Competitive Disclosure of Information to a Rationally Inattentive Agent

Vasudha Jain* Mark Whitmeyer†

January 6, 2022

Abstract

We study competitive disclosure of information on idiosyncratic product quality by two firms to a rationally inattentive consumer. Unless attention costs are low, there is an equilibrium in which the firms provide the consumer with as much information as she would process if she controlled information provision. This is not true if there is only one firm. We identify a novel channel through which the interaction of competition and inattention encourages information disclosure: information on one firm substitutes for information on the other, rendering a unilateral withholding of information unprofitable.

Keywords: Bayesian persuasion; Information design; Multiple senders; Competition; Rational Inattention; Search
JEL Classifications: D82; D83

*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 V. Bhaskar, Max Stinchcombe, Caroline Thomas, Joseph Whitmeyer, Tom Wiseman and especially Vasiliki Skreta for very helpful feedback and discussions. Thanks also to participants at numerous seminars and conferences for comments that helped improve this paper.

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. It is natural to expect that firms, when crafting their information disclosure strategies, account for the fact that merely making information available may not suffice to influence beliefs–consumers must find it in their interest to invest in processing that information.

In this paper, we study how much information about product quality is 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 information. A firm's disclosure decision thus places an upper bound, but not a lower bound, on the information a consumer obtains.

A key implication of the consumer's information processing costs is that it might not be optimal for her 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, even if both firms provide perfectly accurate information on the quality of their respective products, the consumer might rationally ignore some of each firm's information. 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, a firm 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.

We find that as long as information processing costs and the prior uncertainty about the senders' types are not too low, the answer is yes. Namely, 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. Further, any symmetric equilibrium in a natural class of strategies is shown to be outcome equivalent to this equilibrium.

Prior literature has shown that the receiver is *not* provided with her first-best level of information in two benchmark scenarios: first, where information processing is costless; second, where information processing is costly but there is only one firm. Thus, both ingredients–competition and information processing costs–are critical for our results, and the key contribution of this paper is to uncover the strategic forces that lead to this. In particular, we illuminate the role played by *substitutability* of information sources: highly precise information on one firm diminishes the value of information about the other.

Our second main result–which is a consequence of the first–is that the presence of information processing costs may actually lead a consumer to ultimately process *more* information (and therefore select the better firm with a higher probability). This suggests that a regulator who wishes to promote the purchase of better quality products should be cautious in assessing the implications of costly attention.

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 and vaccines.

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 or match value) drawn independently.[1] Before acquiring any private information, each of them decides how much information to disclose about his own type.[2] Rather than impose 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.

There is a receiver (e.g. consumer or doctor), who wishes to choose the sender with the higher type. She visits the senders sequentially, and chooses the order of visits after observing how informative each sender's disclosure rule is. Upon visiting the first sender, she may choose to acquire less precise, or coarser, information than what is available. Formally, she is free to choose any mean preserving contraction–or *garbling*–of the sender's Blackwell 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. Think back to the doctor example, and the shortcuts she might take: she could 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 visits the second sender and follows the same protocol: she chooses a garbling of that sender's chosen experiment subject to an information cost. Finally, she selects the sender favored by her posterior beliefs. Each sender wants to maximize the probability of being selected.

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 learn anything about the second sender–she might be so optimistic about the first sender that she selects him without learning further, and she may be so pessimistic that she chooses

[1] Our focus is on the 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 about his competitor.

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.[3] 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 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.

We ask whether in our game with strategic senders, there is an equilibrium in which senders might voluntarily provide as much information as the receiver would acquire in her first-best scenario; and find that the answer is yes under general conditions. In particular, for any prior, there is such an equilibrium[4] as long as an attention cost parameter is above a threshold.[5] 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.

1.1 Related Literature

Our work relates thematically to several strands of the literature.

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

[4]In fact, there is a continuum of such equilibria and they are all equivalent not only in terms of receiver payoffs, but also in terms of sender payoffs.

[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.

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. (2021) provide an excellent review of this literature and its wide range of applications.

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 in order to manipulate 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-type, binary-action model with a single principal who has state independent preferences and an exogenous threshold of acceptance for the agent. He finds that the principal restricts the agent's learning below her first-best level. In the present paper, we show how competitive forces change this.

Bloedel and Segal (2021) 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. The receiver observes the realized perception but 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. Matyskova and Montes (2021) study a persuasion model where the receiver, after observing the sender's signal realization, 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 (which restricts the receiver's welfare below her first-best level).[6]

Au and Whitmeyer (2021) extend the competitive persuasion scenario to a sequential (directed) search setting, and allow for fixed search (visit) costs. They find that search costs may encourage information provision by the sellers because firms compete to be visited first.

[6]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), 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.

In contrast, we find that information processing costs enable the consumer to make a credible threat that sustains a highly informative equilibrium–if a seller deviates, the consumer will learn nothing from him, rendering that deviation unprofitable.

Board and Lu (2018) also look at sellers competing through information to entice buyers. However, there, the number of sellers is uncountable, search is random (and so the order of visits is not endogenously chosen), and the decision by a seller of what information to disclose is made upon the buyer's visit. Moreover, the value of each seller's goods to the buyer are perfectly correlated whereas here they are independent.

Costs borne by sender: Gentzkow and Kamenica (2014) look at optimal persuasion mechanisms when a single sender pays higher costs (proportional to entropy reduction) of designing more informative experiments. Le Treust and Tomala (2019) consider constraints on the sender's information transmission channel and 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, which is analyzed in Section 3. Section 4 compares our main results to two benchmarks. Section 5 discusses how the receiver's welfare varies with attention costs. Section 6 illustrates the robustness of our results to three modifications of the baseline model: one where the senders are *ex ante* heterogeneous, one where attention costs directly depend on senders' disclosure rules, and one where each sender's type is an arbitrary real-valued random variable and the receiver is risk-neutral. Section 7 concludes. The appendix contains omitted proofs.

2 Baseline Model

There are two senders indexed by $i \in \{1, 2\}$, and a receiver. Sender i has type $\omega_i \in \Omega_i := \{0, 1\}$, 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. 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

[7] Our results continue to hold with a low, positive outside option.

$\{0,1\}$, 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 observes each sender's choice of experiment (but not its realization[8]) and decides which sender to visit first. Say she visits Sender 1 first. Then she 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; and incurs an attention cost proportional to the variance of her posterior belief:

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

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 visits the other sender (say Sender 2) and chooses a garbling q_2 of p_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 a sender and her payoff is equal to the type of the selected sender. She may select a sender even if she learned nothing about him, i.e., if she chose the uninformative garbling of his distribution.

Figure 1 summarizes 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. That is, 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. Note also that 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 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 $\{\nu_1, \nu_2\} \iff supp(q) \subseteq [\min\{\nu_1,\nu_2\}, \max\{\nu_1,\nu_2\}]$.

[8]In fact, one can think that realizations from the senders' experiments are never drawn.

[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, the receiver always has the option of choosing $q_1 = p_1$ or $q_1 = \delta_\mu$.

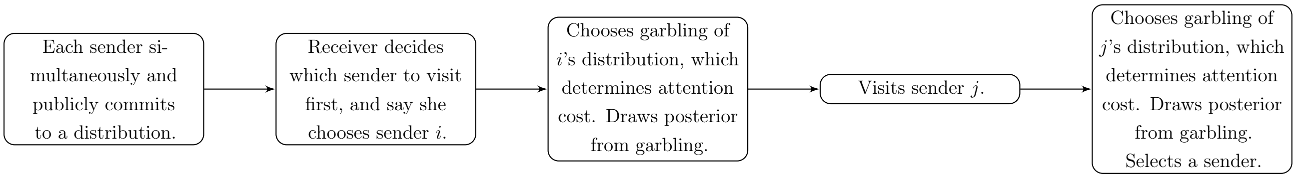

Figure 1: Sequence of moves

Strategies and solution concept. A pure strategy for Sender i is a choice of a distribution $p_i \in \Delta[0,1]$ with mean μ. A pure strategy for the receiver consists of the following for each pair of distributions offered by the senders: i) a choice of which sender to visit first; ii) a choice of garbling of the first sender's distribution; iii) a choice of garbling for the second sender's distribution for each belief that can be drawn about the first sender; iv) a choice of which sender to select at the end for each pair of posterior beliefs that can be drawn.

We assume that if the receiver holds the same belief about both senders, she selects each of them with equal probability.[10] We also restrict senders to pure strategies and disallow the receiver from mixing over garblings–this eases notation but does not have substantive implications for our results.

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 learning that a sender has deviated, the receiver does not update her belief about either sender's type.

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 of that. Her costs would depend on how much of the research she chooses to read, not on

[10] We need not assume anything about the tie-breaking rule for our main analysis, but we do so to maintain consistency with other papers on competitive persuasion which serve as a benchmark.

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.[11] Accordingly, the cost specified by equation 1 is simply proportional to the variance of the posterior belief under the receiver's chosen garbling.[12]

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. 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 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 types–i.e., to introduce correlation between the two experiments she learns from.[13] An option to do so might be valuable for the receiver by reducing the expected cost of acquiring information of a certain gross value.

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

[12] This cost function belongs to the family of Tsallis-entropy-based-cost functions (see Tsallis (1988)). It is also familiar from Lipnowski et al. (2020b) and Eliaz and Eilat (2021).

[13] 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.

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 Equilibrium Analysis

In this section, we begin by analyzing the game described in Section 2 for arbitrary $k > 0$ and $\mu \in (0, 1)$. As will become clear, we may have a multiplicity of equilibria that correspond to the same distribution of payoffs. We focus our attention on equilibrium payoff distributions, and in particular on existence (or non-existence) of equilibria that give the receiver her *first-best* payoff. This is the highest payoff that she may secure across all profiles of sender behavior (but when she is still subject to attention costs); equivalently, it is the payoff that she would attain if the senders' decisions in the game were delegated to her. We discuss reasons for our focus on this class of equilibria in Section 3.2.

Since every distribution with expectation equal to the prior is a garbling of the fully informative distribution, the receiver has greatest latitude when both senders offer the fully informative distribution. Thus, she attains her first-best payoff when both senders offer full information. However, the same payoff may be attained even when senders choose other, less informative distributions: at the heart of this is the fact that due to attention costs, the receiver might not use full information even when allowed to. As an example, suppose that when offered full information, the following is a best response for the receiver: visit Sender 1 first, learn according to 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, the receiver gets to follow the same protocol and secure her first-best payoff. 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 3.1 (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.*

3.1 Equilibrium with full information provision

An implication of Proposition 3.1 is that in order to establish conditions for existence of an equilibrium where the receiver gets her first-best payoff, we need only look for existence of an equilibrium where full information is provided. The next proposition tackles precisely this question.

Proposition 3.2 (Full information equilibrium)**.** *There is an equilibrium in which both senders offer full information if and only if* $k > \frac{1}{2}$ *and* $\mu \in \left[\frac{1}{4k}, 1 - \frac{1}{4k}\right]$.

Notice that k, the parameter in the attention cost function, is crucial. If it 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. The following corollary, which trivially follows from propositions 3.1 and 3.2, embodies our main result.

Corollary 3.2.1. *There is an equilibrium in which the receiver gets her first-best payoff if and only if* $k > \frac{1}{2}$ *and* $\mu \in \left[\frac{1}{4k}, 1 - \frac{1}{4k}\right]$.

A second corollary states the same result (as in Proposition 3.2) differently.

Corollary 3.2.2. *For any* $\mu \in (0, 1)$*, there is an equilibrium in which both senders offer full information if and only if* $k \geq \max\left\{\frac{1}{4\mu}, \frac{1}{4(1-\mu)}\right\}$ *(strict inequality if* $\mu = \frac{1}{2}$*).*

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 stay entirely ignorant 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 one sender. The proof of Proposition 3.2, which follows next, will clarify how this fact plays a crucial role. We assume here that $k = 1$; the structure of the proof is the same for a generic $k > 0$, and the details are relegated to the appendix.[14]

Given $k = 1$, suppose that each sender offers full information, i.e., each sender chooses the fully informative distribution with support $\{0, 1\}$. We first analyze the receiver's best response and then show that a sender can gain by unilaterally deviating to a different distribution if and only if $\mu \notin \left[\frac{1}{4}, \frac{3}{4}\right]$, thereby confirming Proposition 3.2 for $k = 1$.

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

Recall that one element of the receiver's response is a choice of whom to visit first. Here, since both senders offer the same distribution, she is indifferent between the two orders of visit, and we suppose that with probability $q \in [0,1]$ Sender 1 is the first to be visited. Henceforth, whenever we refer to the *first sender*, we will mean the sender visited first, and correspondingly for the *second sender*.

To analyze how the receiver optimally learns at each stage, we proceed in two steps–first, we determine her stage 2 best response for each belief that could be drawn at stage 1; second, we use that to solve for her optimal stage 1 behavior. The concavification technique, popularized in this literature by Kamenica and Gentzkow (2011), comes in handy here.

Receiver's optimal learning strategy at stage 2: Say the belief drawn at stage 1, about the first sender, is $x \in [0,1]$. Then, the receiver selects the second sender if the stage 2 draw y turns out to be higher than x (and tosses a coin when $y = x$).[15] 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 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 $U_2(y;x) := \max\{x,y\} - (y-\mu)^2$ 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 $[0,1]$.[16] 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 3.3 (Stage 2 optimal garbling)**.** *Suppose that the receiver's stage 1 draw is $x \in [0,1]$. Then her stage 2 optimal garbling is either degenerate or binary, and its support is as follows.*

1. *If $\mu < \frac{1}{2}$,*

$$\begin{cases} \{x - \frac{1}{4}, x + \frac{1}{4}\} & if \; \frac{1}{4} \leq x < \mu + \frac{1}{4} \\ \{0, \sqrt{x}\} & if \; \mu^2 < x < \frac{1}{4} \\ \{\mu\} & if x \leq \mu^2 \; or \; x \geq \mu + \frac{1}{4} \end{cases}$$

[15] As mentioned previously, the tie-breaking rule does not matter for our results. This is because x would never be in the support of the receiver's optimal stage 2 garbling.

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

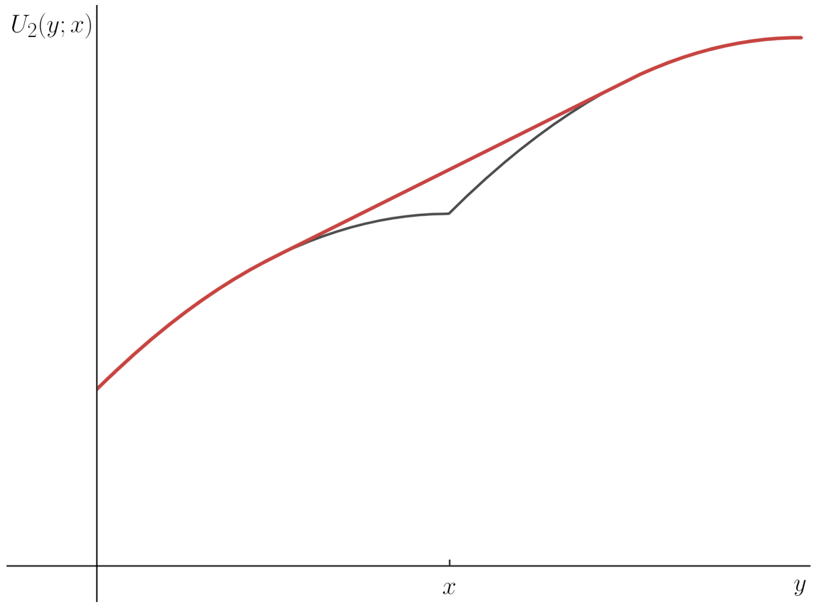

Figure 2: The receiver's stage 2 payoff function and its concavification (red).

2. If $\mu = \frac{1}{2}$,

$$\begin{cases} \{x - \frac{1}{4}, x + \frac{1}{4}\} & if\ \frac{1}{4} < x < \frac{3}{4} \\ \{\mu\} & if\ x \leq \frac{1}{4}\ or\ x \geq \frac{3}{4} \end{cases}$$

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

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

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 type 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 type 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 at the second stage 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 beliefs into one.

Receiver's optimal learning strategy at stage 1: 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 3.4 (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.[17]

The multiplicity of best responses for $\mu \in \left[\frac{1}{4}, \frac{3}{4}\right]$ captures the key 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.

No profitable deviation for a sender when $\boldsymbol{\mu} \in \left[\frac{\mathbf{1}}{\mathbf{4}}, \frac{\mathbf{3}}{\mathbf{4}}\right]$: 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 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 3.3, 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 would rather not put off learning until the next stage.

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

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 first sender. If the belief drawn is $\mu - \frac{1}{4}$, select the second sender without learning anything about him. If the belief drawn is $\mu + \frac{1}{4}$, select the first sender without learning anything about the second one.

Having specified on-path behavior, we examine what happens in the event of a unilateral deviation by a sender to a less-than-fully informative distribution. First note that if $\left\{\mu - \frac{1}{4}, \mu + \frac{1}{4}\right\}$ is a garbling of the distribution that the sender deviates to, then the receiver need not change her behavior, and payoffs are unaffected. If instead $\left\{\mu - \frac{1}{4}, \mu + \frac{1}{4}\right\}$ is *not* a garbling of the new distribution, then we can specify the following protocol for the receiver. She continues with the same order of visits as on path (i.e., with probability q she visits Sender 1 first). If the deviating sender is visited second, then she adopts the same learning behavior as on path (this is clearly feasible regardless of what the deviation is). If the deviating sender is visited first, then she learns nothing from him; instead, she moves on and learns according to the garbling $\left\{\mu - \frac{1}{4}, \mu + \frac{1}{4}\right\}$ about the second (non-deviating) sender (following which, she selects the deviating sender if and only if the lower belief is realized).

Such a response to the deviation is sequentially rational for the receiver: our preceding analysis tells us that this protocol is among the receiver's best responses to full information, and so it preserves her on-path payoff (which in turn is her highest attainable payoff in this game).

Now it remains to be shown that this off-path response renders the sender's deviation unprofitable. The next lemma leads us in that direction.

Lemma 3.5. *For all $\mu \in \left[\frac{1}{4}, \frac{3}{4}\right]$, conditional on being visited first, a sender's expected payoff is equal to $1/2$ for any of the receiver responses specified in Lemma 3.4, provided that her subsequent behavior is dictated by Lemma 3.3.*

Combined with the fact that a deviation affects neither the probability of being visited first, nor the expected payoff conditional on being visited second; this implies that in fact it is not only the receiver, but also the deviating sender himself, whose payoff is unaffected by the deviation. Thus, providing full information is indeed an equilibrium.

Profitable deviation for a sender when $\boldsymbol{\mu \notin \left[\frac{1}{4}, \frac{3}{4}\right]}$: When $\mu \notin \left[\frac{1}{4}, \frac{3}{4}\right]$, the analogous reasoning does not apply because the receiver's best response is unique and involves learning from both senders on path. In this case, as illustrated below, we can construct a profitable deviation for a sender. For low priors a sender gains by deviating to provide no information; for high priors a sender gains by inducing the receiver to visit him first.

If $\mu < \frac{1}{4}$, then on path at least one sender must get a payoff weakly lower than $\frac{1}{2}$. It is easy to see that by unilaterally deviating to the uninformative distribution, this sender can attain a payoff strictly above $\frac{1}{2}$ (Lemma 3.3 tells us that the receiver would learn according to $\{0, \sqrt{u}\}$ from the non-deviating sender).

If $\mu > \frac{3}{4}$, then we have seen that on path the receiver chooses support $\{y_2(\mu), 1\}$ at stage 1. Following belief 1 she immediately selects the first sender, and following belief $y_2(\mu)$, she learns according to $\left\{1 - \sqrt{1 - y_2(\mu)}, 1\right\}$ at stage 2. Now, there must be at least one sender who is visited first with probability less than 1. This sender can profitably deviate to a distribution $\{l, 1\}$ with $1 - \sqrt{1 - y_2(\mu)} < l < y_2(\mu)$, which poses a binding constraint for the receiver only at stage 2. Upon observing this deviation, it is optimal for the receiver to visit the deviating sender first with probability 1 and simply adopt the same learning protocol as on path. That this helps the deviating sender follows from the observation that under the receiver's on-path learning protocol, the first sender's payoff is higher than the second sender's.

3.1.1 Other equilibria

We began this section by asserting that we are interested in equilibria that give the receiver her first-best payoff, and in our search for conditions under which any such equilibrium exists, Proposition 3.1 afforded us the convenience of restricting focus to full information provision. Now, when an equilibrium with full information provision exists, what can we say about *other* first-best equilibria for the receiver? In particular, how might they differ in terms of *sender* payoffs, or in terms of how much information the receiver processes?

In fact, on both these counts, any first-best equilibrium for the receiver is equivalent to the equilibrium with full information provision. Indeed, if the receiver achieves her first-best payoff, her on-path learning protocol must be optimal under full information provision too. It is easy to see that this in turn implies that expected payoffs for senders are also the same in any such equilibrium: for $k = 1$, Lemma 3.5 implies that each sender's expected payoff is $1/2$ when the receiver follows any of her first-best protocols (and the appendix generalizes this for other values of k).

First-best equilibria for the receiver do differ in terms of the information initially provided by senders, but as we have argued, that does not translate to any difference in *outcomes*. In fact, there is a continuum of such equilibria: essentially, since the receiver never chooses to process full information even when it is available,[18] we obtain an equilibrium whenever senders withhold information that the receiver would discard under availability of full information. The next proposition formalizes this claim.

[18]We showed this for $k = 1$, but the appendix shows that it is true in general: in response to full information, the receiver only learns from one sender according to support $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$.

Proposition 3.6 (Equilibria outcome equivalent to full information provision)**.** *Suppose that $k > \frac{1}{2}$ and $\mu \in \left[\frac{1}{4k}, 1 - \frac{1}{4k}\right]$. For $i \in \{1, 2\}$, 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 provision.*

Turning next to the question on equilibria that do not give the receiver her first-best payoff, we have the following result that rules out their existence in the class of *symmetric, binary* equilibria (that is, equilibria where both senders offer the same binary support distribution). For any parameters, if there is a symmetric binary equilibrium, then it must fall within the continuum of first-best equilibria for the receiver that was identified in Proposition 3.6.

Proposition 3.7 (Symmetric binary 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 3.2.

Recalling from Section 2 the characterization of the set of garblings of a binary support distribution, we observe that by offering a binary distribution, a sender essentially places bounds on the receiver's posterior belief about his type, and within those bounds the receiver is free to acquire information flexibly. For example, if a sender offers the distribution $\{1/5, 4/5\}$, then the only restriction on the receiver's learning is that she cannot become more than 80% sure of that sender's type. Given this intuitive interpretation of binary distributions, it is natural to focus on them. Beyond this, characterizing *all* equilibria of the game is beyond the scope of this paper, due largely to the challenge of characterizing the set of garblings of a non-binary distribution.

3.2 Discussion

We have established that given k, as long as the prior is not extreme, there is a continuum of equilibria where the receiver gets her first-best payoff. In all such equilibria, the receiver processes the same amount of information, and expected sender payoffs are the same. In one such equilibrium–the one we focus on for convenience–each sender provides full information.

Importantly, as k grows, the range of "extreme" priors diminishes (and vanishes in the limit). Moreover, there is no symmetric equilbrium where senders follow "simple" strategies of placing bounds on the receiver's belief, and which does not fall within the aforementioned continuum of equilibria.

The primary reason for our focus on first-best equilibria for the receiver is that our analysis of these equilibria highlights the novel economic force that operates in this environment.

Indeed, let us take a closer look at the intuition behind our proof of the existence of an equilibrium with full information provision. Even when full information is provided by both senders, attention costs make it optimal for the receiver to learn only about one sender. A unilateral deviation by the sender about whom the receiver plans to learn can indeed become a binding constraint for the receiver and force a change of behavior, but it cannot prevent the receiver from switching the identity of the sender she learns about: she may always walk away from the deviating sender and learn about the other one instead. This off-path behavior, though, allows the receiver to maintain her first-best payoff: the senders are *ex ante* identical, which means that if it is optimal to learn only about one sender, then it does not matter *which* sender that is. Thus, the receiver is able to fully compensate for under-provision of information by one sender, so that her probability of eventually selecting the better sender remains unaffected. Intuitively, it is only this probability that matters for a sender's expected payoff, and so a sender does not stand to gain (or lose) by deviating.

Our results stand in sharp contrast to two benchmark scenarios: first, where only one sender may provide information; second, where both senders may provide information but the receiver is not subject to attention costs. We turn to a discussion of these benchmarks next.

4 Benchmarks

4.1 Single sender

To highlight the role played by competing senders, suppose that only one sender (say Sender 1) may provide information about his type. In other words, there is a single strategic sender who faces a receiver with outside option μ. Further suppose that $k > \frac{1}{2}$ and $\mu \in \left[\frac{1}{4k}, 1 - \frac{1}{4k}\right]$, so that in our main model with two senders there is a first-best equilibrium for the receiver.

In this setup, if Sender 1 were to offer full information, then the receiver would choose the garbling with support $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$,[19] yielding an expected payoff of 1/2 to Sender 1. The first-best scenario for the receiver is one where Sender 1 offers a distribution that can be garbled down to this binary distribution.

In equilibrium, does Sender 1 provide the receiver with her first-best level of information? The answer is no: a deviation to a distribution with support $\{0, h\}$ is profitable if $h \in (\mu, \mu + \frac{1}{4k})$. Lemma A.1 tells us that the garbling chosen by the receiver in response to this deviation has

[19] This is the receiver's optimal garbling at stage 2 in the main game when the stage 1 draw is μ. We have already computed this for $k = 1$, and Lemma A.1 in the appendix generalizes it.

support $\left\{h - \sqrt{\frac{h-\mu}{k}}, h\right\}$. The resulting probability that Sender 1 is selected is the probability that the receiver's realized belief is h, which is greater than $1/2$. Thus, a single sender does not permit the receiver her first-best payoff.

It turns out that as a consequence of the receiver's outside option being exactly equal to the prior, an optimal distribution for Sender 1 does not exist, and the probability of Sender 1 being selected gets arbitrarily close to 1 as h approaches μ.

Wei (2020) studies a more interesting scenario where the receiver's outside option is *above* the prior belief about a single sender: $\lambda \in (\mu, 1)$. Allowing for any posterior-separable attention cost function, he shows that a sender-preferred equilibrium exists, and in such an equilibrium the receiver processes less information (in the Blackwell sense) than she would in her first-best. In fact, it can easily be shown that the "sender-preferred" qualifier may be omitted.[20] That is, a sender who does not face competition in information provision restricts (in any equilibrium) the receiver's learning, keeping her payoff strictly below the first-best level.

This result may be explained as follows. If the receiver never garbles the sender's distribution, then this is a canonical Bayesian persuasion problem and we know that the sender-optimal distribution has support $\{0, \lambda\}$. When the receiver garbles due to attention costs, this distribution is no longer sender-optimal since the receiver would garble it to the uninformative distribution.[21] Consequently, the sender must provide the receiver with more information–in particular, he must allow her to generate beliefs strictly above λ. Despite this, Wei (2020)'s results tell us that the sender need not go to the extent of providing as much information as the receiver would choose in her first-best. It is optimal for the sender to place an upper bound on the receiver's posterior belief, which is above λ but is a binding constraint for her.

4.2 No attention costs ($k = 0$)

We now return to our model with competitive information provision but suppose that $k = 0$, so that attention is costless for the receiver. Here, the receiver does not garble the senders' experiments; rather, she processes all information made available by the senders. Our discussion below draws upon earlier literature that has analyzed this model, and that we have discussed in our literature review.

Clearly, here the receiver attains her first-best payoff if and only if she learns at least one

[20] This is a consequence of the fact that the receiver has a unique optimal garbling in response to any binary distribution, as we show.

[21] The support of any feasible garbling lies on $[0, \lambda]$, but the outside option is weakly preferred at all of these beliefs. Thus, there is no benefit of learning, while there is a cost.

sender's type with certainty. We argue that there cannot be an equilibrium where either sender offers full information. For the sake of contradiction, suppose that Sender 1 offers full information, and recall our assumption that in case of ties the receiver selects each sender with equal probability. A bit of reflection reveals that in fact a best response for Sender 2 does not exist: he would like to choose the distribution $\{\epsilon, 1\}$ with $\epsilon > 0$ arbitrarily close to 0. At a realization of ϵ, Sender 2's probability of being selected is $1 - \mu$. On the other hand, if ϵ were chosen to be equal to 0 then this probability would be $\frac{1-\mu}{2}$. Evidently, the distribution $\{\epsilon, 1\}$ with a small, positive ϵ yields a higher expected probability of being selected than $\{0, 1\}$.

The qualitative differences that arise in the presence of attention costs may be appreciated by observing that given $k > 0$, the receiver's behavior following a posterior of 0 is not different from her behavior following a small, positive posterior ϵ: in either case she is sure enough about the type of that sender that she does not wish to learn anything about the other sender.

The game with $k = 0$ has a unique equilibrium as described in the claim below. The proof can be found in Boleslavsky and Cotton (2015) and is omitted here.

Claim 4.1 (Unique equilibrium without attention costs)**.** *Let $k = 0$. Then there is a unique equilibrium in which:*

1. *Each sender chooses the uniform distribution on $[0, 2\mu]$ if $\mu \leq \frac{1}{2}$.*
2. *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 if $\mu > \frac{1}{2}$.*

5 Welfare Effects of Costly Attention

We now explore an intriguing possibility that emerges from our analysis. We have seen that with positive attention costs the receiver might be able to elicit full information from both senders in equilibrium, although she does not process all of it. On the other hand, without attention costs, there is a unique equilibrium where senders do not provide full information, but where all available information is processed. One might wonder, then, if it is possible that the receiver ends up basing her decision on better quality information when she *is* subject to attention costs than when she is not. Our next result indeed answers this in the affirmative.

Proposition 5.1 (Welfare effects of attention costs)**.** *Suppose that $k > 0$ and parameters are such that a full information equilibrium exists. Then, the ex ante probability that the better sender is eventually selected by the receiver in this equilibrium is strictly higher than* *the corresponding probability in the unique equilibrium under $k = 0$ if and only if k is below a threshold $\overline{k}(\mu)$.*

To understand why the inequality reverses when $k > \overline{k}(\mu)$, note that even as we hold the full information equilibrium constant, the amount of information the receiver chooses to process in this equilibrium diminishes as k grows, so that beyond a certain point the receiver does better with the limited amount of information provided under $k = 0$.

6 Extensions

Here, we illustrate that our main result on existence of a full information equilibrium is robust to different modeling choices. In each case, familiar lines of reasoning can be invoked to see that there is nothing sacrosanct about full information *per se*: there is a continuum of equilibria that are outcome equivalent.

6.1 *Ex ante* heterogeneity

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 extends to settings where the prior beliefs on the two senders are different but sufficiently close. Setting $k = 1$ for expositional convenience, we have the following result, whose proof follows steps analogous to the one for the homogeneous means case.

Proposition 6.1. *There is an equilibrium in which each sender offers full information 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* $i \neq j$.

6.2 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 sender is, the costlier it is for the receiver 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 (at Sender 1) by

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

where p_1 is Sender 1's chosen distribution and q_1 is the receiver's chosen garbling. 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 positive.

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 6.2. *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 senders 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. ■

6.3 Non-binary types

The spirit of our main result also extends to a scenario in which each sender's type is an i.i.d. real-valued random variable ω_i that is distributed on some subset of the real line according to cdf F. We assume that ω_i has a finite mean, μ, and that F is not the degenerate mass point on μ (or else all results would be trivial). Furthermore, we specify that the receiver is risk-neutral, so that she only cares about the distribution of posterior means. If each sender offers full information, the receiver's choice of distribution of posterior means is equivalent (see e.g. Kolotilin (2018)) to a choice of mean-preserving contraction (MPC) $G \in \mathcal{M}(F)$ of the prior F. Information remains costly, and the receiver incurs a cost according to the cost functional $C : \mathcal{M}(F) \to \mathbb{R}_+$. Mirroring our main specification, we assume a specific quadratic functional form:

$$C(G) = k \int (x - \mu)^2 \, dG.$$

Now let us think back to the main specification, and recall that the receiver's first-best information acquisition where she acquires the distribution with support $\left\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\right\}$ at the first sender and the degenerate distribution at the second is such that she acquires strictly less information at the first sender as k increases. Moreover, it is easy to see that for all k sufficiently high, this distribution is an MPC of F and must be optimal, since it solves a problem with fewer constraints.[22] Consequently, an analog of Proposition 3.2 holds

Proposition 6.3. *If k is sufficiently high, there is an equilibrium in which both senders offer full information.*

7 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 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, there are equilibria where the senders offer *at least* as much information to the receiver as she would choose if she herself could control information provision. All such equilibria are outcome equivalent, and there is no symmetric binary equilibrium that leads to a different distribution of outcomes. Moreover, we find that the receiver may base her final decision on better quality information when she is subject to attention costs than when she is not.

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 type of one sender partially substitutes for information on the type of the other. 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 about the *other* sender. The deviating sender

[22] As k goes to infinity, the distribution over posteriors at the first sender converges to the degenerate mass at μ, which is by assumption a strict MPC of F.

thus has limited ability–if any–to affect the overall quality of the receiver's information across the two alternatives.

Our model can be applied to several other settings beyond our leading example in which pharmaceutical companies strategically disclose information to prescribing physicians. For instance, the receiver could be a buyer sequentially visiting two used car dealerships, taking test drives and gathering information on the features of each alternative. 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.

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.

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. 2021. "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.

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

Eliaz, Kfir and Ran Eilat. 2021. "Collective Information Acquisition." *Working paper*.

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.

Kolotilin, Anton. 2018. "Optimal information disclosure: A linear programming approach." *Theoretical Economics*, 13: 607–635.

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, Filip Matejka, and Mirko Wiederholt. 2021. "Rational inattention: A review."

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

Matyskova, Ludmila and Alphonso Montes. 2021. "Bayesian persuasion with costly information acquisition." *Working paper.*

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

Tsallis, Constantino. 1988. "Possible generalization of Boltzmann-Gibbs statistics." *Journal of Statistical Physics.*

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

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

A Appendix: 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 $x \in [l, h]$. Her stage 2 optimal garbling is either degenerate or binary, and its support is as follows.*

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

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

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

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

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 $l \leq y_1 < x < y_2 \leq h$. By the definition of concavification of a function, we must have[23]

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

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 $\mu < l + \frac{1}{4k}$.*

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

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 $l+k(\mu-l)^2$ and concave everywhere else (weakly on $\left(l+\frac{1}{4k}, \mu+\frac{1}{4k}\right)$). Then, the concavification must join points $z_1 \leq l+k(\mu-l)^2$ and $z_2 > l+k(\mu-l)^2$ (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 A.2.1. *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* $1/2$ *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 3.7

Suppose each sender offers support $\{l, h\}$, with $l \in [0, \mu)$ and $h \in (\mu, 1]$. Let $q \in [0, 1]$ be the probability that on path the receiver visits Sender 1 first.

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. 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. Say 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.

Suppose that a a sender deviates to a different distribution, and that the receiver does not change her order of visits in response. Then the deviating sender's 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, on path the receiver chooses a unique binary garbling at stage 1, and exactly one belief in the support is followed by learning at the second sender.

Denote the stage 1 belief following which the receiver learns 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 the 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 $\left\{h - \sqrt{\frac{h-w}{k}}, h\right\}$. Everything is as in possibility 1, except that l' is chosen such that $h - \sqrt{\frac{h-w}{k}} < l' < w$.

Possibility 3: If $w > \mu$ and is followed by a stage 2 best response $\left\{l, l + \sqrt{\frac{w-l}{k}}\right\}$ or $\{l, h\}$. There must be at least one sender, say sender i, who gets payoff $\leq \frac{1}{2}$ on path. That sender can profitably deviate to the uninformative distribution: in response the receiver would only learn from the other sender as dictated by Lemma A.1 (setting $x = \mu$). It is verified that the deviating sender's payoff is then $> \frac{1}{2}$. ■

A.3 Proof of Proposition 3.1

"Only if": Suppose that there is no equilibrium in which both senders offer full info. Then, Proposition 3.2 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 3.2

See the proof of Proposition 3.7, setting $l = 0, h = 1$.

A.5 Proof of Lemma 3.3

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

A.6 Proof of Lemma 3.4

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

A.7 Proof of Lemma 3.5

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

A.8 Proof of Proposition 3.6

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 $\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\}$ at stage 1 and to learn nothing at stage 2.

Suppose sender i offers a distribution of which $\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\}$ 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 $\{\mu - \frac{1}{4k}, \mu + \frac{1}{4k}\}$ for him. Exactly as in the proof for existence of a full information equilibrium (Proposition 3.7 for $h = 1, l = 0$), Lemma A.3 can be used to argue that the deviation cannot be profitable. ■

A.9 Proof of Proposition 5.1

Proof. When $k = 0$, Claim 4.1 tells us that for $\mu \leq \frac{1}{2}$ the expected type of the selected sender is $\frac{4}{3}\mu$, and for $\mu > \frac{1}{2}$ this is

$$\frac{4}{3}(1-\mu)\left(\frac{1}{\mu}-1\right)^2 + \left[1 - \left(\frac{1}{\mu}-1\right)^2\right].$$

Suppose that $k > 1/2$ and $\mu \in \left[\frac{1}{4k}, 1 - \frac{1}{4k}\right]$. The expected type of the selected sender in the full information equilibrium is

$$\frac{1}{2}\mu + \frac{1}{2}\left(\mu + \frac{1}{4k}\right).$$

Straightforward algebra then yields the result. ■

A.10 Proof of Proposition 6.1

Suppose that both senders offer full information. We first determine the receiver's best response. Suppose that Sender 2 is visited second (we'll revisit this assumption shortly).

Lemma A.4 (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* $\mu_1 \in \left[\frac{1}{4}, \mu_2 + \frac{1}{4}\right]$.

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

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 A.4 when Sender 2 is visited second, then they satisfy one of the conditions for Lemma A.4 when Sender 2 is visited first. Formally,

Lemma A.5. *One of the four parametric restrictions in Lemma A.4 holds if and only if one of the four parametric restrictions in Lemma A.4 holds in which μ_1 and μ_2 are replaced with each other.*

Proof. Let us begin by looking at the parametric conditions given in bullet points 1 and 3 of Lemma A.4. 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. ■

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 A.4, 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 3.2:

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. ■