particular, say for instance $\mu < \frac{1}{4}$. Recall that in response to full information, the receiver chooses support $\{0, y_1(\mu)\}$ at stage 1. Following belief 0 she immediately accepts the second sender, and following belief $y_1(\mu)$, she chooses support $\left\{0, \sqrt{y_1(\mu)}\right\}$ at stage 2.

It can be shown that there exists $p_2 \in (y_1(\mu), \sqrt{y_1(\mu)})$ such that if a sender deviates to $\{0, p_2\}$ and is visited second (by the receiver holding a belief $y_1(\mu)$ from stage 1), the receiver chooses $\{0, p_2\}$ instead of $\left\{0, \sqrt{y_1(\mu)}\right\}$. If he is instead visited first, the receiver's best response is unchanged. Evidently this deviation increases the probability of being selected, and is therefore profitable.

4.2.2 Other equilibria for $k > 0$

The analysis so far tells us that for any k, first, an uninformative equilibrium always exists; and second, a full information equilibrium exists for a large class of parameter values.

Our focus on full information is not misplaced, in spite of the fact that the receiver never makes use of it even when it is on offer.[22] The reason we focus on it–as we have discussed at length–is Proposition 4.2: full information is an equilibrium if and only if the receiver can get her first-best outcome in an equilibrium.

Our interest in equilibria where the receiver gets her first-best is natural–first, the existence of such equilibria is surprising; second, in many situations it is appropriate to select receiver-preferred equilibria.

This brings us to two questions. First, what can we say about other equilibria (besides full information) that give the receiver her first-best payoff? And second, what can we say about equilibria that do not give her the first-best?

Answering the first question is easy. There is a whole class of such equilibria, where the distributions offered by the senders allow her to respond exactly as under full information.

Claim 4.2 (Equilibria outcome equivalent to full information)**.** *Suppose $k > \frac{1}{2}$ and $\mu \in \left[\frac{1}{4k}, 1 - \frac{1}{4k}\right]$. For $i \in \{0, 1\}$, let $p_i \in \Delta[0, 1]$ be any distribution with expectation μ, and of which the distribution with support $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$ is a garbling. Then, there is an equilibrium in which sender i offers the distribution p_i. Such an equilibrium is outcome equivalent to full information.*

This can be used to construct specific examples such as the following.

Corollary 3.

1. *Let $\mu \le \frac{1}{2}$. Then there is an equilibrium in which both senders offer the uniform distribution on $[0, 2\mu]$ if $k \ge \frac{1}{2\mu}$.*
2. *Let $\mu > \frac{1}{2}$. Then there is an equilibrium in which both senders offer a CDF with a continuous portion $F(x) = \frac{x}{2\mu}$ on $[0, 2(1 - \mu)]$ and a point mass of size $2 - \frac{1}{\mu}$ on 1 if $k \ge \frac{1}{2\mu}$ for $\mu \le \frac{2}{3}$, and if $k \ge \frac{1}{4(1-\mu)}$ for $\mu \ge \frac{2}{3}$.*

We choose to present these examples to highlight a key distinction: Recall from Claim 4.1 that these distributions are also equilibria in the $k = 0$ scenario, where full information is *not* an equilibrium. In contrast, here these equilibria are in fact outcome equivalent to full

[22] We showed this for $k = 1$, but Appendix A.2 shows that it is true generally: in response to full information, she visits only one sender and picks support $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$.

information. The difference arises since with attention costs, both full information and these distributions are garbled to the same distribution the receiver.

Turning to the question about equilibria that do not give the receiver her first-best, we have the following sharp result that rules out their existence in the class of *binary, symmetric* equilibria.[23] Essentially, it implies that if a symmetric binary equilibrium exists (for any parameters), so must the full information equilibrium, and in fact it must be outcome equivalent to the full information one–so that the receiver must be getting her first-best.

Proposition 4.6 (Binary symmetric equilibria)**.** *Let the distribution p have support $\{l, h\}$ with $l \in [0, \mu)$ and $h \in (\mu, 1]$. There is an equilibrium where both senders offer p if and only if $k > \frac{1}{2(h-l)}$ and $\mu \in \left[l + \frac{1}{4k}, h - \frac{1}{4k}\right]$.*

The proof uses arguments similar to those for the full information equilibrium, and for $h = 1$, $l = 0$ this proposition is identical to Proposition 4.3.

Beyond this, characterizing *all* equilibria of the game is beyond the scope of this paper, a major reason being that little of practical use can be said about the set of garblings of a non-binary distribution.

5 Extensions

Here, we illustrate that our main results continue to hold under a variety of different modelling choices. We begin by demonstrating that they continue to hold if instead the senders' experiment choices are public.

5.1 Publicly Chosen Experiments

So far we have considered a scenario where the receiver does not observe the experiment chosen by a sender until she visits him. This is a natural assumption in the context of our leading example–it would not be known to a doctor how detailed the research published on a drug is until she takes a look through it. As another example, one does not know how many customer reviews a seller has allowed to be posted on his website until one visits the website.

However, there can be other applications where one might expect the senders' experiments to be observed when they are chosen, i.e. at stage 0 of our game. In this case, the strategic considerations remain the same except for an important difference–posted experiments allow

[23] That is, equilibria where both senders offer the same distribution, that has binary support.

a sender's deviation to be observed by the receiver, and hence deviations affect her order of visits. For example, if both senders are expected to choose the same experiment, then the receiver is indifferent between the order of visits; by deviating, a sender can break this indifference. We show that our main results continue to hold in this scenario.

Proposition 5.1. *Suppose that the experiments chosen by the senders are observed by the receiver at stage 0 of the game. Then Proposition 4.6 still holds.*

A full proof of Proposition 5.1 follows in the Appendix, but by recalling the arguments that lead up to existence of a full information equilibrium for some parameters in the baseline model, it is easy to see why that result is robust: If the receiver's best response to full information in the baseline model is to learn from exactly one sender, then a deviation cannot affect payoffs even if it is observed at stage 0. Put succinctly, if the receiver knows that a sender has deviated, she can simply visit the *other* sender and learn from him.

Given that Proposition 4.6 holds, it is unsurprising that Proposition 5.1 does as well. As noted in Au and Whitmeyer (2021), posted information policies, like posted prices, encourages competition, which may be to the searcher's benefit. We find in this paper that–unlike Au and Whitmeyer (2021) or the frictionless models of Boleslavsky and Cotton (2015), Albrecht (2017), and Au and Kawai (2020)–full information is provided in an equilibrium, even when information policies are not posted.

Considering this extensive form allows us to gain additional insight into how attention costs matter for welfare. It turns out that when $k = 0$, the *unique* equilibrium in this game is given by Claim 4.1 (for a proof, see, e.g., Boleslavsky and Cotton (2015)). Neither sender offers a fully informative experiment in this equilibrium, but since attention is costless, the receiver processes all information on offer.

In contrast, we have seen that in the presence of attention costs, there can be equilibria where full information is provided but is garbled by the receiver. The following result establishes conditions under which the amount of information the receiver ultimately processes is higher when attention is costly than when it is costless. In particular, it highlights that attention costs may end up enabling the receiver to make a better choice between the senders on average.

Proposition 5.2 (Welfare effects of attention costs)**.** *Given $\mu \in (0, 1)$, suppose that $k > \max\left\{\frac{1}{4\mu}, \frac{1}{4(1-\mu)}\right\}$, so that there is an equilibrium in which senders provide full information. Then there exists $\overline{k}(\mu) > \max\left\{\frac{1}{4\mu}, \frac{1}{4(1-\mu)}\right\}$ such that the expected type of the sender eventually* *selected by the receiver (i.e., her gross payoff) in this equilibrium is strictly higher than in the unique equilibrium under $k = 0$ if and only if $k < \overline{k}(\mu)$.*

Proof. The expected type of the selected sender in the full information equilibrium is $\frac{1}{2}\mu + \frac{1}{2}(\mu + \frac{1}{4k})$. When $k = 0$, for $\mu \leq \frac{1}{2}$ this is $\frac{4}{3}\mu$, and for $\mu > \frac{1}{2}$ this is $\frac{4}{3}(1-\mu)(\frac{1}{\mu}-1)^2+[1-(\frac{1}{\mu}-1)^2]$. Straightforward algebra then yields the result. ■

To understand why the inequality reverses when $k > \overline{k}(\mu)$, note that even as we hold the full information equilibrium constant, the amount the receiver chooses to learn in this equilibrium diminishes as k grows.

5.2 Different Means

Our baseline model assumes that the distributions of the senders' types have identical means. There, the receiver's problem is the most interesting, since *ex ante* she has very little information to base her choice on.

In this section, we show that our main result applies to settings where the prior beliefs on the two senders are different but sufficiently close. Namely, if the two means μ_1 and μ_2 lie in a particular interval, then there is an equilibrium in which both senders offer full information.

Once again for expositional convenience, let $k = 1$. We proceed in the same vein as in the proof for Lemma 4.2: using Lemma 4.1 we can obtain the receiver's first stage continuation payoffs for an arbitrary first stage draw x, which we then concavify to obtain the receiver's first stage optimal garbling. Suppose that sender 2 is visited second (we'll revisit this assumption shortly).

Lemma 5.1 (Stage 1 optimal garbling)**.** *Any distribution with expectation μ and support drawn from the following sets is optimal for the receiver at stage 1.*

1. $\left[\mu_2 - \frac{1}{4}, \frac{3}{4}\right] \cup \left\{\mu_2 + \frac{1}{4}\right\}$ *if* $\mu_2 \in \left[\frac{1}{2}, \frac{3}{4}\right]$ *and* $\mu_1 \in \left[\mu_2 - \frac{1}{4}, \mu_2 + \frac{1}{4}\right]$.
2. $\left[\mu_2 - \frac{1}{4}, \frac{3}{4}\right]$ *if* $\mu_2 \in \left[\frac{3}{4}, 1\right]$ *and* $\mu_1 \in \left[\mu_2 - \frac{1}{4}, \frac{3}{4}\right]$.
3. $\left[\frac{1}{4}, \mu_2 + \frac{1}{4}\right] \cup \left\{\mu_2 - \frac{1}{4}\right\}$ *if* $\mu_2 \in \left[\frac{1}{4}, \frac{1}{2}\right]$ *and* $\mu_1 \in \left[\mu_2 - \frac{1}{4}, \mu_2 + \frac{1}{4}\right]$.
4. $\left[\frac{1}{4}, \mu_2 + \frac{1}{4}\right]$ *if* $\mu_2 \in \left[0, \frac{1}{4}\right]$ *and* $\mu_1 \in \left[\frac{1}{4}, \mu_2 + \frac{1}{4}\right]$.

Note that we have not yet determined which sender should be visited first. Our first step is to show that if μ_1 and μ_2 satisfy one of the conditions for Lemma 5.1 when μ_2 is visited second, then they satisfy one of the conditions for Lemma 5.1 when μ_2 is visited first. Formally,

Lemma 5.2. *One of the four parametric restrictions in Lemma 5.1 holds if and only if one of the four parametric restrictions in Lemma 5.1 holds in which μ_1 and μ_2 are replaced with each other.*

Assuming that one of the parametric conditions hold, the second step is to show that the receiver's expected payoff under any optimal protocol in which we assume that sender 1 is visited first is the same as her expected payoff under any optimal protocol in which we assume sender 2 is visited first. Hence, it does not matter which sender she visits first, and so she can break ties in that manner in any way that she chooses.

This step requires just a couple sentences to prove: in each of the four cases described in Lemma 5.1, there is a stage 1 optimal distribution in which the receiver learns nothing at the first sender. Her expected payoff under the optimal search protocol is thus

$$\mu_1^2 + \mu_2^2 + \frac{\mu_1 + \mu_2}{2} - 2\mu_1\mu_2 + \frac{1}{16}$$

which is invariant to an exchange of μ_1 and μ_2. Finally, we arrive at the heterogeneous means analog to Proposition 4.3:

Proposition 5.3. *There is a full information equilibrium if $|\mu_2 - \mu_1| \leq \frac{1}{4}$ and*

1. $\mu_1, \mu_2 \in \left[\frac{1}{4}, \frac{3}{4}\right]$*; or*
2. $\mu_i \leq \frac{3}{4} \leq \mu_j$ *for* $i, j \in \{1, 2\}$ *and* $i \neq j$*; or*
3. $\mu_i \leq \frac{1}{4} \leq \mu_j$ *for* $i, j \in \{1, 2\}$ *and* $i \neq j$.

5.3 Sender Experiment-Dependent Cost Function

In our baseline model, attention costs depend only on the information structure the receiver takes a draw from, and not *directly* on a sender's experiment.

Here, we incorporate the possibility that the more informative the sender's experiment, the less costly a given information structure is for the receiver. This corresponds to the following intuition: the less informative the seller is, the costlier it is for the buyer to maintain a particular information structure, since she is forced to pay closer attention.

We do so by allowing the receiver's information processing cost at a sender to itself depend on the distribution chosen by the sender. That is, we amend the attention cost so that it is now given by (at sender 1)

$$C(q_1, p_1) = k(p_1) \int_{[0,1]} (x-\mu)^2 dq_1(x), \tag{2}$$

where p_1 is the choice of distribution of posterior beliefs by sender 1 and q_1 is the garbling of that distribution chosen by the receiver. To capture the intended intuition, we assume that k is weakly decreasing in the Blackwell order.

We define $k_F := k(p_1^B)$ to be the minimum cost parameter, where p_1^B is the Bernoulli distribution begotten by full information provision by the sender. Naturally, we stipulate that k_F is non-negative.

With this modified cost function, our main result continues to hold. Namely, provided the attention cost is sufficiently high, there is an equilibrium in which both senders offer full information:

Proposition 5.4 (Full information equilibrium)**.** *For all $k_F > \frac{1}{2}$, if $\mu \in \left[\frac{1}{4k_F}, 1 - \frac{1}{4k_F}\right]$ then there* is *an equilibrium in which both sellers offer full information.*

Proof. On path, where each sender provides full information, the analysis is unchanged from earlier sections (with k_F *in lieu* of k), and the receiver's optimal protocol is unaltered. Moreover, should a sender deviate, then again the receiver can behave optimally by learning nothing at the deviating sender, eliminating the possibility for a sender to deviate profitably. ■

6 Conclusion

We study a model of information disclosure by two senders who compete to be selected by a receiver. The receiver, instead of passively accepting the information disclosure rule adopted by a sender, may choose to garble it before drawing a belief. The lower the informativeness of the chosen garbling, the lower her attention costs.

We show how for a large class of parameters, it is an equilibrium for the senders to offer *at least* as much information to the receiver as she would choose for herself, if she could control information provision. In particular, full disclosure by both senders is an equilibrium. Moreover, there is no binary symmetric equilibrium (for any value of parameters) that does not give the receiver this first-best outcome. We prove robustness to various modeling assumptions.

Our results stem from an interesting trade-off that generalizes beyond the specifics of our model. Due to attention costs, the receiver never finds it worthwhile to learn either sender's type perfectly. That is, even with access to full information, she leaves some scope for further learning about each. Moreover, since her task is to choose between the senders, information on the quality of one sender partially substitutes for information on the quality of the other. For example, learning a lot about the quality of one drug on the market can be just as good (for the accuracy of a doctor's decision) as learning a bit about both alternatives.

Consequently, starting from a situation of full disclosure by both senders, if either sender deviates and restricts the receiver's learning, she has an opportunity to make up for it by using some of the "surplus" information (so far unused) about the *other* sender. The deviating sender thus has limited ability–if any–to affect the overall quality of the receiver's information across the two alternatives.

This channel clearly breaks down in the absence of attention costs (so that the receiver always uses all available information), or if there is only one sender (so that there is no notion of substitutability). Thus our model, by being the first one to combine these elements, identifies novel strategic incentives for greater information disclosure.

Our leading example is that of pharmaceutical companies strategically disclosing information to prescribing physicians. The assumptions of high attention costs, low outside option for the receiver, and relative unimportance of price differences are reasonable in this context.

However, this model can be applied to numerous other settings. For instance, the receiver could be a buyer sequentially visiting two used car dealerships, taking test drives and gathering information made available on the features of each alternative. Our results suggest that a single dealership might not gain by being the only one to conceal information. Other settings include provision of information about insurance or pensions plans, and the design of political campaigns.

References

Albrecht, B.C.. 2017. "Political Persuasion." *Working paper.*

Alper, Brian S, Jason A Hand, Susan G Elliott, Scott Kinkade, Michael J Hauan, Daniel K Onion, and Bernard M Sklar. 2004. "How much effort is needed to keep up with the literature relevant for primary care?" *Journal of the Medical Library association*, 92(4): 429.

Armstrong, Mark. 2017. "Ordered Consumer Search." *Journal of the European Economic Association*, 15(5): 989–1024, 06.

Au, Pak Hung and Keiichi Kawai. 2020. "Competitive information disclosure by multiple senders." *Games and Economic Behavior*, 119: 56 – 78.

Au, Pak Hung and Mark Whitmeyer. 2021. "Attraction Versus Persuasion." *arXiv e-prints*: arXiv–1802.

Battaglini, Marco. 2002. "Multiple referrals and multidimensional cheap talk." *Econometrica*, 70(4): 1379–1401.

Blackwell, David. 1951. "Comparison of Experiments." in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*: 93–102, Berkeley, Calif.: University of California Press.

——— 1953. "Equivalent comparisons of experiments." *The annals of mathematical statistics*: 265–272.

Bloedel, Alexander W and Ilya Segal. 2020. "Persuading a Rationally Inattentive Agent." *Working paper.*

Board, Simon and Jay Lu. 2018. "Competitive information disclosure in search markets." *Journal of Political Economy*, 126(5): 1965–2010.

Boleslavsky, Raphael and Christopher Cotton. 2015. "Grading Standards and Education Quality." *American Economic Journal: Microeconomics*, 7(2): 248–79, 4.

——— 2018. "Limited capacity in project selection: Competition through evidence production." *Economic Theory*, 65(2): 385–421.

Burdett, Kenneth and Kenneth L Judd. 1983. "Equilibrium price dispersion." *Econometrica: Journal of the Econometric Society*: 955–969.

Caplin, Andrew, Mark Dean, and John Leahy. 2019. "Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy." *Working paper.*

Diamond, Peter A. 1971. "A model of price adjustment." *Journal of economic theory*, 3(2): 156–168.

Fudenberg, Drew and Jean Tirole. 1991. "Perfect Bayesian equilibrium and sequential equilibrium." *journal of Economic Theory*, 53(2): 236–260.

Gentzkow, Matthew and Emir Kamenica. 2014. "Costly persuasion." *American Economic Review*, 104(5): 457–62.

Goldacre, Ben. 2014. *Bad pharma: how drug companies mislead doctors and harm patients*: Macmillan.

Hébert, Benjamin M and Michael Woodford. 2019. "Rational Inattention when Decisions Take Time." *Working paper.*

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

Le Treust, Maël and Tristan Tomala. 2019. "Persuasion with limited communication capacity." *Journal of Economic Theory*, 184: 104940.

Lipnowski, Elliot, Laurent Mathevet, and Dong Wei. 2020a. "Attention management." *American Economic Review: Insights*, 2(1): 17–32.

——— 2020b. "Optimal Attention Management: A Tractable Framework." *arXiv preprint arXiv:2006.07729.*

Mackowiak, Bartosz Adam, Filip Matejka, Mirko Wiederholt et al.. 2020. "Rational Inattention: A Review." *CEPR Discussion Papers*(15408).

Matějka, Filip and Alisdair McKay. 2012. "Simple market equilibria with rationally inattentive consumers." *American Economic Review*, 102(3): 24–29.

——— 2015. "Rational inattention to discrete choices: A new foundation for the multinomial logit model." *American Economic Review*, 105(1): 272–98.

Matyskova, Ludmila. 2018. "Bayesian persuasion with costly information acquisition." *Working paper.*

Montes, Alphonso. 2020. "Bayesian Persuasion with Rational Inattentive Receiver." *Working paper.*

Sims, Christopher A. 2003. "Implications of rational inattention." *Journal of monetary Economics*, 50(3): 665–690.

Wei, Dong. 2020. "Persuasion under costly learning." *Journal of Mathematical Economics.*

Whitmeyer, Joseph and Mark Whitmeyer. 2020. "Mixtures of mean-preserving contractions." *Journal of Mathematical Economics*: 102450.

Zhong, Weijie. 2019. "Optimal dynamic information acquisition." *Working paper.*

A Proofs

In our proofs, we frequently invoke the receiver's optimal protocol (and senders' payoffs) when both senders offer the same binary distribution (which may be fully informative). Therefore, we begin by studying this.

A.1 The receiver's optimal protocol (and sender payoffs) when both senders offer the same binary distribution

Consider any $k > 0$ and $\mu \in (0, 1)$. Let each sender offer support $\{l, h\}$, with $l \in [0, \mu)$ and $h \in (\mu, 1]$.

Lemma A.1 (Optimal stage 2 garbling). *Suppose that the receiver's stage 1 draw is $x \in [l, h]$ and she visits the sender at stage 2. The receiver's stage 2 optimal garbling is either degenerate or binary, and its support is as follows.*

1. *If $k > \frac{1}{2(h-l)}$ and $\mu \le \min\left\{h - \frac{1}{2k}, l + \frac{1}{2k}\right\}$:*

$$\begin{cases} \{\mu\} & if\ x \in [l, l + k(\mu - l)^2] \\ \left\{l, l + \sqrt{\frac{x-l}{k}}\right\} & if\ x \in (l + k(\mu - l)^2, l + \frac{1}{4k}) \\ \left\{x - \frac{1}{4k}, x + \frac{1}{4k}\right\} & if\ x \in \left[l + \frac{1}{4k}, \mu + \frac{1}{4k}\right) \\ \{\mu\} & if\ x \in \left[\mu + \frac{1}{4k}, h\right] \end{cases}$$

2. *If $k > \frac{1}{2(h-l)}$ and $\mu \ge \max\left\{h - \frac{1}{2k}, l + \frac{1}{2k}\right\}$:*

$$\begin{cases} \{\mu\} & if\ x \in \left[l, \mu - \frac{1}{4k}\right] \\ \left\{x - \frac{1}{4k}, x + \frac{1}{4k}\right\} & if\ x \in \left(\mu - \frac{1}{4k}, h - \frac{1}{4k}\right] \\ \left\{h - \sqrt{\frac{h-x}{k}}, h\right\} & if\ x \in \left(h - \frac{1}{4k}, h - k(h - \mu)^2\right) \\ \{\mu\} & if x \in [h - k(h - \mu)^2, h] \end{cases}$$

3. If $l + \frac{1}{2k} \le \mu \le h - \frac{1}{2k}$:

$$\begin{cases} \left\{x - \frac{1}{4k}, x + \frac{1}{4k}\right\} & if\ x \in \left(\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right) \\ \{\mu\} & if\ x \in \left[l, \mu - \frac{1}{4k}\right] \cup \left[\mu + \frac{1}{4k}, h\right] \end{cases}$$

4. If $k > \frac{1}{2(h-l)}$ and $h - \frac{1}{2k} < \mu < l + \frac{1}{2k}$:

$$\begin{cases} \{\mu\} & if\ x \in [l, l + k(\mu - l)^2] \\ \left\{l, l + \sqrt{\frac{x-l}{k}}\right\} & if\ x \in (l + k(\mu - l)^2, l + \frac{1}{4k}) \\ \left\{x - \frac{1}{4k}, x + \frac{1}{4k}\right\} & if\ x \in \left[l + \frac{1}{4k}, h - \frac{1}{4k}\right] \\ \left\{h - \sqrt{\frac{h-x}{k}}, h\right\} & if\ x \in (h - \frac{1}{4k}, h - k(\mu - h)^2) \\ \{\mu\} & if\ x \in [h - k(\mu - h)^2, h] \end{cases}$$

5. If $k \le \frac{1}{2(h-l)}$:

$$\begin{cases} \{\mu\} & if\ x \in [l, l + k(\mu - l)^2] \\ \left\{l, l + \sqrt{\frac{x-l}{k}}\right\} & if\ x \in (l + k(\mu - l)^2, l + k(h - l)^2) \\ \{l, h\} & if\ x \in [l + k(h - l)^2, h - k(h - l)^2] \\ \left\{h - \sqrt{\frac{h-x}{k}}, h\right\} & if\ x \in (h - k(h - l)^2, h - k(\mu - h)^2) \\ \{\mu\} & if\ x \in [h - k(\mu - h)^2, h] \end{cases}$$

Proof. The receiver's stage 2 payoffs for a stage 2 belief y are given by

$$U_2(y; x) = \max\{x, y\} - k(y - \mu)^2.$$

This is piecewise concave. We first obtain the concavification of $U_2(y; x)$ over $[l, h]$ and then use it to find the optimal garbling.

The concavification of $U_2(y; x)$ is obtained by joining two points y_1, y_2 (in a straight line) with $l \le y_1 < x < y_2 \le h$. By the definition of concavification of a function, we must have[24]

$$U_2'(y_1; x) \le \frac{U_2(y_2; x) - U_2(y_1; x)}{y_2 - y_1} \le U_2'(y_2; x), \tag{3}$$

[24] The best way to see this is to assume it is does not hold and see that the definition of concavification is violated.

with the first inequality holding with equality if $y_1 > l$ and the second one holding with equality if $y_2 < h$.

The solution to Inequation 3 with both equalities is

$$y_1 = x - \frac{1}{4k},\ y_2 = x + \frac{1}{4k}.$$

If $l + \frac{1}{4k} < x < h - \frac{1}{4k}$, the concavification is given by $y_1 = x - \frac{1}{4k},\ y_2 = x + \frac{1}{4k}$.

If $x \leq \min\left\{l + \frac{1}{4k}, h - \frac{1}{4k}\right\}$, the lower bound l binds and the concavification has $y_1 = l$. $y_2 = l + \sqrt{\frac{x-l}{k}}$ is obtained from the second equality in Inequation 3.

If $x \geq \max\left\{h - \frac{1}{4k}, l + \frac{1}{4k}\right\}$, the upper bound h binds and the concavification has $y_2 = h$. $y_1 = h - \sqrt{\frac{h-x}{k}}$ is obtained from the first equality in Inequation 3.

If $h - \frac{1}{4k} < x < l + \frac{1}{4k}$, the concavification is:

1. $y_1 = l$, $y_2 = l + \sqrt{\frac{x-l}{k}}$ if $l + \sqrt{\frac{x-l}{k}} \leq h$.
2. $y_2 = h$, $y_1 = h - \sqrt{\frac{h-x}{k}}$ if $h - \sqrt{\frac{h-x}{k}} \geq l$.
3. $y_1 = l$, $y_2 = h$ otherwise.

Having obtained the concavification for any x, the optimal stage 2 garbling has support $\{y_1, y_2\}$ if $\mu \in (y_1, y_2)$, and support $\{\mu\}$ otherwise. Straightforward algebra then gives us the stated result. ■

Lemma A.2 (Optimal stage 1 garbling)**.**

1. *If $k > \frac{1}{2(h-l)}$ and $\mu \leq \min\left\{h - \frac{1}{2k}, l + \frac{1}{2k}\right\}$, the receiver's optimal stage 1 garbling is*
   - *(a) Any distribution with expectation μ and support drawn from the set $\left\{\mu - \frac{1}{4k}\right\} \cup \left[l + \frac{1}{4k}, \mu + \frac{1}{4k}\right]$ if $\mu \geq l + \frac{1}{4k}$.*
   - *(b) The distribution with support $\{l, y_1(\mu)\}$ with $y_1(\mu) \in \left(\mu, l + \frac{1}{4k}\right)$ if $\mu < l + \frac{1}{4k}$.*
2. *If $k > \frac{1}{2(h-l)}$ and $\mu \geq \max\left\{h - \frac{1}{2k}, l + \frac{1}{2k}\right\}$, the receiver's optimal stage 1 garbling is:*
   - *(a) Any distribution with expectation μ and support drawn from the set $\left[\mu - \frac{1}{4k}, h - \frac{1}{4k}\right] \cup \left\{\mu + \frac{1}{4k}\right\}$ if $\mu \leq h - \frac{1}{4k}$.*

(b) *The distribution with support* $\{y_2(\mu), h\}$ *with* $y_2(\mu) \in \left(h - \frac{1}{4k}, \mu\right)$ *if* $h - \frac{1}{4k} < \mu$.

3. *If* $l + \frac{1}{2k} \leq \mu \leq h - \frac{1}{2k}$*, the receiver's optimal stage 1 garbling is any distribution with expectation* μ *and support on* $\left[\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right]$.

4. *If* $k > \frac{1}{2(h-l)}$ *and* $h - \frac{1}{2k} < \mu < l + \frac{1}{2k}$*, the receiver's stage 1 optimal garbling is:*

   (a) *Any distribution with expectation* μ *and support drawn from* $\left\{\mu - \frac{1}{4k}\right\} \cup \left[l + \frac{1}{4k}, h - \frac{1}{4k}\right] \cup \left\{\mu + \frac{1}{4k}\right\}$ *if* $l + \frac{1}{4k} \leq \mu \leq h - \frac{1}{4k}$.

   (b) *The distribution with support* $\{l, y_1(\mu)\}$ *with* $y_1(\mu) \in \left(\mu, l + \frac{1}{4k}\right)$ *if* $\mu < l + \frac{1}{4k}$.

   (c) *The distribution with support* $\{y_2(\mu), h\}$ *with* $y_2(\mu) \in \left(h - \frac{1}{4k}, \mu\right)$ *if* $h - \frac{1}{4k} < \mu$.

5. *If* $k \leq \frac{1}{2(h-l)}$*, then*

   (a) *If* $\mu \leq \frac{l+h}{2}$*, the receiver's optimal stage 1 garbling is* $\{l, y_1(\mu)\}$*, where* $y_1(\mu) > \mu$ *is either on* $(l + k(\mu - l)^2, l + k(h - l)^2)$ *or on* $[l + k(h - l)^2, h - k(h - l)^2]$.

   (b) *If* $\mu > \frac{l+h}{2}$*, the receiver's optimal stage 1 garbling is* $\{y_2(\mu), h\}$*, where* $y_2(\mu) < \mu$ *is either on* $[l + k(h - l)^2, h - k(h - l)^2]$ *or on* $(h - k(h - l)^2, h - k(\mu - h)^2)$.

Proof. Let $U_1(x)$ be the receiver's first stage continuation payoffs for a first stage belief x. Say the stage 2 distribution following x has support $\{y_1, y_2\}$, with $y_1 \leq y_2$ and $\nu y_1 + (1-\nu) y_2 = \mu$. Then $U_1(x) = \nu U_2(y_1; x) + (1-\nu) U_2(y_2; x) - k(x - \mu)^2$. The concavification of U_1 over $[l, h]$ is used to obtain the stage 1 optimal distribution.

For any μ, U_1 is continuous. Note that U_1 is affine over any interval of x for which the stage 2 optimal garbling is $\left\{x - \frac{1}{4k}, x + \frac{1}{4k}\right\}$.

Remark. If the stage 1 optimal garbling is unique, then it cannot have support $\{\mu\}$.

The reason for this is the following. If the stage 1 unique optimal garbling is degenerate, then it is verified from Lemma A.1 that the stage 2 optimal garbling has binary support, say $\{y_1, y_2\}$. But then, choosing the garbling $\{y_1, y_2\}$ at stage 1 and $\{\mu\}$ at stage 2 must give the same expected payoff, and hence must be optimal. This is a contradiction.

Now, first let $k > \frac{1}{2(h-l)}$ and $\mu \leq \min\left\{h - \frac{1}{2k}, l + \frac{1}{2k}\right\}$.

Then U_1 is strictly convex in a right neighborhood of $k(\mu - l)^2$ and concave everywhere else (weakly on $\left(l + \frac{1}{4k}, \mu + \frac{1}{4k}\right)$). Then, the concavification must join points $z_1 \leq k(\mu - l)^2$

and $z_2 > k(\mu - l)^2$ (in a straight line), with z_1, z_2 determined by a condition analogous to Inequation 3.

Say $\mu \geq l + \frac{1}{4k}$. Then it is verified that $z_1 = \mu - \frac{1}{4k}$ and $z_2 = l + \frac{1}{4k}$. Since $\mu \in \left[l + \frac{1}{4k}, \mu + \frac{1}{4k}\right]$ and U_1 is affine over this interval, a distribution with support on $\left\{\mu - \frac{1}{4k}\right\} \cup \left[l + \frac{1}{4k}, \mu + \frac{1}{4k}\right]$ would be optimal.

Now say $\mu < l + \frac{1}{4k}$. Clearly the lower bound l would bind and $z_1 = l$ must hold. z_2 is obtained from the second equality in Inequation 3, and it must be higher than μ, since otherwise the optimal garbling would uniquely be degenerate, and we ruled that out above. z_2 is denoted by $y_1(\mu)$ in the statement of the Lemma.

Now let $k > \frac{1}{2(h-l)}$ and $\mu \geq \max\left\{h - \frac{1}{2k}, l + \frac{1}{2k}\right\}$. The argument is symmetric to the preceding one.

In this case U_1 is strictly convex in a left neighborhood of $h - k(h-\mu)^2$ and concave everywhere else (weakly on $(\mu - \frac{1}{4k}, h - \frac{3}{4k})$). The concavification is obtained by joining points z_1 and z_2 as before.

It is verified that for $\mu \leq h - \frac{1}{4k}$, $z_1 = h - \frac{1}{4k}$ and $z_2 = \mu + \frac{1}{4k}$. This tells us that a distribution with support on $\left[\mu - \frac{1}{4k}, h - \frac{1}{4k}\right] \cup \left\{\mu + \frac{1}{4k}\right\}$ would be optimal.

For $\mu > h - \frac{1}{4k}$, $z_2 = h$ must hold. Now z_1 is found from the first equality in Inequation 3, and it must be lower than μ, since otherwise the stage 1 optimal garbling would uniquely be degenerate. z_1 is denoted by $y_2(\mu)$ in the statement of the Lemma.

Cases *3* and *4* are dealt with completely analogously.

Finally, let $k \leq \frac{1}{2(h-l)}$.

Then U_1 is strictly convex in a right neighborhood of $l + k(\mu - l)^2$, and in a left neighborhood of $h - k(h-\mu)^2$, and *strictly* concave everywhere else.

Clearly, the concavification must:

1. join points $z_1 \in [l, l + k(\mu - l)^2)$ and $z_2 > l + k(\mu - l)^2$ in a straight line, and
2. join points $z_3 < h - k(h-\mu)^2$ and $z_4 \in (h - k(h-\mu)^2, h]$ in a straight line.

As usual, these points are determined by a condition analogous to Inequation 3. It turns out that $z_1 = l$ and $z_4 = h$, while the positions of z_2 and z_3 depend on parameters. The optimal garbling is either $\{l, z_2\}$ or $\{z_3, h\}$, depending on where μ lies.

■

The previous result immediately gives us the following useful corollary.

Corollary 4. *The following two statements are equivalent:*

1. $\mu \in [l + \frac{1}{4k}, h - \frac{1}{4k}]$ *and* $k > \frac{1}{2(h-l)}$.

2. *There are multiple stage 1 optimal garblings for the receiver, including support* $\{\mu\}$ *and support* $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$.

Lemma A.3 (Sender payoffs)**.** *Suppose* $k > \frac{1}{2(h-l)}$ *and* $\mu \in [l + \frac{1}{4k}, h - \frac{1}{4k}]$*. If the receiver's behavior is as specified in Lemmata A.1 and A.2, then conditional on being the first sender to be visited, the probability of being selected is the same regardless of which stage 1 optimal garbling is chosen by the receiver.*

Proof. We show the proof for $\mu \leq \min\left\{h - \frac{1}{2k}, l + \frac{1}{2k}\right\}$. It is entirely analogous for the other cases from Lemma A.1.

Suppose $l + \frac{1}{4k} \leq \mu \leq \min\left\{h - \frac{1}{2k}, l + \frac{1}{2k}\right\}$ and the receiver's first stage response is a distribution F on $\left\{\mu - \frac{1}{4k}\right\} \cup [l + \frac{1}{4k}, \mu + \frac{1}{4k}]$.

Using Lemma A.1 it is easy to see that the probability of the first sender being selected conditional on a first stage belief x is given by

$$P(x) = \begin{cases} 0 & if\ x = \mu - \frac{1}{4k} \\ 2kx - 2k\mu + \frac{1}{2} & if\ x \in [l + \frac{1}{4k}, \mu + \frac{1}{4k}] \end{cases}$$

Suppose that F places a mass $p \geq 0$ on $\mu - \frac{1}{4k}$. Then conditional on being visited first, a sender's expected probability of being selected is given by

$$V_1 = p * 0 + \int_{l+\frac{1}{4k}}^{\mu+\frac{1}{4k}} P(x)\ dF(x) \tag{4}$$

Next note that

$$p(\mu - \frac{1}{4k}) + \int_{l+\frac{1}{4k}}^{\mu+\frac{1}{4k}} x dF(x) = \mu \tag{5}$$

and

$$\int_{l+\frac{1}{4k}}^{\mu+\frac{1}{4k}} dF(x) = 1 - p \tag{6}$$

Inserting Equations 3 and 4 into Equation 2, we get that $V_1 = \frac{1}{2}$, which is independent of F. ■

A.2 Proof of Proposition 4.6

Suppose each sender offers support $\{l, h\}$, with $l \in [0, \mu)$ and $h \in (\mu, 1]$.

First let $k > \frac{1}{2(h-l)}$ and $\mu \in [l + \frac{1}{4k}, h - \frac{1}{4k}]$.

Given a stage 1 draw x, the receiver's optimal stage 2 garbling is specified in Lemma A.1. If this garbling does not have support $\{\mu\}$, the receiver necessarily visits the second sender. If it is $\{\mu\}$, she is indifferent between visiting him and not, and may choose either way.

At stage 1, she has multiple best responses. The most informative one among them has support $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$, and from Lemma A.2 it is the only one that is necessarily followed by no learning at stage 2. We assume that she breaks her indifference in favor of this distribution.

At belief $\mu - \frac{1}{4k}$ she accepts the first sender with certainty, and at belief $\mu + \frac{1}{4k}$ accepts the other one with certainty.

Then if a sender deviates to a different distribution, his payoffs may be affected only if he is visited first and the distribution he deviates to is such that $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$ is not a garbling of it.

In this case, regardless of the deviation, the receiver can secure a payoff equal to what she gets in the absence of the deviation, by picking $\{\mu\}$ at stage 1, followed by visiting the other sender and choosing $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$. Thus the deviation cannot force the receiver to choose from outside the set of optimal garblings from Lemma A.2.

But then due to Lemma A.3, the deviating sender's payoffs are unaffected. Thus, there does not exist a profitable deviation and we have an equilibrium.

Next say that either $k > \frac{1}{2(h-l)}$ and $\mu \notin [l + \frac{1}{4k}, h - \frac{1}{4k}]$, or $k \leq \frac{1}{2(h-l)}$.

Then from Lemmata A.2 and A.1, the receiver chooses a unique binary garbling at stage 1, and exactly one belief in the support is followed by a visit to the second sender.

Denote the stage 1 belief following which the receiver does learn at stage 2 by w. Under each possibility we show that there is a profitable deviation for a sender.

Possibility 1: If $w < \mu$, then $\exists\, l' \in [\max\left\{0, \mu - \frac{1}{2k}\right\}, \mu)$ s.t $w = l' + k(\mu - l')^2$. Suppose a sender deviates to support $\{l', h\}$. Then from Lemma A.1, if the deviating sender is visited second, the receiver chooses support $\{\mu\}$ and selects the deviating sender with certainty. This does not affect the receiver's behavior if the deviating sender is visited first, since $l' < w$. Thus the sender profits from this deviation.

Possibility 2: If $w > \mu$ and is followed by a stage 2 best response $\left\{l, l + \sqrt{\frac{w-l}{k}}\right\}$, then it must be true that $w \in (l + k(\mu - l)^2, l + \frac{1}{4k})$. $\exists\, h' < l + \sqrt{\frac{w-l}{k}}$ s.t $w < h' - k(h' - l)^2$. Suppose a sender deviates to $\{l, h'\}$. Then $k \leq \frac{1}{2(h'-l)}$, and Lemma A.1 tells us that if the deviating sender is visited at stage 2, the receiver's response changes to $\{l, h'\}$. $w < h' < l + \sqrt{\frac{w-l}{k}}$ implies that this is profitable if visited at stage 2, without affecting what happens if visited at stage 1. Thus the deviation is profitable.

Possibility 3: If $w > \mu$ and is followed by a stage 2 best response of $\{l, h\}$. Then it is seen that $w < h$, so that h does not bind at stage 1. This implies that a sender can increase or decrease h slightly to h' (so that $\{l, h'\}$ is instead chosen), without affecting what happens if he is visited first. This is clearly profitable. ■

A.3 Proof of Proposition 4.2

"Only if": Suppose that there is no equilibrium in which both senders offer full info. Then, Proposition 4.3 tells us that either $k \leq \frac{1}{2}$, or $k > \frac{1}{2}$ and $\mu \notin [\frac{1}{4k}, 1 - \frac{1}{4k}]$. Lemma A.1 and Lemma A.2 tell us the receiver's unique best response (on path) to full info from both senders. Now we need to show that there is no equilibrium where she gets her first-best payoff. For the sake of contradiction, suppose that there is such an equilibrium–and where sender i offers some p_i. From the discussion in the main text, this just means that the receiver's best response on path to (p_1, p_2), is the same as the best response to full information. We argue, however, that the same deviations that we identified for full info, also work for this supposed equilibrium. Recall the nature of those deviations from A.2: they do not make a difference if the deviating sender is visited first, and restrict learning if visited second. Now if p_1, p_2 is the equilibrium under consideration and the same deviation occurs, the receiver's response to this deviation would be as under full info: if she visits the deviating sender first, she would realize she can continue to choose as on path; if she visits him second, she would make the same adjustment as under the full info scenario.

Thus, since the deviation was profitable under full info, it must be profitable here, and p_1, p_2 cannot be an equilibrium. ■

A.4 Proof of Proposition 4.3

See the proof of Proposition 4.6, setting $l = 0, h = 1$.

A.5 Proof of Proposition 4.4

Existence of the uninformative equilibrium is proven in the text. Here we show non-existence of a full information equilibrium.

Suppose that each sender chooses a fully informative distribution. Because each sender has chosen the same distribution (on path), the receiver is indifferent as to whom she visits first. Hence, suppose that she visits sender 1 first with probability $\lambda \in [0, 1]$ and sender 2 with its complement.

If sender 1 is visited first, then upon the receiver's visit, 1 is realized with probability μ. At this point, she will stop and select sender 1. On the other hand, if 0 is realized then she will select sender 2 without visiting. The symmetric statements hold for sender 2 and her payoff is

$$u_2 = \lambda(1 - \mu) + (1 - \lambda)\mu$$

Now suppose that sender 2 deviates and chooses a distribution that consists of 1 with probability $\eta := \mu - \frac{1}{n}$, $n \in \mathbb{N}$, $n > \frac{1}{\mu}$, and ϵ with probability $1 + \frac{1}{n} - \mu$, where $\epsilon := \frac{1}{n+1-\mu n}$. If sender 1 is visited first then again sender 2 obtains an expected payoff of $(1 - \mu)$. If sender 2 is visited first, with probability η, 1 is realized and sender 2 is selected and with probability $(1 - \eta)$, ϵ is realized. At this point the receiver visits sender 1 and obtains a realization of 0 with probability $1 - \mu$, at which point she selects sender 2. Accordingly,

$$u_2 = \lambda(1 - \mu) + (1 - \lambda)\left(\eta + (1 - \eta)(1 - \mu)\right)$$

and so sender 2 has a profitable deviation if and only if

$$\lambda(1 - \mu) + (1 - \lambda)\left(\eta + (1 - \eta)(1 - \mu)\right) > \lambda(1 - \mu) + (1 - \lambda)\mu$$

which reduces to

$$\frac{1 + 2n - \sqrt{1 + 4n}}{2n} > \mu$$

provided $\lambda < 1$. Without loss of generality we may assume this, since otherwise the same argument would suffice for a deviation by sender 1.

The limit of the left hand side goes to 1 as n goes to ∞; hence for any $\mu < 1$ there exists a $\hat{n}$ such that the left hand side is strictly greater than μ for all $n > \hat{n}$. We conclude that for any $\mu < 1$ there exists a profitable deviation, negating the possibility that full information is an equilibrium. ■

A.6 Proof of Claim 4.1

For $\mu \leq \frac{1}{2}$:

Let each sender choose the uniform distribution on $[0, 2\mu]$, and suppose that the receiver visits sender 1 first with probability $\lambda \in [0, 1]$ and sender 2 with its complement.

No matter the realization at stage 1, the receiver will proceed and visit the other sender as well before selecting one of them. Hence, $u_1 = u_2 = \frac{1}{2}$. Next, we check for a profitable deviation. Suppose sender 1 deviates to a distribution that contains a probability measure of size a on $[2\mu, 1]$ and some portion F on $[0, 2\mu)$. It is clear that it is without loss of generality to set a to be a point mass on 2μ.

If sender 1 is visited first then with probability a, he is selected and sender 2 is never visited; and otherwise, sender 2 is visited after which the receiver selects the sender with the highest realization. If sender 2 is visited first, then no matter what, sender 1 is also visited, after which the comparison ensues.

Sender 1's payoff is

$$u_1 = \lambda \left(a + \int_0^{2\mu} \int_0^x dG(y)dF(x) \right) + (1-\lambda) \left(a + \int_0^{2\mu} \int_0^x dGdF \right) = a + \int_0^{2\mu} \int_0^x dG(y)dF(x)$$

where $G(y) = \frac{y}{2\mu}$ is the (on-path) distribution chosen by sender 2 and where $\int_0^{2\mu} dF = 1 - a$ and $\int_0^{2\mu} xdF = 2 - 2\mu a$.

Next, we use the result in Whitmeyer and Whitmeyer (2020) who establish that it suffices to show that 1 has no profitable deviation to any binary distribution. Let F be described by α with probability p and β with probability $1 - p$; where $0 \leq \alpha \leq \mu$, $\mu \leq \beta \leq 2\mu$, and $\alpha p + \beta(1 - p) = \mu$. Consequently, we rewrite u_1, which becomes

$$\begin{aligned} u_1 &= (1-p)F(\beta) + pF(\alpha) \\ &= (1-p)\frac{\beta}{2\mu} + p\frac{\alpha}{2\mu} = \frac{1}{2} \end{aligned}$$

Hence, there is no profitable deviation. ■

For $\mu > \frac{1}{2}$:

On path, sender 1's payoff is

$$u_1 = \lambda\left(2 - \frac{1}{\mu} + \int_0^{2(1-\mu)} \int_0^x \frac{1}{2\mu}\frac{1}{2\mu} dydx\right) + (1-\lambda)\left(\left(\frac{1}{\mu} - 1\right)\left(2 - \frac{1}{\mu}\right) + \int_0^{2(1-\mu)} \int_0^x \frac{1}{2\mu}\frac{1}{2\mu} dydx\right)$$
$$= \frac{2\left(2\mu-1\right)^2 \lambda + \left(1-\mu\right)\left(3\mu-1\right)}{2\mu^2}$$

If sender 1 deviates to 1 with probability μ and 0 with probability $1-\mu$, his payoff from deviating is

$$u_1^D = \lambda\mu + (1-\lambda)\left(\frac{1}{\mu} - 1\right)\mu = 1 + 2\lambda\mu - \lambda - \mu$$

The difference, $u_1^D - u_1$ is

$$\frac{\left(1-\mu\right)^2\left(2\mu-1\right)\left(2\lambda-1\right)}{2\mu^2}$$

Since $\mu > \frac{1}{2}$, this is positive provided $\lambda > \frac{1}{2}$ and negative provided $\lambda < \frac{1}{2}$. Thus, if $\lambda \neq \frac{1}{2}$ there exists a profitable deviation (if $\lambda < \frac{1}{2}$, sender 2 can deviate profitably in the analogous fashion).

It remains to show that this vector of distributions is an equilibrium for $\lambda = \frac{1}{2}$. Substituting $\lambda = \frac{1}{2}$ into u_1, we see that $u_1 = \frac{1}{2}$ on path. Just as for $\mu \leq \frac{1}{2}$, from Whitmeyer and Whitmeyer (2020) we need check only deviations to binary distributions. Let F be described by α with probability p and β with probability $1-p$, where $\alpha p + \beta(1-p) = \mu$ and $0 \leq \alpha \leq \mu$. There are two cases that we need to consider. 1. $\mu \leq \beta \leq 2(1-\mu)$; and 2. $\beta = 1$. In the first case,

$$u_1 = (1-p)F(\beta) + pF(\alpha)$$
$$= (1-p)\frac{\beta}{2\mu} + p\frac{\alpha}{2\mu} = \frac{1}{2}$$

and in the second case

$$u_1 = \frac{1}{2}\left(1 - p + pF(\alpha)\right) + \frac{1}{2}\left(\left(\frac{1}{\mu} - 1\right)(1-p) + pF(\alpha)\right)$$
$$= p\frac{\alpha}{2\mu} + \frac{1-p}{2\mu} = \frac{1}{2}$$

where we used the fact that $\beta = 1$ implies that $1 - p = \mu - p\alpha$. Hence, there is no profitable deviation. ■

A.7 Proof of Lemma 4.1

See the proof of Lemma A.1, setting $l = 0, h = 1$ and $k = 1$.

A.8 Proof of Lemma 4.2

See the proof of Lemma A.2, setting $l = 0, h = 1$ and $k = 1$.

A.9 Proof of Lemma 4.3

See the proof of Lemma A.3, setting $l = 0, h = 1$ and $k = 1$.

A.10 Proof of Claim 4.2

Let $k > \frac{1}{2}$ and $\mu \in [\frac{1}{4k}, 1 - \frac{1}{4k}]$. As shown in Appendix A.2, one of the receiver's best responses to full information ($l = 0, h = 1$) from both senders is to choose the garbling $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$ at stage 1 and to learn nothing at stage 2.

Suppose sender i offers a distribution of which $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$ is a garbling. Then, the aforementioned best response to full information is permissible, and thus continues to be a best response. Suppose the receiver chooses this response.

Then if a sender unilaterally deviates and is the one to be visited first, the receiver may respond by choosing $\{\mu\}$ and visiting the other sender, choosing $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$ for him. Exactly as in the proof for existence of a full information equilibrium (Proposition 4.6 for $h = 1, l = 0$), Lemma A.3 can be used to argue that the deviation cannot be profitable. ■

A.11 Proof of Corollary 3

For $\boldsymbol{\mu \leq \frac{1}{2}}$:

We show that $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$ is a mean preserving contraction of the uniform distribution on $[0, 2\mu]$ when $k \geq \frac{1}{2\mu}$.

Define $l(x)$ as

$$l(x) = \begin{cases} 0 & 0 \leq x < \mu - \frac{1}{4k} \\ \frac{1}{2}x - \frac{\mu}{2} + \frac{1}{8k} & \mu - \frac{1}{4k} \leq x < \mu + \frac{1}{4k} \\ x - \mu & \mu + \frac{1}{4k} \leq x \leq 1 \end{cases}$$

Define $j(x) := \int_0^x G(t)dt$:

$$j(x) = \begin{cases} \frac{x^2}{4\mu} & 0 \leq x < 2\mu \\ x - \mu & 2\mu \leq x \leq 1 \end{cases}$$

It suffices to show that $\mu > \frac{1}{4k}$, that $j(x) - l(x) = 0$ has at most one real root, and that $j\left(\mu + \frac{1}{4k}\right) > l\left(\mu + \frac{1}{4k}\right)$.

Set $j(x) = l(x)$, which holds if and only if

$$x = \frac{4\mu k \pm \sqrt{8k\mu\left(1 - 2k\mu\right)}}{4k}$$

This is imaginary if and only if

$$k > \frac{1}{2\mu}$$

and has a unique root for $k = \frac{1}{2\mu}$ (at μ). $\mu - \frac{1}{4k} \geq \frac{\mu}{2} > 0$ for $k \geq \frac{1}{2\mu}$. It remains to verify that $j\left(\mu + \frac{1}{4k}\right) > k\left(\mu + \frac{1}{4k}\right)$; but it is simple to verify that this must hold. Thus, if $k \geq \frac{1}{2\mu}$, we have the result.

For $\mu > \frac{1}{2}$:

The proof is analogous to the preceding one, with the exception that k must be sufficiently large so that $\mu + \frac{1}{4k} \leq 1$. This holds if and only if $k \geq \frac{1}{4(1-\mu)}$. This constraint binds for $\mu \geq \frac{2}{3}$ and $k \geq \frac{1}{2\mu}$ binds for $\mu \leq \frac{2}{3}$. ■

A.12 Proof of Proposition 5.1

Suppose each sender offers support $\{l, h\}$, with $l \in [0, \mu)$ and $h \in (\mu, 1]$.

Lemma A.1 and Lemma A.2 continue to describe on path behavior. Lemma A.3 still holds.

First let $k > \frac{1}{2(h-l)}$ and $\mu \in [l + \frac{1}{4k}, h - \frac{1}{4k}]$.

On path behavior is exactly as in the baseline model: visit any one sender, pick $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$ and take a decision without learning from the other sender.

Then if a sender deviates to a different distribution, it would be observed. Then the receiver can simply respond by visiting the other, non-deviating sender, and picking $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$ for him and taking a decision immediately.

Due to Lemma A.3, the deviating sender's payoffs are the same as on path. Thus, there does not exist a profitable deviation and we have an equilibrium.

Next say that either $k > \frac{1}{2(h-l)}$ and $\mu \notin [l + \frac{1}{4k}, h - \frac{1}{4k}]$, or $k \leq \frac{1}{2(h-l)}$.

Then from Lemmata A.2 and A.1, on path the receiver chooses a unique binary garbling at stage 1, and exactly one belief in the support is followed by a visit to the second sender.

Denote the stage 1 belief following which the receiver does learn at stage 2 by w. Under each possibility we show that there is a profitable deviation for a sender.

Possibility 1: Say $w < \mu$ and the stage 2 garbling is $\{l, h\}$. There must be a sender, say sender i, who is visited first with probability < 1 on path. Suppose sender i deviates to $\{l', h\}$, where $l < l' < w$. But on observing this deviation, the receiver would choose to visit sender i first. By doing this she could get her first-best. Thus, behavior is as on path, except that the order of visits is changed: sender i is visited first with probability 1. It is easy to verify that the payoff from being visited first is $> \frac{1}{2}$ (i.e. higher than payoff from being visited second), which means that this increase in probability of being visited first is profitable.

Possibility 2: Say $w < \mu$ and the stage 2 garbling is $\left\{h - \sqrt{\frac{h-k}{k}}, h\right\}$. Everything is as in possibility 1, except that l' is chosen such that $h - \sqrt{\frac{h-k}{k}} < l' < w$.

Possibility 3: If $w > \mu$ and is followed by a stage 2 best response $\left\{l, l + \sqrt{\frac{w-l}{k}}\right\}$ or $\{l, h\}$. Then if a sender deviates to no information, clearly the receiver would just learn from the other sender with a threshold of acceptance μ. It is verified that the deviating sender's payoffs then are higher than the payoffs on path, conditional on being visited first as well as conditional on being visited second. ■

A.13 Proof of Lemma 5.1

See the proof of Lemma A.2, setting $l = 0, h = 1$ and $k = 1$ and using μ_2 as the mean for the second sender and μ_1 as the mean for the first sender.

A.14 Proof of Lemma 5.2

Let us begin by looking at the parametric conditions given in bullet points 1 and 3 of Lemma 5.1. By symmetry it suffices to assume that one of these two pairs of conditions holds for the scenario in which sender 2 is visited second, and show that that implies that one of the four pairs of conditions for the scenario in which sender 2 is visited first must hold. Observe that the conditions for bullet points 1 and 3 reduce to $|\mu_1 - \mu_2| \leq \frac{1}{4}$ and $\mu_2 \in \left[\frac{1}{4}, \frac{3}{4}\right]$. It is easy to see that if $\mu_1 \in \left[\frac{1}{4}, \frac{3}{4}\right]$ then we are done. What if $\mu_1 \notin \left[\frac{1}{4}, \frac{3}{4}\right]$? WLOG suppose that $\mu_1 < \frac{1}{4}$. By assumption we must have $\mu_2 - \frac{1}{4} \leq \mu_1$ and $\mu_2 \geq \frac{1}{4}$. Hence, condition 4 (with μ_2 and μ_1 transposed) must hold.

Next, we turn our attention to the conditions given in bullet points 2 and 4. WLOG it suffices to focus on the conditions in bullet point 2. As we did in the previous paragraph, it suffices to assume that these conditions hold for the scenario in which sender 2 is visited second, and show that that implies that one of the four pairs of conditions for the scenario in which sender 2 is visited first must hold. By construction, $\mu_2 \leq \mu_1 + \frac{1}{4}$ and $\mu_1 \in \left[\frac{1}{2}, \frac{3}{4}\right]$. Moreover, $\mu_2 \geq \mu_1 > \mu_1 - \frac{1}{4}$, and so condition 1 (with μ_2 and μ_1 transposed) must hold. ■

A.15 Proof of Proposition 5.3

It suffices to show that conditional on being the first sender to be visited, the probability of being selected is the same regardless of which stage 1 optimal garbling is chosen by the receiver. The remainder of the proof follows analogously to the proof of Lemma A.3. Alternatively, observe that it follows from the fact that probability of the first sender being selected conditional on a first stage belief x is either 0, 1, or a function that is affine in x.
■