To Reveal or Not to Reveal: Privacy Preferences and Economic Frictions

Ned Augenblick and Aaron Bodoh-Creed*

May 2017

Abstract

We model two agents who wish to determine if their types match, but who also desire to reveal as little information as possible to non-matching types. For example, firms considering a merger must determine the merger's profitability, but would prefer to keep their information private if the deal fails. In the model, agents with different traits reveal information to a potential partner to determine if they share the same type, but face a penalty depending on the accuracy of their partner's posterior beliefs. With limited signaling, there is universally-preferred dynamic communication protocol in which traits are sequentially revealed depending on sensitivity of the trait. Interestingly, the rarity of an agent's traits plays no role due to the balance of opposing effects: although revealing a rare trait reveals more information immediately, it also screens more partners from later learning information about other traits. When more complex signaling is allowed, agents prefer an equilibrium in which they signal membership of a small but diverse group, which screens many partners while revealing little information about the agent's individual traits.

Keywords: Privacy; Dynamic Communication; Asymmetric Information
JEL Classification Numbers: D01, D82, D83

*Contact Address: 545 Student Services Builiding #1900, Berkeley, CA 94720, contact email: *ned@haas.berkeley.edu*. The authors are grateful to Ben Hermalin, Michael Katz, and Justin Rao.

1 Introduction

In many strategic situations, agents want to identify partners with whom they can profitably interact, but they are also concerned about revealing sensitive information to other parties. For example, an entrepreneur needs to reveal information about her business plan, technology, and market data to venture capitalists (VCs) in order to solicit financing, but is concerned that a VC could pass the information to another company in the VC's portfolio. Firms may need to reveal information to determine if a potential merger or joint venture is profitable, but the firms are also concerned with the economic losses that could result from revealing information to potential trading partners. A political dissident might want to discover if a new acquaintance shares the same political beliefs, but she does not want to reveal subversive views if that acquaintance is a member of the ruling party. Finally, in social situations, a person with unusual preferences might like to find out if an acquaintance shares those preferences, but worries about revealing her type to someone with different tastes.

In this paper, we present a model of dynamic communication between agents who share the goal of identifying partners with matching traits in order to engage in profitable interactions (which we call *matches*), but who face a penalty for revealing information about their traits. While agents must end up revealing all information about themselves to find a match, they can avoid revealing all of their information to all of their potential partners by dynamically screening out non-matching partners over time using equilibrium communication structures that we call *conversations*. Our main result proves that, given limited signaling, all agent types have the same most-preferred conversational structure, and we provide a characterization of how information is revealed over the course of an optimal conversation.

In the model, a sender (she) and a receiver (he) have private information about their own type, a set of binary traits. For expositional purposes, we focus on the strategic concerns of the sender. In each stage of our game, the sender issues a verifiable message about some of her traits to the receiver. After each message, the receiver updates his beliefs about the sender's type based on the traits verifiably revealed in the message (direct information) and the choice of which traits to verifiably reveal (indirect information). If the receiver realizes that he does not share the sender's type, he ends the conversation as mismatching is costly. The players receive a positive *match value* if they share all of the same trait realizations and match.

Crucially, the sender receives an *information penalty* at the end of the game. The

information penalty for a particular trait is the product of (1) the trait's exogenous *information value* or sensitivity and (2) the receiver's knowledge about the sender as represented by the receiver's posterior belief of the likelihood of the sender's true trait realization. For example, if the sender actually has a realization of 0 for the fifth trait and the receiver believes with high (low) probability that the sender has a realization of 0 for the fifth trait, then the sender suffers a relatively high (low) information penalty. Therefore, revealing a rare realization of a high information value trait increases a player's information penalty more than revealing a common realization of a low information value trait.

We start by focusing on equilibria in which all sender-types reveal groups of traits in a common order. First, we show that all orderings can be sustained in equilibrium as long as the match value is large enough. We then show that all sender-types prefer conversations that are *efficient* (do not include non-informative stages), *complete* (fully reveal all traits by the end of the conversation), *sequential* (reveal traits one-by-one), and *ordered by information value* (the traits are revealed in order of increasing information value). The first three properties arise from the sender's desire to eliminate the possibility of matching with a different type of receiver while revealing as little information about her own type in the process.

The fourth property implies that the rarity of a trait does not affect the order in which traits ought to be revealed. The result arises from two opposing incentives. First, revealing a rare trait results in a large immediate increase in the sender's information penalty because it leads to a large change in the beliefs of every receiver-type in the conversation. Second, the release of a rare trait is also a powerful tool for screening out non-matching receivers, who will learn no additional information about the sender in later stages. Given rational expectations, the change in the receiver's beliefs from learning a sender's specific trait must equal to the percentage of the population who does not have that trait (and are then screened from the remainder of the conversation when that trait is revealed). As the sender's information penalty is bilinear in the receiver's beliefs in our baseline specification, these opposing effects balance.

We then discuss the effect of allowing different sender-types to use different conversation structures, which opens the possibility of the sender signaling information about her type through the choice of which traits to reveal at each stage of the conversation. We explore sender preferences over a restricted set of equilibrium conversation structures that allow some forms of signaling to better understand whether the uniform

sender preference for a particular grammar is robust to the presence of some signaling possibilities. Given this restriction, senders continue to have a universal preference for the common-order equilibrium discussed above. Intuitively, senders prefer slow revelation of information, and the use of different orderings unnecessarily signals additional information at each stage.

The uniform sender-preference for a common-order equilibrium breaks down when we allow any sender-type to use any conversation structure. Specifically, we provide an example in which senders – through their choice of conversation structure – signal that they belong to a *small* but *diverse* group of sender-types, which screens out many non-matching receivers without revealing much information about the sender's individual traits. While the equilibrium is complex, the example suggests that very thoughtful coordination can allow senders to do better than the baseline common-ordering equilibrium.

In the Appendix, we provide an additional result and examine the consequences of five modifications to our model. The result shows that as the number of traits grow, many common-grammar equilibria approach the same payoff because conversations mostly end before many traits are revealed. The first extension allows for a profitable match when agents have identical realizations of some, but not all, traits, and we show that all sender-types continue to prefer the equilibrium where traits are revealed in order of increasing information value. In our second extension, we partially characterize the solution to our model when senders incur significant costs to send messages. As messaging costs increase, senders prefer to reveal multiple traits in a single message, with message sizes growing (almost) exponentially as the conversation proceeds. In the third extension, we study how altering the information penalty function affects the balance between the screening and privacy incentives, and we show that extreme imbalances between the privacy and screening incentives can result in different sender-types preferring to reveal traits in different orders. In the fourth extension we allow the sender's messages to be cheap-talk. This opens up possibilities for new kinds of deviations, and we discuss when and how these deviations may alter the equilibria discussed in the paper. The final extension of our model concerns the case where the agents can swap the roles of sender and receiver exogenously. We prove that the sender-optimal equilibrium derived for the benchmark model remains the unique sender-optimal equilibrium when we allow the sender to break equilibria that are bad for the sender in a way we make precise in the discussion.

There is, of course, a large literature on the effects of asymmetric information. The friction in this model is distinct from more standard issues caused by asymmetric information. For example, in moral hazard and adverse selection models, agents are vertically differentiated and unable to directly reveal their types, which leads to frictions because "low" types try to mimic "high" types. In cheap talk games, frictions arise because senders cannot verifiably reveal information and the sender and receiver differ in their preferences over outcomes. In our model, agents are horizontally differentiated, agree on optimal outcomes, and are able to verifiably reveal information. The economic friction arises in our model because agents simply prefer to reveal less information about their own type to other agents.

There is a related literature on the structure of dynamic communication in different situations, such as firms desiring to retain ideas for private use (Stein [37] and Ganglmair and Tarantino [12]), an agent who desires to gather information of value to another agent (Hörner and Skrzypacz [20]), an agent who can communicate private information to another agent through a series of messages (Aumann and Hart [2] and [1]), and two agents that can exchange information relevant to the others' decision problem (Rosenberg, Solan, and Vieille [33]). Other work formalizes notions of communication complexity and studies the economic consequences (Green and Laffont [17]; Melamud, Mookherjee, and Reichelstein [26]; McAfee [29]; Blumrosen, Nisan, and Segal [6]; Fadel and Segal [11]; Kos [22] and [23]; Blumrosen and Feldman [5]; Mookherjee and Tsumagari [31]). Our model is distinguished from these works by our inclusion of a privacy preference directly in the agents' utility function and our focus on the preferences of all sender-types on both the kind and volume of information conveyed in each period.[1]

The closest paper to our work is Dziuda and Gradwohl [10], who contemporaneously developed a model of the exchange of informative messages that focuses on screening between vertically differentiated types - one productive ("viable") type and many unproductive ("unviable") partner types.[2] The model fixes the order in which

[1]The paper also has links to the computer science literature on *differential privacy*, which focuses on the amount of information revealed by an algorithm (see the survey by Dwork [9]), including incorporating bounds on differential privacy into mechanism design problems (e.g., Gradwohl [16], Nissim, Orlandi, and Smorodinsky [32], Chen et al. [7], Xiao [40]). Unlike this literature, we assume that the sender has preferences over the receiver's knowledge of her type. This aspect of our work has some similarities to the literature on psychological games (Geanakoplos et al. [13]).

[2]The main results provide conditions under which the joint surplus of viable types is maximized by one of two potential equilibrium communication protocols. In the first, players engage in a dynamic communication protocol with the amount of information revealed in each stage determined by the distribution of unviable types and the cost of revealing additional pieces of information. In the

information is revealed, which allows more general results on the equilibrium set than we can provide. However, the fixed order obviously prevents them from studying the order in which information is revealed, which is our primary interest. In their model, the information of viable agents has an "intrinsic value" (an increase of the utility of all types of receiver immediately upon receipt) and an "extrinsic value" (an increase in the expected utility of matching as communication proceeds successfully). Our model shares the idea of extrinsic value, but there is no clean analog to intrinsic value since agents are horizontally differentiated and therefore all agents are viable given a matching receiver.

We start by outlining the theoretical model in Section 2. In Section 3 we analyze the sender-types' preferences over the set of equilibria wherein all agents reveal traits in the same order, and in Section 4 we argue that our results extend to a class of equilibria that allow different sender-types to reveal traits in different orders, and then show an example with more complex signaling. Finally, we conclude in Section 5. All proofs are contained in Appendix A.

2 Model

There are two players, a sender and receiver, indexed by $i \in \{S, R\}$.[3] The payoff-relevant characteristics of the players are defined by N binary traits, so agent types are drawn from the set $\Omega = \{0, 1\}^N$. Our focus on binary traits is solely for expositional purposes - it is easy to extend our results to the case where each trait can assume any of a finite number of values. A generic agent type is denoted $\omega \in \Omega$ with the j^{th} trait denoted $\omega_j \in \{0, 1\}$. The probability that trait j has a realized value of 1 is ρ_j and the realization of each trait is stochastically independent. For example, the probability that $\omega = [1\ 0\ 1]$ is realized is denoted $\Pr(\omega) = \rho_1 \cdot (1 - \rho_2) \cdot \rho_3$. We assume as a convention that $\rho_j \in [\frac{1}{2}, 1)$, so $\omega_j = 1$ is the common trait realization and $\omega_j = 0$ the rare trait realization. Therefore, high values of ρ_j increase the rarity of $\omega_j = 0$ as well as the homogeneity of the population with respect to that trait. We denote player i's type as $\omega^i \in \Omega$. Initially neither player has any information regarding the type of the

second, one player reveals all of his information in the first stage. This second equilibrium can be jointly optimal (even though the first player may receive a low payoff) if the marginal cost of revealing information falls relative to the marginal benefit of screening out more unviable types as the message size grows.

[3] In Section B.6 we extend the model to the case where the agents take turns playing the role of sender and receiver or have these roles switch in an exogenous, random fashion.

other party, but the values ρ_j are common knowledge.

The sender reveals messages about her traits over multiple *stages,* indexed by $t \in \{1, 2, ..., T\}$ with $T \gg N$. In each stage, the sender reveals a message of the form $m \in \{\varnothing, 0, 1\}^N$, where (for example) $m = (\varnothing, \varnothing, 1, \varnothing, 0)$ is a decision to reveal $\omega_3 = 1$ and $\omega_5 = 0$ to the receiver. As shorthand, we denote a message by its revealed traits, such as $\{\omega_3 = 1, \omega_5 = 0\}$. We assume that these messages are verifiable and cannot contain false information.[4,5] We call this dynamic exchange a *conversation.*

In many of the real-life environments, the sender could choose to not show up to a conversation in the first place. To explicitly model this decision, we allow the sender to issue a *Not Attend* message in the first stage, which ends the game. Sending any other message is labeled as *Attending* and equilibrium with all sender-types attending is called a *full-participation equilibrium.*

In each stage the receiver chooses to *Continue* the conversation, *End* the conversation, or *Match* with the sender after receiving the sender's message. If the receiver decides to continue the conversation, the game proceeds to the next stage. If the receiver chooses to end the conversation, the game ends and both agents get a match payoff of 0. Finally, if the receiver chooses Match, the game ends, the sender's type is revealed to the receiver, and both agents receive a match payoff of $M > 0$ if the sender and receiver share the same type and $-L < 0$ otherwise.

The sender's strategy σ is a mapping from a history h_{t-1} to a set of traits that will be revealed in stage t of the game for all possible t and histories h_{t-1}, where h_0 denotes the beginning of the game. To more concisely represent a sender-type's messages in equilibrium given a strategy σ, define $\widetilde{m}_1$ as $\sigma(h_0)$ and recursively define $\widetilde{m}_t$ as $\sigma(\widetilde{m}_1, \widetilde{m}_2, ...\widetilde{m}_{t-1})$ given that the game has continued. Define a sender-type's *grammar* g in an equilibrium as the sequence of sets of *traits* revealed in each stage. For example if type $\omega =$ [1 1 0 1] issues messages $\widetilde{m}_1 = \{\omega_1 = 0, \omega_4 = 1\}$, $\widetilde{m}_2 = \{\}$, and $\widetilde{m}_3 = \{\omega_2 = 1, \omega_3 = 0\}$, then $g = (\{1, 4\}, \{\}, \{2, 3\})$ is the grammar for that type. Intuitively, a grammar captures the sequence of traits revealed by the sender in equilibrium, but does not describe the realizations of those traits. Therefore, senders of different types can follow the same grammar but send different messages because they have different trait realizations. We will largely not discuss the receiver's strategy, which is a mapping from past histories to the three possible receiver actions.

[4]We consider the case of cheap talk messages in appendix B.5.

[5]Some papers refer to this information as *hard information.* The salient feature is that the information cannot be falsified, not that it can be confirmed by a third party enforcer (e.g., a court).

We use perfect Bayesian equilibria (PBE) as our solution concept. In every PBE, each history h is associated with the beliefs of the receiver regarding the sender's traits, which are determined by Bayes Rule where possible. We define $\mu(\omega_j^S = 1|h)$ as the equilibrium belief held by the receiver at history h that the sender's realization of trait j is equal to 1. Finally, we define $\mu(\omega_j^S = \omega_j^{S*}|h)$ as the probability that the receiver places on the sender's true realization of trait j, where we use the notation ω_j^{S*} to emphasize that ω_j^{S*} is the true realization known to the sender. For example, if $\omega^S = [1\ 0]$ and $\mu(\omega_1^S = 1) = .2$ and $\mu(\omega_2^S = 1) = .2$, then $\mu(\omega_1^S = \omega_1^{S*}) = .2$ and $\mu(\omega_2^S = \omega_2^{S*}) = .8$. $\mu(\omega_j^S = \omega_j^{S*}|h)$ describes the amount of knowledge the receiver has about the sender's j^{th} trait.

The novel component of our model is that senders have direct preferences over the information that they reveal.[6] Specifically, we assume that a sender of type ω suffers an *information penalty* of the following form if the game ends at history h:

$$\text{Information penalty at } h\text{: } -\sum_{j=1}^{N} \mu(\omega_j^S = \omega_j^{S*}|h) \cdot v_j \tag{2.1}$$

where $v_j \geq 0$ is an exogenous *information value* assigned to trait j. For expositional purposes, we adopt the trait labeling convention that $v_{j+1} \geq v_j$.[7] Information penalties enter utility additively with match payoffs ($M, -L$, or 0). Note that we assume that the information penalty applies even in the event that the agents match. This allows our formulas to be more transparent and better demonstrates the intuitions underlying our result.[8]

In our setup, the information penalty of the sender increases as the receiver places more probability on the sender's true trait realizations. We interpret:

$$\mu(\omega_j^S = \omega_j^{S*}|h) - \mu(\omega_j^S = \omega_j^{S*}|h_0)$$

as the amount of information revealed (i.e., privacy lost) through the conversation at history h about the sender's trait j, with this value multiplied by v_j representing the

[6] In a previous version of the paper, the receiver was modeled as having a preference for privacy like the sender. In addition, the agents switched communication roles in each period. See Appendix B.6 for more details.

[7] For the bulk of the paper, we assume $\mu(\circ|h)$ enters utility linearly, but we explore the effect of nonlinear functions of $\mu(\circ|h)$ in Section B.4.

[8] One alternative is that the sender only suffers the information penalty if her type is different than the receiver's type - our results are completely unchanged under this assumption. A second alternative is that the sender only suffers the information penalty if the receiver does not choose Match, but this would not change our results. We discuss these issues in detail in appendix D.

utility representation of the preference to avoid this privacy loss. In the event that a match does not occur, the sender would clearly prefer the receiver place probability 1 on the sender having different trait realizations than she does (although this is impossible in equilibrium).

The information penalty can be interpreted as an agent's preferences over the beliefs of others, which is plausible in many social settings such as professional networking or dating. In other circumstances, the information penalty is a reduced-form method for capturing the effect of the information on future outcomes. The latter interpretation is more appropriate in the context of bargaining over mergers or other contracts between firms that require the firms to release information about competitive strategy or trade secrets, which can be used by the other agent to reap an advantage in future market competition.[9,10]

Finally, consider two informationally-equivalent grammars, such as $(\{1,4\},\{2,3\})$ versus $(\{1,4\},\{\},\{2,3\})$. These yield the same payoff, but if we had included any cost of communication or time discounting, the first grammar would be strictly preferred to the second. In lieu of adding the notational complexity of such a modification to our model, we assume that both players have a lexicographic preference over grammars that breaks ties between grammars in favor of the grammar that involves fewer stages of communication.

3 Common Grammars

In this section, we consider equilibria in which all sender-types use a common grammar, which removes any signaling of traits through grammar choice. The usage of a common grammar across types could be the result of a social norm, and the simplicity resulting from a lack of signaling may make common-grammar equilibria focal. We first show that, given a high enough match payoff M, any grammar can be sustained in equilibrium. We then discuss sender-types' preferences over these equilibria, culminating with the result that the most-preferred equilibrium is the same for all sender-types.

[9]We do not include the potential advantage from learning about the other agent's type in the player's utility function. Although more complicated, our results would remain qualitatively similar.

[10]We provide a formal microfoundation of the information penalty function in Appendix E

3.1 The Large Set of Equilibria

We start with the result that, for any grammar, there is an equilibrium in which all senders attend the conversation and use that grammar as long as the match payoff is large enough. For this to occur, the match payoff must be large enough so that each sender-type receives a higher payoff from participating in the conversation than deviating and not attending (given the information penalty associated with the receiver's beliefs after observing this deviation). To create the largest penalty from deviation, off-the-path receiver beliefs are assigned such that the receivers infer that the sender- and receiver-types do not match, so the receiver chooses to end the conversation following a sender's deviation.[11]

Proposition 1. *For any grammar g, there exists $M > 0$ such that there is a full participation equilibrium in which all sender-types use grammar g.*

If M is not large enough to sustain a common-grammar equilibrium, there can exist equilibria in which some players choose to not attend the conversation. As a simple example, consider the case in which there are two traits ($N = 2$) with $\rho_1 = 0.75$, $\rho_2 = 0.5$, $v_1 = 1$ and $v_2 = 2$. Given these parameters, an equilibrium with full attendance using the common grammar $(\{1\}, \{2\})$ does exist if $M \geq 8$ and the receiver beliefs following any deviation are $\mu(\omega_1^S = 1|h_{NA}) = 0$ and $\mu(\omega_2^S = 1|h_{NA}) = 0.5$.

However, if $M < 8$, then there is no equilibrium with a common grammar in which all players attend the conversation. There is, however, an equilibrium in which sender-types [0 1] and [0 0] choose to Not Attend, the receiver correctly infers the sender-type's first trait after observing Not Attend, and sender-types [1 1] and [1 0] participate in the conversation using grammar $(\{1\}, \{2\})$. In this equilibrium, sender-types with a rare first-trait realization choose to Not Attend the conversation and forgo

[11] As noted, our proof constructs PBE wherein the receiver's beliefs following a deviation by the sender depend on the receiver's type, which violates the *consistency* condition imposed on beliefs in a sequential equilibrium (SE). However, our focus on PBE strengthens our main results, Propositions 2 and 3, which find the best equilibrium out of a subset of PBE, only a subset of which can be SE. In addition, one can support the common grammar equilibrium that these propositions identify as optimal using beliefs that satisfy the SE consistency requirement for sufficiently large M. To see this, suppose that all agents believed that deviating senders have a realization of 0 for all of the traits that had not previously been revealed verifiably, which are beliefs that satisfy the consistency requirement. A sender that did not share these trait realizations would refuse to deviate from the common grammar equilibrium (for large enough M) since she would lose the opportunity to match profitably. A sender with those trait realizations would refuse to deviate as she would be revealing information to more receiver types than in the common grammar equilibrium, resulting in a higher information penalty than is received in the common grammar equilibrium.

any matching payoff in order to avoid revealing information about their type. Note that any equilibrium of this type requires multiple sender-types to choose Not Attend because otherwise the receiver could infer the full type of the sender.

In the next section, we turn to sender preferences over common-grammar equilibria. Consequently, we continue the analysis under a maintained assumption which guarantees the existence of at least one common-grammar equilibrium: $M \geq \sum_{j=1}^{N} \frac{\rho_i}{\prod_{k=j}^{n}(1-\rho_k)} v_i$ ($M \geq 8$ in the previous example). However, we imagine that non-participation equilibria might be common in real-life, with rare types avoiding information-revealing conversations and never finding matches because (1) the likelihood of finding a match is low for rare types and (2) the differential information costs of revelation are higher for rare types.

3.2 Universally-Preferred Equilibrium

While there are many equilibria given a high enough match value, we now show that all sender-types prefer a unique common-grammar equilibrium. As an initial step, we define a series of suboptimal grammars, and the natural "improvement steps" one might use to find a more optimal grammar.

Definition 1. *1) Grammar g is **inefficient** if any stage contains a null message or a trait revealed in a previous stage. The **efficient** version of grammar g is a grammar g' that is identical to grammar g except where stages containing null messages or repeated traits are removed.*

*2) Grammar g is **incomplete** if it does not reveal all traits. The **complete** version of incomplete grammar g is a grammar g' that is identical to grammar g except that a final stage is appended in which all unrevealed traits in grammar g are revealed.*

*3) Grammar g is **non-sequential** if multiple traits are revealed in one stage. The **sequential** version of non-sequential grammar g is a grammar g' in which each stage that reveals multiple traits is replaced by individual stages that each reveal a single trait.*

*4) Grammar g is **misordered** if the grammar is sequential and traits are not revealed in order of increasing information value. The **ordered** version of grammar g is a grammar g' that reveals traits in order of increasing information value.*

Given this, we present one of our main results, which implies that our improvement steps are universally preferred by all sender-types.

Proposition 2. *Suppose M is sufficiently large that a common grammar entailing participation by the sender exists. Among all common-grammar equilibria, all sender-types prefer the equilibrium using:*

1) the efficient version of grammar g to the inefficient grammar g.

2) the complete version of grammar g to an incomplete grammar g.

3) the sequential version of grammar g to a non-sequential grammar g.

4) the ordered version of grammar g to the misordered grammar g.

The first three results of Proposition 2 are straightforward. Claim (1) follows immediately from our assumption that senders have a lexicographic preference for shorter conversations ceteris paribus. Claim (2) notes a preference for conversations in which the types are fully revealed before a match occurs. In an equilibrium employing an incomplete grammar, receiver-types choose Match without knowing the sender's full type because the receiver (correctly) believes that the sender will send no more messages. Meanwhile, the sender sends no more messages because she (correctly) believes that the receiver will not choose Match if she reveals any more messages. However, the sender would prefer to issue a final message revealing all of her traits because it avoids the negative payoff from mismatching ($-L < 0$).[12,13] Claim (3) implies that senders prefer equilibria wherein messages do not reveal multiple traits at once. If the sender reveals multiple traits and the receiver does not match on one of these, the sender has revealed more information than required and unnecessarily increased the size of her information penalty. Claims (1) - (3) suggest that senders prefer the conversation in which, in each stage, the sender reveals the smallest amount of information that will screen out at least some receivers until her full type is revealed.

Claim (4) is more surprising. A natural (and incorrect) conjecture is that players prefer to reveal traits in an order that minimizes the immediate increase in their information penalty, implying that the preferred trait revelation order is dependent on

[12]In any equilibria with matching that occurs when *multiple* traits are unrevealed, the information penalty also pushes a sender to prefer an additional stage that reveals more traits, because this reduces the number of mismatched receiver-types who choose to match and consequently observe all of the sender's traits. However, when there is only trait left to reveal, the remaining receiver-types will learn all traits of the sender regardless. Therefore, the mismatch penalty is necessary to break the indifference of the sender between complete sequential revelation and an equilibrium where all traits *but one* are revealed and then the receiver always chooses to match.

[13]There are certainly situations in which the sender might prefer an incomplete grammar. For example, imagine a third-party eavesdropper who can listen to the conversation, but cannot observe the process once a match is initiated. Given privacy preferences over the eavesdropper's beliefs, the sender prefers to gradually reveal traits publicly until a critical stage (prior to the revelation of all traits) where she prefers a match to be initiated to cutoff revelation to the eavesdropper.

both the information value of a trait and the rarity of an agent's particular trait realization. For example, suppose $\rho_1 = 0.6$ and $\rho_2 = 0.9$, with v_1 only slightly smaller than v_2. Our (false) conjecture would suggest that sender-type [1 1] would prefer grammar $(\{2\}, \{1\})$ to $(\{1\}, \{2\})$ because it reveals less information, shifting the mismatched receiver-types' prior from 0.9 to 1 on the second trait (rather than from 0.6 to 1 on the first). Of course, if this conjecture were correct, sender-type [0 0] would have the opposite preference, and more generally, senders of different types would not agree on the optimal grammar.

However, this logic ignores a dynamic benefit: revealing a rare trait realization ends the conversation earlier for more receiver-types, which reduces the information senders reveal in later stages to non-matching partners. For example, consider the full benefit of the two grammars for sender type [1 1] relative to the situation in which all receivers observe her full type. By revealing the second trait first, she stops .1 of the population (who have a different second trait) from learning her first trait, which would have moved their prior for that trait by .4. Alternatively, initially revealing the first trait stops .4 of the population from learning the second trait, which would have moved the prior by .1. Therefore, the population-times-movement-in-prior savings are the same for both grammars. The effects balance because the sender's information penalty is linear in the receiver's beliefs and, given that the receiver has rational expectations, the movement in beliefs from learning that the sender holds a trait must be equal to the percentage of the population that does not hold that trait. However, because $v_2 > v_1$, the information penalty puts more weight on the latter case, leading the sender to prefer the grammar $(\{1\}, \{2\})$. The proof of Proposition 2 uses this logic to demonstrate that all sender-types prefer the ordering in which traits with higher information values are revealed later.

Given Proposition 2, the structure of the grammar universally preferred by all sender-types is obvious.

Corollary 1. *Among all common-grammar equilibria, all sender-types prefer the one with the efficient, complete, and sequential-ordered grammar* $g = (\{1\}, \{2\}, ..., \{N\})$

Now that we have derived the optimal grammar, we present Example 1 to show how a conversation unfolds over time, using a format designed to make all of the possible paths-of-play explicit (although only describing the potential histories of play for a single type).

Example 2: (Three Trait Conversation)

Assume that N=3 and that ρ_1=.8, ρ_2=.6, and ρ_3=.7.
Focus on sender-type [110]. Consider sender-optimal conversation:

Stage 1

Sender	Message	Reciever (Potential Types)	Reciever Inference (About Sender)	Reciever Response	Sender Payoff
	"I am a 1.."	000	100,101,110, or 111	Leave	$-v_1-.6v_2-.3v_3$
	"I am a 1.."	001	100,101,110, or 111	Leave	$-v_1-.6v_2-.3v_3$
	"I am a 1.."	010	100,101,110, or 111	Leave	$-v_1-.6v_2-.3v_3$
	"I am a 1.."	011	100,101,110, or 111	Leave	$-v_1-.6v_2-.3v_3$
	"I am a 1.."	100	100,101,110, or 111	Confirm	[Game Continues]
	"I am a 1.."	101	100,101,110, or 111	Confirm	[Game Continues]
110	"I am a 1.."	110	100,101,110, or 111	Confirm	[Game Continues]
	"I am a 1.."	111	100,101,110, or 111	Confirm	[Game Continues]

Stage 2

Sender	Message	Reciever (Potential Types)	Reciever Inference (About Sender)	Reciever Response	Sender Payoff
		000	100,101,110, or 111		$-v_1-.6v_2-.3v_3$
		001	100,101,110, or 111		$-v_1-.6v_2-.3v_3$
		010	100,101,110, or 111		$-v_1-.6v_2-.3v_3$
		011	100,101,110, or 111		$-v_1-.6v_2-.3v_3$
	"I am a .1."	100	110 or 111	Leave	$-v_1- \ v_2-.3v_3$
	"I am a .1."	101	110 or 111	Leave	$-v_1- \ v_2-.3v_3$
110	"I am a .1."	110	110 or 111	Confirm	[Game Continues]
	"I am a .1."	111	110 or 111	Confirm	[Game Continues]

Stage 3

Sender	Message	Reciever (Potential Types)	Reciever Inference (About Sender)	Reciever Response	Sender Payoff
		000	100,101,110, or 111		$-v_1-.6v_2-.3v_3$
		001	100,101,110, or 111		$-v_1-.6v_2-.3v_3$
		010	100,101,110, or 111		$-v_1-.6v_2-.3v_3$
		011	100,101,110, or 111		$-v_1-.6v_2-.3v_3$
		100	110 or 111		$-v_1 \ -v_2-.3v_3$
		101	110 or 111		$-v_1 \ -v_2-.3v_3$
110	"I am a ..0"	110	110	Match	$M-v_1 \ -v_2 \ -v_3$
	"I am a ..0"	111	110	Leave	$-v_1 \ -v_2 \ -v_3$

Expected payoff to sender-type [110]: $.144M - v_1 - .92v_2 - .636v_3$
Expected payoff, static communication: $.144M - v_1 - v_2 - v_3$

4 Type-Specific Grammars: Indirect Information

In the previous section we focused on equilibria in which all types of senders use the same grammar. If senders instead use different grammars, then receivers will infer information (in equilibrium) about traits that are not explicitly revealed based on the chosen grammar. For example, consider an equilibrium in which there are two traits and sender-types [1,0] and [0,1] use grammar $(\{1\},\{2\})$ while sender-types [1,1] and [0,0] use grammar $(\{2\},\{1\})$. Although the sender only reveals one trait in the first stage, the receiver infers the sender's full type in equilibrium after this one message: for example, the receiver can infer that the sender's type is [0,1] after observing the message $m_1 = \{\omega_1 = 0\}$, as only that type sends that message in equilibrium. As a second example, consider a different two-trait equilibrium in which sender-types [1,1], [1,0] and [0,1] use grammar $(\{1\},\{2\})$, while sender-type [0,0] uses grammar $(\{2\},\{1\})$. In this case, the receiver infers both traits of sender-types [0,1] and [0,0] after the first message, but only learns the first trait of sender-types [1,1] and [1,0].

Considering equilibria where sender-types use different grammars significantly complicates the analysis, so we proceed in two steps. First, we introduce a refinement that eliminates some types of indirect information revelation and show that senders continue to prefer the previously-discussed optimal common-grammar equilibrium. Second, we allow for arbitrary signaling and provide an example of a non-common-grammar equilibrium that is universally preferred by senders.

4.1 No Type Prefers to Signal About Traits Only

In our first step, we allow sender-types to use different grammars, but restrict the combination of these grammars by only allowing equilibria that satisfy a refinement that we call *block inference*.

Definition 2. *At any history h along the path of play of an equilibrium satisfying block inference, we can define $K(h), U(h) \subset \{1,..,N\}$ that denote the sets of known and unknown traits at history h. We require that $K(h) \cup U(h) = \{1,..,N\}$ and $K(h) \cap U(h) = \varnothing$. For all $j \in K(h)$ we have $\mu_R(\omega_j^S = 1|h) \in \{0,1\}$. The receiver believes all traits within $U(h)$ are distributed independently and $\mu_R(\omega_j^S = 1|h) = \rho_j$.*

Loosely speaking, block inference requires that all indirect information revealed through the use of different grammars be achievable in some equilibrium in which all

information is directly revealed.[14] For example, in the second example of this section, there is a payoff-equivalent equilibrium in which all senders directly reveal all traits. This equilibrium has the form:

Sender-type	First Message	Second Message
[1,1]	$\{\omega_1 = 1\}$	$\{\omega_2 = 1\}$
[1,0]	$\{\omega_1 = 1\}$	$\{\omega_2 = 0\}$
[0,1]	$\{\omega_1 = 0, \omega_2 = 1\}$	N/A
[0,0]	$\{\omega_1 = 0, \omega_2 = 0\}$	N/A

Note that no signaling occurs, and the receiver has the same beliefs about the various sender-types at the close of each period as in the second example.

We pursue the analysis of block inference primarily to illustrate that Proposition 2 does not hinge on restricting ourselves to equilibria wherein agents use the same grammar. Given the relatively simple pattern of posterior beliefs along the equilibrium path, this type of equilibrium might be focal among signaling equilibria.

While restrictive, block inference admits a richer variety of signaling phenomena than prior work, allowing for a wide range of indirect inference in equilibrium.[15] In Appendix C, we show that block inference allows for the majority of traits to be revealed by signaling. An example of an equilibrium that does not satisfy block inference requires more than two traits. Consider a three-trait equilibrium in which types [1,0,0], [1,1,0], and [1,1,1] are the only types that transmit $m_1 = \{\omega_1 = 1\}$. Given this message, the receiver knows that if the sender's second trait is a "0," then the third trait is also a "0." This type of correlational information cannot be transmitted in an equilibrium with only direct revelation of traits.

Proposition 3 shows that senders continue to prefer the common-grammar equilibrium highlighted in Corollary 1 over any equilibria with a non-common grammar that satisfies block inference. The logic follows that of part (3) of Proposition 2: all sender-types prefer not to signal traits indirectly through the choice of grammar in the same

[14] More formally, under strategy σ a set of traits $\sigma(h)$ is revealed at history h. In a block inference equilibrium, an additional set of traits $m_{\text{Signal}}(h)$ is signaled to the receiver. Consider the alternative strategy $\widetilde{\sigma}(h) = \sigma(h) \cup m_{\text{Signal}}(h)$. It is obvious that no additional information is revealed under $\widetilde{\sigma}(h)$ relative to $\sigma(h)$ - the information is just revealed verifiably under $\widetilde{\sigma}$. The subtle difference between $\widetilde{\sigma}$ and σ is that σ allows more deviations, an issue related to the cheap-talk message model discussed in Section B.5.

[15] Past papers restrict the inferences that can be made by, for example, exogenously requiring agents to reveal information in a prespecified order (e.g., Stein [37], Dziuda and Gradwohl [10]).

way that they prefer not to reveal multiple traits directly in one stage. This implies that senders prefer to use a common grammar that eliminates signaling — specifically, the senders all prefer the common-grammar noted in Corollary 1:

Proposition 3. *Among all equilibria satisfying block inference, all sender-types prefer the one with common grammar* $g = (\{1\}, \{2\}, ..., \{N\})$.

4.2 Signalling about Type Without Revealing Traits

Expanding the set of equilibria beyond those satisfying block inference complicates the analysis to the point that it is difficult to make general statements about sender preferences. Instead, we show by example that senders can now universally prefer an equilibrium with a non-common grammar. In this equilibrium – outlined in Example 2 – the first message reveals a large amount of information about the type of the sender, but leads to a low information penalty for the sender because there is little information revealed about the sender's individual traits. For example, the message $\{\omega_2 = 1\}$ reveals that the sender is either type [0 1 0 0] or [1 1 1 1], narrowing down the potential sender-types from 16 types to 2 types. However, the receiver does not update his beliefs about the probabilities of traits one, three, or four.

Example 2 (Non-Block Inference Equilibrium):

Assume that $N = 4$ **and that** $\rho_1 = \rho_2 = \rho_3 = \rho_4 = .5$

Consider the following potential equilibrium behavior:

Types [0 0 1 0], [0 1 0 1],[1 0 1 1],[1 1 0 0]:	$g = (\{1\}, \{2,3,4\})$
Types [0 0 0 1], [1 0 1 0],[0 1 0 0],[1 1 1 1]:	$g = (\{2\}, \{1,3,4\})$
Types [1 1 1 0], [0 0 1 1],[0 0 0 0],[1 1 0 1]:	$g = (\{3\}, \{1,2,4\})$
Types [0 1 1 0], [1 0 0 0],[0 1 1 1],[1 0 0 1]:	$g = (\{4\}, \{1,2,3\})$

Then after the first stage:

If $m_1 = \{\omega_1 = 0\}$,then $\mu(\omega_1^S = 1|m_1) = 0$, $\mu(\omega_j^S = 1|m_1) = .5$ for $j \in \{2,3,4\}$
(as [**0** 0 1 0] and [**0** 1 0 1] are the players that send that message in equilibrium.)

If $m_1 = \{\omega_1 = 1\}$,then $\mu(\omega_1^S = 1|m_1) = 1$, $\mu(\omega_j^S = 1|m_1) = .5$ for $j \in \{2,3,4\}$
(as [**1** 0 1 1] and [**1** 1 0 0] are the players that send that message in equilibrium.)

In general:

If $m_1 = \{\omega_j = k\}$,then $\mu(\omega_j^S = k|m_1) = 1$, $\mu(\omega_l^S = 1|m_1) = .5$ for $l \neq j$

The example is a Pareto optimal equilibrium for the given parameters because it leads to a small information penalty for the sender (revealing one trait realization) in the first round and screens out the majority of receivers (in this case, $\frac{7}{8}$ of receivers).

We conjecture that Pareto optimal equilibria for other parameters share the same basic features: senders group in a way such that the receiver learns a great deal about the sender's type, but not much about the sender's specific trait realizations.

This example also demonstrates the costs and benefits of our particular parameterization of the penalty function, in which privacy is solely a function of marginal beliefs about individual traits (and not about correlation between traits). The benefit of modeling type-as-traits is the ability to analyze the structure of conversation when different pieces of information have different privacy values to the sender. The cost is the circumvention of situations in which correlation information may be additionally important for the sender's privacy. For example, in Example 3 the first message sent does not provide any information about the marginal likelihood of any other of the sender's traits. However, the message does imply that the unrevealed traits are perfectly correlated, which could be seen as a loss of privacy for the sender. Unfortunately, other parameterizations of the penalty function which take correlation into account appear either unwieldy or lack the ability to differentiate information by value and rarity.

5 Conclusion

Our paper analyzes situations in which agents must exchange information to discover whether they can productively match, but the agents have a preference for privacy and would like to reveal as little information about their type as possible. Such a concern for privacy in the context of information exchange is pervasive in business, political, and social settings. Our goal is to provide a realistic model of information exchange in these settings in order to discover how the preference for privacy structures communication in terms of both the quantity and the kind of information disclosed as the conversation progresses.

Focusing on the set of equilibria in which all senders-types use a common trait revelation sequence, we find that there is a universally-prefered conversation structure in which individual traits are revealed sequentially, and more sensitive data are disclosed once the agents have become more confident that a profitable match is possible. The universal preference across types with different rarity of traits is driven by the dynamic nature of the conversation - early disclosures of rare trait realizations are more likely to incur an immediate information penalty, but also screen out many non-matching

receiver-types, which reduces the expected information penalties for traits revealed in later stages. We then show that this preference continues when senders can indirectly signal information about traits by using different grammars. We close our analysis with a brief description of an equilibrium that illustrates some of the possibilities realized when more complex signaling is allowed. In this equilibrium, the sender signals membership in a small but diverse group through her initial message, which causes many receiver-types to leave the conversation without learning much about many of the sender's individual traits.

While our principal goal is to study how preferences for privacy influence the structure and timing of information exchange, future work could incorporates this type of analysis into more general models that endogenize the match value. Privacy is often a concern in mechanism-design settings where the agents may interact later. For example, spectrum auctions usually involve a small number of national companies as bidders that are engaged in competitive interactions that can span decades. Other potential settings include bargaining over merger decisions in models of market competition, policy debates in political economy settings, and principal-agent problems that require information exchange between the actors.

References

[1] Amitai, M. (1996) "Cheap-Talk with Incomplete Information on Both Sides," *mimeo.*

[2] Aumann, R. and S. Hart (2003) "Long Cheap Talk," *Econometrica*, 71 (6), pp. 1619–1660.

[3] Bernheim, B.D. (1994) "A Theory of Conformity," *The Journal of Political Economy*, 102 (5), pp. 841 - 877.

[4] Blume, A. (2000) "Coordination and Learning with a Partial Language," *Journal of Economics Theory*, 95, pp. 1-36.

[5] Blumrosen, L. and M. Feldman (2013) "Mechanism design with a restricted action space," *Games and Economic Behavior*, 82, pp. 424 - 443.

[6] Blumrosen, L.; N. Nisan; and I. Segal (2007) "Auctions with Severely Bounded Communication," *Jounral of Artificial Intelligence Research*, 28, pp. 233 - 266.

[7] Chen, Y.; S. Chong; I.A. Kash; T. Moran; and S. Vadhan (2013) "Truthful Mechanisms for Agents that Value Privacy, " *EC '13*, pp. 215 - 232.

[8] Crawford, V.P. and J. Sobel (1982) "Strategic Information Transmission," *Econometrica*, 50, pp. 1431 - 1451.

[9] Dwork, C. (2008) "Differential Privacy: A Survey of Results," *Theory and Applications of Models of Computation: Lecture Notes in Computer Science*, 4978, pp. 1-19.

[10] Dziuda, W. and R. Gradwohl (2012) "Achieving Coordination Under Privacy Concerns," *mimeo.*

[11] R. Fadel and I. Segal (2009) "The communication cost of selfishness," *Journal of Economic Theory*, 144 (5), pp. 1895 - 1920.

[12] Ganglmair, B. and Tarantino, E. (2013) "Conversation with Secrets," *mimeo.*

[13] Geanakoplos, J.; D. Pearce and E. Stacchetti (1989) "Psychological Games and Sequential Rationality," *Games and Economic Behavior*, 1, pp. 60 - 79.

[14] Ghosh, P. and D. Ray (1996) "Cooperation in Community Interaction without Information Flows," *Review of Economic Studies*, 63, pp. 491-519.

[15] Glazer, J. and A. Rubinstein (2003) "Optimal Rules for Persuasion," *Econometrica*, 72 (6), pp. 1715 - 1736.

[16] Gradwohl, R. (2012) "Privacy in Implementation," *mimeo.*

[17] Green, J. and J-L Laffont (1987) "Limited Communication and Incentive Compatibilit." in *Information, Incentives, and Economic Mechanisms: Essays in Honor of Leonid Hurwicz*, edited by Theodore Groves, Roy Radner, and Stanley Reiter, 308–29. Minneapolis: Univ. Minnesota Press.

[18] Hayek, F. (1945) "The Use of Knowledge in Society," *American Economic Review*, 35 (4), pp. 519 - 530.

[19] Honryo, T. (2011) "Dynamic Persuasion," *mimeo.*

[20] Hörner, J. and A. Skrzypacz (2011) "Selling Information," *mimeo.*

[21] Kamenica, E. and M. Gentzkow (2011) "Bayesian Persuasion," *American Economic Review*, 101, pp. 2590–2615.

[22] Kos, N. (2012) "Communication and efficiency in auctions," *Games and Economic Behavior*, 75, pp. 233 - 249.

[23] Kos, N. (2014) "Asking questions," *Games and Economic Behavior*, 87, pp. 642-650.

[24] Krishna, V. and J. Morgan (2004) "The Art of Conversation, Eliciting Information from Experts through Multi-Stage Communication," *Journal of Economic Theory*, 117 (2), pp. 147-179.

[25] McAfee, R.P. (2002) "Coarse Matching," *Econometrica*, 70 (5), pp. 2025 - 2034.

[26] Melamud, N.; D. Mookherjee; and S. Reichelstein (1992) "A theory of responsibility centers," *Journal of Accounting and Economics*, 15, pp. 445 - 484.

[27] Milgrom, P. (1981) "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, Vol. 12, pp. 380-391.

[28] Milgrom, P. (2008) "What the Seller Won't Tell You: Persuasion and Disclosure in Markets," *The Journal of Economic Perspectives*, 22 (2), pp. 115-132.

[29] Milgrom, P. (2010) "Simplified mechanisms with an application to sponsored-search auction," *Games and Economic Behavior*, 70, pp. 62-70.

[30] Milgrom, P. and J. Roberts (1986) "Relying on the Information of Interested Parties," *The RAND Journal of Economics*, 17 (1), pp. 18-32.

[31] Mookherjee, D. and M. Tsumagari (2014) "Mechanism Design with Communication Constraints," *Journal of Political Economy*, 122 (5), pp. 1094-1129.

[32] Nissim, K.; C. Orlandi; and R. Smorodinsky (2012) "Privacy-Aware Mechanism Design," *EC '12*, pp. 774 - 789.

[33] Rosenberg, D.; E. Solan; and N. Vieille (2013) "Strategic information exchange," *Games and Economic Behavior*, 82, p. 444 - 467.

[34] Rubinstein, A. (1996) "Why Are Certain Properties of Binary Relations Relatively Common in Natural Language?" *Econometrica*, 64, pp. 343 - 355.

[35] Rubinstein, A. (2000) *Economics and Language: Five Essays*, Cambridge University Press: Cambridge.

[36] Sher, I. "Persuasion and Dynamic Communication," *mimeo.*

[37] Stein, J.C. (2008) "Conversations among Competitors," *America Economic Review*, 98, pp. 2150 - 2162.

[38] Van Zandt, Timothy (2007) "Communication Complexity and Mechanism Design," *Journal of the European Economic Association*, 5 (2-3), pp. 543 - 553.

[39] Watson, J. (2002) "Starting Small and Commitment," *Games and Economic Behavior*, 38, pp. 1769-199.

[40] Xiao, D. (2013) "Is Privacy Compatible with Truthfulness?" *ITCS '13*, pp. 67 - 86.

A Proofs

It will be convenient to denote the probability of a given sender's realization on trait j as ρ_j^*, so that for type $\omega = [1\ 0\ 1]$, $\rho_1^* = \rho_1, \rho_2^* = (1 - \rho_2)$, and $\rho_3^* = \rho_3$, and therefore $\Pr(\omega) = \prod_{j=1}^{N} \rho_j^*$. We use this notation throughout the proof appendix.

Proposition 1. *For any grammar g there exists $M > 0$ such that there is a perfect Bayesian equilibrium in which all sender-types use grammar g.*

Proof. Along the equilibrium path, assume that the sender-types follow the complete grammar defined by the strategy for their type. If the sender follows a grammar assigned to some sender-type, then the receiver's beliefs are generated by Bayes's rule.

If the sender ever deviates to a grammar used by no other type of sender, then at every successor history the receiver's beliefs place probability 1 on the sender having a particular type that does not match the receiver's. Therefore, the receiver will choose End after such a deviation. Given these events off-the-path, the sender finds it optimal to play as required on the path for large enough M since to do otherwise would foreclose the possibility of a profitable match. □

Proposition 2. *Suppose M is sufficiently large that a common grammar entailing participation by the sender exists. Among all common-grammar equilibria, all sender-types prefer the equilibrium using:*

1) the efficient version of grammar g to the inefficient grammar g
2) the complete version of grammar g to an incomplete grammar g
3) the sequential version of grammar g to a non-sequential grammar g
4) the ordered version of grammar g to the misordered grammar g

Proof. As a preliminary step, note that since the choice to Attend is part of the grammar, equilibria with a common grammar entail either every sender-type or no sender-type chooses Attend. Because we assume that M is sufficiently large that the sender-types choose Attend, we consider only the former case.

To start our proof of **claim 1**, note that aside from the cost of delaying the close of a conversation, the efficient and inefficient grammars yield the same payoffs. Since we lexicographically drop ties of this form in favor of the shorter grammar, the shorter efficient grammar is strictly preferred to the longer inefficient grammar. This yields claim 1.

To prove **claim 2**, consider an incomplete grammar g and its complete version g'. The only difference in the outcomes of these two grammars occurs at history h at which (a) there are unrevealed sender traits and (b) no further messages are sent. Since we are assuming full participation by both sender- and receiver-types, the only possible

outcome following history h under grammar g is Match,[16] so all of the sender's traits are revealed to the receiver. This is also the case under g', so the information penalties are identical under each grammar. However, g' screens out all non-matching receiver-types, which means the sender need never incur the loss L that results from a mismatch between the sender and receiver types. Therefore g' is preferred by all sender-types to g, which yields claim 2.

To prove **claim 3**, consider a non-sequential grammar g and its sequential version g'. For the purposes of our argument, assume that the only difference between g and g' is that a single message m in g reveals two or more traits. We will prove that dividing this messages into two messages, one revealing a single trait and the second revealing the remainder of the content of m, is preferred by all sender-types. A straightforward (and omitted) induction argument then gives us claim 3.

Throughout this proof we use $g_1 \oplus g_2$ to denote the concatenation of g_2 to the end of g_1. We can write $g = g_1 \oplus m \oplus g_2$ for appropriately chosen g_1 and g_2, where g_1 or g_2 might be empty if m is the first or last message respectively. Suppose m reveals trait i as well as a set of other traits with indices in set $\mathcal{S}$. Let m_1 be a message that reveals trait i, while m_2 is a message that reveals the traits with indices in S. Consider the grammar $g' = g_1 \oplus m_1 \oplus m_2 \oplus g_2$. Let $\mathcal{T}$ denote the indices of traits revealed in g_1.

The payoff to grammars g and g' differ only in the event that the sender and receiver differ on a trait revealed in m. The payoff to following grammar g' in this event can be written

$$\begin{aligned} -(1-\rho_i^*)* &\left(\textstyle\sum_{t\in\mathcal{T}} v_t + v_i + \sum_{t\notin\mathcal{T}\cup\{j\}} \rho_t^* * v_t\right) \\ &- \rho_i^*(1-\prod_{s\in\mathcal{S}}\rho_s^*)\left(\textstyle\sum_{t\in\mathcal{T}} v_t + v_i + \sum_{s\in\mathcal{S}} \rho_s^* v_s + \sum_{t\notin\mathcal{T}\cup\mathcal{S}} \rho_t^* * v_t\right) \end{aligned} \tag{A.1}$$

Grammar g has an expected payoff following g conditional on differing on either trait j or k is equal to

$$-(1-\rho_i^*\prod_{s\in\mathcal{S}}\rho_s^*)\left(\textstyle\sum_{t\in\mathcal{T}} v_t + v_i + \sum_{s\in\mathcal{S}} v_s + \sum_{t\notin\mathcal{T}\cup\mathcal{S}} \rho_t^* * v_t\right) \tag{A.2}$$

Subtracting Equation A.2 from A.1 yields

$$(1-\rho_i^*)\prod_{s\in\mathcal{S}}(1-\rho_s^*)\,v_s > 0$$

The sender thus has a strict preference for grammar g' to grammar g, proving our claim.

To prove **claim 4**, consider a misordered grammar g. Let message m_j in grammar g reveals trait $\beta(j)$. Suppose for some $j \in \{1,..,N-1\}$ we have $v_{\beta(j)} > v_{\beta(j+1)}$. We

[16] If a receiver-type instead chose End, then that receiver-type could never match. Because of the lexicographic preference for shorter conversations, this receiver-type would instead choose Not Attend, which would violate our assumption of full participation.

show that senders prefer the grammar $g' = (m_1, .., m_{j-1}, m_{j+1}, m_j, m_{j+2}, ., m_N)$. A straightforward (and omitted) induction argument suffices to then prove our claim.

Note that the only difference between g and g' is that under g trait $\beta(j)$ is revealed before $\beta(j+1)$, whereas under g' trait $\beta(j+1)$ is revealed before trait $\beta(j)$. The sender's payoff only differs between the grammars on the event where the sender and receiver have different realizations of either trait $\beta(j)$ or $\beta(j+1)$ and match on all previously revealed traits. Conditional on the sender and receiver having different values of trait $\beta(j)$ or $\beta(j+1)$ and the same values for traits $\beta(1)$ through $\beta(j-1)$, the sender has an expected utility under grammar g equal to

$$\begin{aligned} &-\left(1-\rho^*_{\beta(j)}\right) * \left(\textstyle\sum_{k=1}^{j} v_{\beta(k)} + \sum_{k=j+1}^{N} \rho^*_{\beta(k)} v_{\beta(k)}\right) \\ &\qquad -\left(1-\rho^*_{\beta(j+1)}\right)\rho^*_{\beta(j)} * \left(\textstyle\sum_{k=1}^{j+1} v_{\beta(k)} + \sum_{k=j+2}^{N} \rho^*_{\beta(k)} v_{\beta(k)}\right) \end{aligned} \tag{A.3}$$

Conditional on the sender and receiver having different values of trait $\beta(j)$ or $\beta(j+1)$ and the same values for traits $\beta(1)$ through $\beta(j-1)$, the sender has an expected utility under grammar g' equal to

$$\begin{aligned} &-\left(1-\rho^*_{\beta(j)}\right)\rho^*_{\beta(j+1)} * \left(\textstyle\sum_{k=1}^{j+1} v_{\beta(k)} + \sum_{k=j+2}^{N} \rho^*_{\beta(k)} v_{\beta(k)}\right) \\ &\qquad -\left(1-\rho^*_{\beta(j+1)}\right) * \left(\textstyle\sum_{k=1}^{j-1} v_{\beta(k)} + v_{\beta(j+1)} + \rho^*_{\beta(j)} v_{\beta(j)} + \sum_{k=j+2}^{N} \rho^*_{\beta(k)} v_{\beta(k)}\right) \end{aligned} \tag{A.4}$$

Subtracting Equation A.3 from Equation A.4 yields

$$\left(1-\rho^*_{\beta(j)}\right)\left(1-\rho^*_{\beta(j+1)}\right)\left(v_{\beta(j)} - v_{\beta(j+1)}\right) > 0$$

where the inequality follows from our assumption that $v_{\beta(j)} > v_{\beta(j+1)}$, which implies that the sender strictly prefers g' to g. □

Proposition 3. *All sender-types prefer the equilibrium in which all sender-types use grammar $g = (\{1\}, \{2\}, ..., \{N\})$ to any other equilibrium satisfying block inference.*

Proof. Our argument consists of two steps. First, we prove that sender-types prefer that receivers have what we term "straightforward beliefs" that do not infer anything from a message other than what is revealed explicitly (i.e., there is no signaling). Once this has been shown, Corollary 1 provides our result immediately.

With this argument structure in mind, we need the following definition:

Definition 3. *The receiver's beliefs satisfy* ***straightforward inference*** *if the beliefs following any history of play are conditioned only on the verifiable information contained in the messages received.*

Given any history h of an equilibrium that satisfies block inference, it must be the case that the sender's utility would be higher if the receiver's beliefs satisfy straightforward inference rather than the equilibrium beliefs. To see this, note first that for traits

have been explicitly revealed in a message in equilibrium, the information penalty is the same. For all other traits one of the following is true in equilibrium: (1) the receiver has not updated his prior beliefs or (2) the receiver places probability 1 on the true realization of the sender's trait. In the first case, the payoff for the sender is the same under equilibrium and straightforward inferences. In the second case, the payoff for the sender is higher under straightforward inferences than equilibrium inferences.

Having made the improvement step (for any equilibrium) from equilibrium inferences to straightforward inferences, Corollary 1 implies that we can maximally improve payoffs by having all sender-types use the grammar that reveals information in order of increasing information value. Given that all agents use the same grammar, straightforward inferences now are the equilibrium inferences made by the receiver, meaning that we have found the best possible PBE for all sender-types. □

APPENDIX: For Online Publication Only

B Extensions

This section notes an additional result and five extensions to the model, which we now summarize here.

The additional result shows that, although all sender-types all prefer the same common grammar, many different grammars converge on the same payoff as the number of traits rises.

Our first extension considers situations in which successful matches require matching on some (but not necessarily all) traits. We show that the senders still prefer to release traits in order of increasing information value.

Our second extension provides a partial characterization of our model when sending messages is costly. When the cost of sending messages is small, the results above continue to hold. That is, agents prefer to reveal one trait at a time because the screening benefit outweighs the cost. Higher messaging costs provide an incentive for the sender to reveal multiple traits in a single message, and this incentive pushes against the dynamic screening incentives analyzed in our main model. We argue that the sender balances the cost of multiple messages and the dynamic screening concerns by revealing messages with a message size that grows almost exponentially as the conversation proceeds.

Throughout the main paper, we assume that the information penalty for trait j is the product of the information value and the other player's beliefs. Our third extension allows for non-linear transformations of beliefs in this calculation using an increasing function $\pi(\cdot)$. We show that if π is moderately nonlinear in the other player's beliefs, our results continue to hold. When π is more nonlinear, a sender-type's preferred grammar will depend on a particular combination of the information value of the traits and the trait realizations' rarity. This result implies that sender types may disagree on the optimal grammar, but that all sender-types will prefer to reveal traits in increasing information value if the traits have sufficiently different information values.

Our fourth extension considers situations where the sender's messages are cheap-talk. In our baseline model, senders can economize on the information penalty by choosing Not Attend. If the messages are cheap-talk, then the sender could also reduce her information penalty by mimicking another type. However, this behavior exposes the sender to the risk that a receiver of the mimicked type will be "tricked" into choosing Match, which causes the sender to suffer a welfare loss of $-L$. We can eliminate these deviations (and therefore allow the sender's messages to be cheap-talk) if L is sufficiently large.

Our final extension considers two-sided conversations where the roles of sender and receiver swap exogenously between the agents. We assume the receiver is also a bona fide strategic agent who suffers an information penalty from revealing his type. To extend our assumption of verifiable messages to a model where the receiver is a strategic

agent, we assume that the receiver's messages must verifiably reveal whether his traits match the values revealed by the sender's message.[17] In this setting, all of the types of each player agree on the best equilibria, but the players can have different opinions over the best equilibria depending on their role at each stage.[18]

B.1 Many Equilibria Efficient in the Limit

While sender-types have preferences over finite conversations, many common-grammar equilibria approach the same payoffs as the number of traits is large. The basic intuition is that as $N \to \infty$, the conversation will end with high probability before most of the traits have been revealed. Regardless of which grammar is employed, the expected percentage change in the information penalty is small.

Proposition 4. *Consider a sequence of traits with associated parameters (ρ_j, v_j) and a grammar in which at most K traits are revealed in each stage. Assume that v_j is bounded from above by $\overline{v}$ and $\rho_j < \overline{\rho} < 1$.*[19] *Then*

$$\begin{aligned} \frac{1}{N}\sum_{j=1}^{N}\left\{E\left[\mu_R\left(\omega_j^S = 0|h\right)|\omega_j^S = 0\right] - (1-\rho_j)\right\}v_j &= O\left(\frac{K}{N}\right) \\ \frac{1}{N}\sum_{j=1}^{N}\left\{E\left[\mu_R\left(\omega_j^S = 1|h\right)|\omega_j^S = 1\right] - \rho_j\right\}v_j &= O\left(\frac{K}{N}\right) \end{aligned}$$

Proof. Suppose h represents the history at which the game ends, and trait ω_j is revealed at stage t. Let $\mathcal{S}_t$ denote the indices of the traits revealed before ω_j. Then we can write

$$E\left[\mu_R\left(\omega_j^S = 0|h\right)|\omega_j^S = 0\right] = \prod_{s\in\mathcal{S}_t}\rho_s^* + \left(1 - \prod_{s\in\mathcal{S}_t}\rho_s^*\right)(1-\rho_j^*)$$

which implies that

$$E\left[\mu_R\left(\omega_j^S = 0|h\right)|\omega_j^S = 0\right] - (1-\rho_j) = \rho_j^*\prod_{s\in\mathcal{S}_t}\rho_s^*$$

[17] It turns out to not be particularly important that the receiver reveals whether or not he matches the individual traits revealed by the sender (as opposed to the receiver revealing whether or not *some* trait revealed by the sender is incompatible with the receiver's type).

[18] For example, suppose there are three traits and the agents exchange the role of sender and receiver in each stage. The preferred equilibrium of the agent who is the receiver in stage 1 is to have the sender reveal her entire type in the first stage. The preferred equilibrium of the agent who is the sender in the first stage is for the first sender to reveal the lowest information value trait and then have the other player reveal the remaining two traits in the second stage.

[19] We can allow v_j to grow, but the crucial point is that v_j cannot grow too quickly. If v_j grows exponentially, then convergence may fail entirely.

Using this result we find that

$$\begin{aligned}\frac{1}{N}\sum_{j=1}^{N}\left\{E\left[\mu_R\left(\omega_j^S=0|h\right)|\omega_j^S=0\right]-(1-\rho_j)\right\}v_j &= \frac{1}{N}\sum_{j=1}^{N}\rho_j^* v_j\prod_{s\in\mathcal{S}_t}\rho_s^* \\ &\leq \frac{K}{N}\overline{\rho v}\sum_{j=1}^{N}\overline{\rho}^j \\ &< \frac{K}{N}\frac{\overline{\rho v}}{1-\overline{\rho}}=O\left(\frac{K}{N}\right)\end{aligned}$$

□

Our bound is tight. Consider a case where $(\rho_j, v_j) = \left(\frac{1}{2}, 1\right)$ and one-trait is revealed in each stage. In a multi-stage conversation a few traits must be revealed with high probability, but it is rare that later traits will be revealed. The $O\left(\frac{1}{N}\right)$ bound (which is tight in this case) reflects the speed with which the average discounts the welfare cost of fully revealing those early traits. Loosely, this result suggests that, while the common-order grammar discussed in Proposition 2 is strictly preferred by all senders, the relative penalty from using a sub-optimal grammar that gradually reveals traits drops as the number of traits increases.

B.2 Matching With "Close" Types

In the model above we assumed that the sender and receiver must have the same type for a match to be profitable, whereas in most economic interactions agents find matching profitable if they have similar, but not identical, types. In this subsection we examine equilibria wherein senders and receivers find it profitable to match if and only if they discover that they share $K \leq N$ trait realizations. We show that sender optimal equilibria that satisfies block inference continue to have the same structure as described in Section 4.[20]

Our candidate sender optimal equilibrium is one in which traits are revealed by all sender-types at all histories in order of increasing information value. Since choosing Match is incentive compatible as soon as the sender and receiver are known to share K trait realizations, we focus on equilibria in which the receiver chooses Match as soon as K common trait realizations are observed. We denote the resulting pair of equilibrium strategy for the sender and beliefs for the receiver as (σ^*, μ_R^*). We now argue that (σ^*, μ_R^*) is the most preferred equilibrium by all sender-types.

Given a history h, let N_h be the number of traits revealed by reaching h, and K_h be the number of revealed traits where the sender and receiver have matching values.[21] We refer to any history h where $N - N_h = K - K_h$ as a *last chance history*. At any

[20] Some components of our game, such as the definition of a history, need to be elaborated in obvious ways to account for the fact that the agents can fail to match on some traits and still choose to match in equilibrium. We ignore these technical issues.

[21] The sender and receiver mismatch for the remaining $N_h - K_h$ traits.

history where $K_h \geq K$, the receiver chooses Match, which implies that both the sender and receiver pay the full information penalty. This implies that the sender is indifferent to the set of traits that were revealed or the order in which the traits were revealed along that history. At any last chance history, the remaining traits must match or the receiver will choose to end the conversation with the action End. In any subgame that reaches a last chance history, the sender and receiver are in effect playing the multi-stage conversation game analyzed in the main text.

These two observations imply that when assessing whether (σ^*, μ_R^*) is the most preferred equilibrium by all sender-types, we can restrict attention to improvements along paths of play that terminate at last chance histories. As argued in Section 4, at any last chance history the sender weakly prefers to reveal hidden traits in order of increasing information value. Since σ^* reveals traits in order of increasing information value at every history, this argument implies that (σ^*, μ_R^*) is the most preferred equilibrium by all sender-types.

(σ^*, μ_R^*) is one of several payoff equivalent equilibria of our game. Consider some history h, and let B_h denote a set of traits that is revealed before any last chance successor history.[22] The sender-types are indifferent to the number of messages or order of messages used to reveal the traits in B_h as long as they are revealed before any other traits. For example, consider the beginning of the game, h_0. If $K = N - 1$ then in our equilibrium B_{h_0} contains the two lowest value traits. Since B_{h_0} are surely revealed in the most preferred equilibrium by all sender-types, the sender-types are indifferent with regard to the order in which these traits are revealed. The information value of these traits is a "sunk cost" that is paid by the sender. With this notion of the low value traits as a sunk cost, it is intuitive that the sender types would seek to minimize these costs by revealing the low information value traits first.

B.3 Costly Messages

In our model, there is no cost to sending or receiving a message. Because of this, the sender has an incentive to drag out the conversation by revealing a single trait in each message. There are often real costs, however, to transmitting and processing messages. This suggests that senders may have an incentive to reveal more than one trait in each message to economize on the communication costs. In this section we focus on how these costs influence the structure of the most preferred equilibrium.[23]

First, note that when we consider costly messages, there is a question of how the traits are packed into messages and the order in which these messages are released. We call the first problem the *packing problem*. If we consider equilibria where all of the sender types use the same grammar, then a simple modification of the proof of

[22] This definition is subtle - whether a successor history to h, denoted h', is a last chance history depends on whether the sender and receiver match on the traits revealed between h and h'.

[23] For very small message costs, the sender's incentive to dynamically screen out non-matching receivers obviously dominates the incentive to save on messaging costs by revealing multiple traits in a single message

Proposition 2 would imply that the messages are revealed in order from lowest to highest total information value, where the total information value is the sum of the information values of the individual traits revealed within a given message. In effect, we can consider each of these aggregated messages as a nonbinary trait.

To simplify our discussion of the packing problem when costs get larger, we focus on a completely symmetric model where $v_j = 1$ and $\rho_j = \frac{1}{2}$. The symmetry implies that we can ignore the signaling concerns that dominated our analysis of the more general model, and focus on the effect of costs on how much information is conveyed by each message. We describe solutions to the packing problem with J messages as $P = (p_1, ..., p_J)$ where $p_s \geq 1$ describes an integer number of traits to reveal in stage s.

The packing problem involves complicated combinatorics, which means it is not amenable to closed-form analysis. However, there are a few incentives of interest that we can highlight even given the limited tractability of the problem. Two tensions are at play in our analysis. First, the sender wants to reveal as little information as possible in early messages to avoid information penalties in the event that the receiver does not share the trait realizations revealed in the first message. Second, in order to effectively screen out a large fraction of the non-matching receivers, the sender is required to reveal multiple traits in the first message.

Suppose that K traits are to be revealed within two messages. What is the optimal way of dividing the K traits between the two messages? Suppose we reveal $a \in \{1, .., K-1\}$ traits in the first message with a chosen to solve

$$\min_a \left(1 - \frac{1}{2^a}\right)\left(a + \frac{K-a}{2}\right) + \frac{1}{2^a}K$$

which seeks to minimize the expected information penalty due to a mismatch on one of these K traits. Simplifying this we find the equivalent problem

$$\min_a a + \frac{K-a}{2^a}$$

If we treat a as a continuous variable, then from the concavity of the problem we know that the optimal choice of a satisfies the following first order condition

$$2^a = 1 + (K-a)\ln 2$$

In our two message case, the balance of the two tensions implies that a^* grows at a rate slightly slower than $\log_2 K$. Therefore, as K grows a large number (but an arbitrarily small fraction) of the traits are revealed in the first message.

Our analysis can be easily extended to settings where more than 2 messages are employed. The more general conclusion we reach is that early messages reveal fewer traits than later messages, and (holding fixed the total number of traits) when the number of messages used decreases (i.e., message costs rise) the ratio of the number of traits revealed in early messages relative to later messages grows roughly exponentially.

B.4 Alternative Information Penalty Functions π

In the baseline model, the sender's information penalty for each trait is equal to the receiver's posterior on the sender's true realization for that trait multiplied by the information value v_j of that trait, which assumes that the information penalty function π is linear. In this section we consider nonlinear specifications of π.

When π is linear, all sender-types prefer that traits be ordered in conversation from low information value to high information value (Proposition 2). This result is independent of the rarity of the sender-type's trait realizations due to the trade-off between the privacy and screening incentives: revealing an unlikely trait realization reveals more information, but has a higher chance of removing a non-matching partner from the conversation. When π is not linear, these opposing forces are still present, but no longer perfectly cancel. As a result, senders with different types may have different preferences over common-grammar equilibria.

Let $\rho_j^*(\omega)$ be the ex-ante probability that type $\omega's$ value of trait j is realized, which can be written $\rho_j^*(\omega) = \rho_j$ if $\omega_j = 1$ and $\rho_j^*(\omega) = 1 - \rho_j$ otherwise.

Proposition 5. *Given straightforward inferences, type ω prefers to reveal trait j before trait k if and only if*

$$v_j \frac{1 - \pi\left(\rho_j^*(\omega)\right)}{1 - \rho_j^*(\omega)} \leq v_k \frac{1 - \pi\left(\rho_k^*(\omega)\right)}{1 - \rho_k^*(\omega)} \tag{B.1}$$

Proof. Suppose some type of sender ω^S has the most preferred grammar $g \in \mathcal{G}^*$ of the form $g = \{m_1, m_2, ..., m_N\}$ where message m_j reveals trait $\beta(j)$. Suppose for some $j \in \{1, .., N-1\}$ we have $v_{\beta(j)} > v_{\beta(j+1)}$ contradicting our claim for senders of type ω^S. We show that senders of type ω^S prefer the grammar $g' = (m_1, .., m_{j-1}, m_{j+1}, m_j, m_{j+2}, ., m_N)$ to g which contradicts our assumption that g was the ideal grammar of type ω^S and establishes our claim.

Note that the only difference between g and g' is that under g trait $\beta(j)$ is revealed before $\beta(j+1)$, whereas under g' trait $\beta(j+1)$ is revealed before trait $\beta(j)$. Note that the sender's payoff only differs between the grammars in the event where the sender and receiver have different realizations of either trait $\beta(j)$ or $\beta(j+1)$.

Conditional on the sender and receiver having different values of trait $\beta(j)$ or $\beta(j+1)$ and the same realizations of traits $\beta(1)$ through $\beta(j-1)$, the sender has an expected utility under grammar g equal to

$$\begin{aligned} &-\left(1 - \rho_{\beta(j)}^*\right) * \left(\textstyle\sum_{k=1}^{j} v_{\beta(k)} + \sum_{k=j+1}^{N} \pi\left(\rho_{\beta(k)}^*\right) v_{\beta(k)}\right) \\ &\quad - \left(1 - \rho_{\beta(j+1)}^*\right) \rho_{\beta(j)}^* * \left(\textstyle\sum_{k=1}^{j+1} v_{\beta(k)} + \sum_{k=j+2}^{N} \pi\left(\rho_{\beta(k)}^*\right) v_{\beta(k)}\right) \end{aligned} \tag{B.2}$$

Conditional on the sender and receiver having different values of trait $\beta(j)$ or $\beta(j+1)$,

the sender has an expected utility under grammar g' equal to

$$-\left(1-\rho^*_{\beta(j)}\right)\rho^*_{\beta(j+1)} * \left(\sum_{k=1}^{j+1} v_{\beta(k)} + \sum_{k=j+2}^{N} \pi\left(\rho^*_{\beta(k)}\right) v_{\beta(k)}\right) - \tag{B.3}$$
$$\left(1-\rho^*_{\beta(j+1)}\right) * \left(\sum_{k=1}^{j-1} v_{\beta(k)} + v_{\beta(j+1)} + \pi\left(\rho^*_{\beta(j)}\right) v_{\beta(j)} + \sum_{k=j+2}^{N} \pi\left(\rho^*_{\beta(k)}\right) v_{\beta(k)}\right)$$

Subtracting Equation B.2 from Equation B.3 yields a positive quantity if and only if

$$\frac{v_{\beta(j+1)}}{v_{\beta(j)}} \leq \frac{\left(1-\rho^*_{\beta(j+1)}\right)\left(1-\pi\left(\rho^*_{\beta(j)}\right)\right)}{\left(1-\rho^*_{\beta(j)}\right)\left(1-\pi\left(\rho^*_{\beta(j+1)}\right)\right)}$$

which is equivalent to Equation B.1 for traits $\beta(j)$ and $\beta(j+1)$. □

The non-linearity of π combined with the fact that different types of senders have different values of $\rho^*_j(\omega)$ drives the senders to have different ideal grammars. The information values, v_j, are common across types and push different sender types to have the same most preferred grammar. The following corollary captures this tension by illustrating that the nonlinearity of $\pi(\rho)$ is not an issue when traits have sufficiently different information values. It is only when the information values are of a comparable level that the relative rarity of a trait realization drives disagreement between the sender-types.

Corollary 2. *All agents prefer to reveal trait j before trait k if both of the following hold*

$$\begin{aligned} v_j\frac{1-\pi\left(\rho_j\right)}{1-\rho_j} &\leq \min\left\{v_k\frac{1-\pi\left(\rho_k\right)}{1-\rho_k}, v_k\frac{1-\pi\left(1-\rho_k\right)}{\rho_k}\right\} \\ v_j\frac{1-\pi\left(1-\rho_j\right)}{\rho_j} &\leq \min\left\{v_k\frac{1-\pi\left(\rho_k\right)}{1-\rho_k}, v_j\frac{1-\pi\left(1-\rho_k\right)}{\rho_k}\right\} \end{aligned} \tag{B.4}$$

Proof. Equation B.4 follows by requiring Equation B.1 to hold for all possible realizations of traits j and k. □

Equation B.1 implies that a sender of type ω has preferences over the order of trait revelation that depend jointly on the probability of each ω_j, $\rho^*_j(\omega)$, and the value of the trait revealed, v_j. In the case where the traits have equal information values, we can describe sender preferences over different grammars in terms of the convexity of π and the probability of the trait realizations.

Corollary 3. *Assume $v_j = v_k$ for all j and k and that π is differentiable. If π is strictly concave, type ω prefers the equilibrium in which all agents reveal traits from highest $\rho^*_j(\omega)$ to lowest $\rho^*_j(\omega)$. If π is strictly convex, type ω prefers the equilibrium in which all agents reveal traits from lowest $\rho^*_j(\omega)$ to highest $\rho^*_j(\omega)$.*

Proof. For a single agent, Equation B.1 can be written

$$\frac{1-\pi\left(\rho_j^*\right)}{1-\rho_j^*} \leq \frac{1-\pi\left(\rho_k^*\right)}{1-\rho_k^*}$$

Since π is differentiable we can write

$$\frac{d}{d\rho}\left[\frac{1-\pi\left(\rho\right)}{1-\rho}\right] = \frac{-\pi'(\rho)}{1-\rho} + \frac{1-\pi(\rho)}{(1-\rho)^2}$$

Note that this term is negative (positive) for all ρ if π is concave (convex). Therefore when π is concave (convex) agents prefer to release traits in decreasing (increasing) order of ρ. □

To understand the intuition for this corollary, consider the case in which π is strictly concave. Concavity raises the relative cost of revealing the rare trait realization (0) versus the more common realization (1) of a trait. For example if $\rho_1 = .8$, revealing $\omega_1 = 0$ leads to four times the information penalty of revealing $\omega_1 = 1$ when π is linear ($1 - .2$ vs. $1 - .8$). If π is strictly concave, revealing $\omega_1 = 0$ must lead to more than four times the penalty of revealing $\omega_1 = 1$ ($\pi(1) - \pi(.2)$ vs. $\pi(1) - \pi(.8)$). This is demonstrated in the following example where all senders are assumed to use the same grammar in equilibrium.

Example 4 (Preferences over grammars given non-linear π):

Assume $N = 2$**,** $\rho_1 = .8$**,** $\rho_2 = .6$**, and** $v_1 = v_2$**. Focus on sender-type [1 1].**

Grammar 2: $g = [\{1\}, \{2\}]$ (*t=1*: Reveal trait 1, *t=2*: Reveal trait 2)

Expected info penalty:
$$\begin{aligned} &= -.48(\pi(1) + \pi(1)) - .32(\pi(1) + \pi(1)) \\ &\quad -.12(\pi(1) + \pi(.6)) - .08(\pi(1) + \pi(.6)) \\ &= -1.40\pi(1) - .20\pi(1) - .20\pi(.6) \end{aligned}$$

Grammar 3: $g = [\{2\}, \{1\}]$ (*t=1*: Reveal trait 2, *t=2*: Reveal trait 1)

Expected info penalty:
$$\begin{aligned} &= -.48(\pi(1) + \pi(1)) - .32(\pi(1) + \pi(1)) \\ &\quad -.12(\pi(1) + \pi(1)) - .08(\pi(.8) + \pi(1)) \\ &= -1.40\pi(1) - .40\pi(.8) \end{aligned}$$

Payoffs are equal if π is linear.
Payoff from Grammar 2 is greater if π is concave.
Payoff from Grammar 3 is greater if π is convex.

Once types have different preferences over grammars, it becomes more difficult to select between the multiplicity of equilibria.[24] We leave further analysis of this difficult

[24]Principally, this difficulty is due to the lack of a clear criteria for choosing between the equilibria.

problem for future work.

B.5 Cheap-talk Messages

In the main model we assumed that the sender's messages are verifiable. If all types of sender have a positive payoff from participating in the conversation, then there exists an equilibrium where the senders will continue to use truthful messages in order to preserve the possibility of profitably matching. However, cheap-talk messages can eliminate truthful equilibria, and in particular equilibria with efficient matching, when payoffs from participation are low.

If some sender-type ω^S obtains a negative payoff from participating, she may still be willing to participate given the receiver's beliefs about senders that choose Not Attend. The optimal deviation from truthfulness for this sender may be to mimic the behavior of the sender with the opposite trait realizations. The receiver, believing the messages to be truthful, will believe that the sender is actually of the mimicked type, which reduces the sender's information penalty to 0. The only incentive to not follow this deviation is the possibility that a receiver is of this mimicked type and chooses Match, which means the sender gets a total payoff of $-L$. However, if L is sufficiently small, it can be worth deviating in this way and risking suffering the $-L$ payoff.

B.6 Two-Sided Conversations

We have focused on the case where the roles of sender and receiver are fixed over time. Many forms of information exchange in the real-world involve the sequential revelation of information - in other words, the roles of sender and receiver are exchanged as the conversation progresses. In this section we consider a model where the roles are exchanged in each stage in an exogenous fashion (although it can be stochastic). We break our discussion into two parts. First, we argue that the sender and the receiver have differing preferences over the optimal equilibrium, but that we can still use the logic of Proposition 3 to refine our equilibrium set. Second, we discuss the implications of weakening our assumption of verifiable messages in a two-sided conversation, which we find has more complex consequences for two-sided conversations than for one-sided conversations (Section B.5).

Standard signaling equilibrium refinements are of little use in our model.

B.6.1 Modeling Two-Sided Conversations

In this setting, we assume that both players suffer an information penalty and that both the sender's and the receiver's message are verifiable. However, since the receiver suffers an information penalty in this model, we need to be concrete about what information the receiver conveys. Given that the sender has revealed potentially multiple traits, there are two natural possibilities. One modeling choice is to assume the receiver must verifiably reveal whether or not the receiver has learned from the message that a profitable match is impossible. The other modeling choice is to assume the receiver must verifiably reveal the realization of each trait revealed by the sender.

First let us consider a model wherein the receiver must verifiably reveal whether or not the receiver has learned from the message that a profitable match is impossible.[25] For example, suppose that the sender issues a message $\{\omega_3 = 1, \omega_5 = 0\}$, and the receiver has a type $\omega_3^R = 0$. The receiver must verifiably reveal that the sender and receiver types do not match, but does not need to reveal whether it is the case that $\omega_3^R = 0$, $\omega_5^R = 1$, or both. Informally speaking, the sender has revealed more information than the receiver in this stage.

In this setting, the players can have different opinions over the best equilibria depending on their role in each stage. For example, suppose there are three traits and the agents exchange the role of sender and receiver in each stage. The preferred equilibria of the agent who is the receiver is to have the sender reveal her entire type in the first stage, which would give the sender the lowest payoff possible and the receiver the highest payoff possible. The preferred equilibria of the agent who is the sender in the first stage is for the first sender to reveal the lowest information value trait and then have the other player reveal the remaining two traits in the second stage.

This issue remains even under the second modeling choice, wherein the receiver must verifiably reveal all of the traits revealed by the sender. In other words, if the sender issues a message $\{\omega_3 = 1, \omega_5 = 0\}$, the receiver must verifiably reveal traits 3 and 5. Since the sender and receiver reveal the same verifiable information in each stage, one might have assumed that it was impossible for the sender and receiver to have different information sets in the event End is chosen by the receiver. This intuition neglects the fact that the sender can reveal information through signaling. For example, suppose the sender reveals message $\{\omega_3 = 1, \omega_5 = 0\}$, which signals $\omega_1 = 1$ and $\omega_2 = 1$. The receiver is forced to reveal traits 3 or 5, but he is not forced to reveal traits 1 and 2.

[25] For now let us leave aside the issue of how this verification occurs.

If the receiver chooses End, the sender will never learn the receiver's value of traits 1 and 2. In this model agents prefer equilibria wherein:

- As the sender, all types of the agent reveal the same trait, which minimizes the verifiable information revealed and signals nothing.
- As the receiver, the sender signals as much information as possible.

Given this to what extent can we use Proposition 3 as a refinement device in our two-sided communication setting? First, let us consider only equilibria that satisfy block inference. The assumption of block inference on the equilibrium path insures that at all non-terminal histories of the conversation the agents have the same beliefs about the other agent's trait realizations. In addition, we will focus on equilibria that satisfy straightforward inference off of the equilibrium path. Straightforward inference off of the equilibrium path enhances the sender's "bargaining power" over the desired equilibrium by minimizing the inferences made by the receiver following a sender's deviation from an equilibrium strategy, which (to put it loosely) increases the incentive for the sender to break an equilibrium that she finds undesirable.

We now provide an argument for the fact that the strategies described by Proposition 3 form the unique equilibrium that satisfies block inference on the equilibrium path and straightforward inference off of the equilibrium path.

Proposition 6. *The unique equilibrium strategy that satisfies block inference on the equilibrium path and straightforward inference off of the equilibrium path requires that at each history the sender reveal the trait with the lowest information value among all previously unrevealed traits.*

Proof. As a terminological note, we refer to a trait as revealed when it has been verifiably revealed in a message or is known via signaling. In addition, when we refer to a sender revealing the lowest information value trait at a given history, we implicitly mean the lowest information value trait amongst all traits that remain unrevealed at that history. We now build an induction argument for our claim. To establish the base case of our inductive argument, consider a history where a single unrevealed trait remains. At any such history the sender must necessarily reveal this final trait, which is exactly the action prescribed by the strategy defined above.

Now we prove the induction step. Assume that for any history where K or fewer traits are unrevealed, the strategy described in the proposition is the unique equilibrium

strategy for play following that history. Now consider a history with $K+1$ unrevealed traits, and suppose (by way of contradiction) that there is an equilibrium strategy that dictates that the sender issue a message other than the one that reveals the single, lowest information value trait.

Two subtle points need to be made. First, note that since we have assumed straightforward beliefs off of the path, the sender can make a deviation without the receiver forming disadvantageous beliefs about the sender's type. Using extreme beliefs about the sender's type was the exact "trick" used in Proposition 1 to enforce a large set of equilibria. Second, since equilibrium play in all future stages requires the revelation of the lowest information value trait, the sender in the current period of the two-period game gets the same utility as the sender in the one-sided communication game.[26]

With these two insights in mind, it is merely a matter of altering the notation in the algebra of the proofs of Propositions 2 and 3 to show that if a purported equilibrium strategy dictates that the sender issue a message other than the one revealing the lowest information value trait, the sender can profitably deviate to that message in the current stage. This contradiction implies that at any history with $K+1$ unrevealed traits in equilibrium, the sender must follow the strategy described in the proposition. □

Our argument suggests that the basic patterns of conversation we have identified will occur regardless of how the identity of the sender and receiver evolve over time. Crucially, conversations will proceed with minimal amounts of information revealed in each stage, and the information that is revealed in each stage will have the lowest information value possible. Given our equilibrium refinement, agents are indifferent between playing the role of the sender and the receiver - the information revealed about their types is effectively the same. We leave elaborations of the model that draw further differentiation between these roles or endogenizes the role of sender to future work.

B.6.2 Cheap-talk Messages in Two-Sided Conversations

Section B.5 shows that the potential problems caused by assuming messages are non-verifiable in the one-sided communication game are limited. Primarily this is because the receiver, who is assumed to be nearly mechanical in the one-sided model, has the

[26]The receiver in the current stage of the two-sided communication game may get a different payoff than the sender in the one-sided communication game, but only if the sender in the current stage sends a message other than the one that reveals the lowest information value trait.

majority of the interesting deviations when the verifiability assumption is weakened. We define a *truthful equilibrium* to be one where the sender offers messages that contain information about her type that are truthful, while the receiver issues only messages that truthfully confirm that his type matches the information conveyed by the sender. Throughout this section we focus only on the equilibrium identified in Proposition 6.

If messages are cheap talk in the two-sided model, there will not be an equilibrium with two or more stages of conversation where the agents communicate truthfully following all histories unless the agents can offer transfers that act as a credible signal of truthful behavior. To see this, suppose that there were a truthful equilibrium with two or more stages of nonverifiable messages in which information is conveyed and no transfers are employed. In any such equilibrium, there is a positive probability of a history being realized wherein the sender issues a message such that the receiver knows that the sender and receiver types do not match. In a truthful equilibrium, the receiver must then issue a message revealing that the types do not match, and both the sender and receiver suffer information penalties. However, the receiver can limit his information penalty by lying about his type and "tricking" the sender into thinking a match remains possible.

For an example of when truthful behavior is suboptimal for the receiver when there is no verifiability, suppose that $N = 2$, $v_1 = 1, v_2 = 2$, $\rho_1 = \rho_2 = \frac{1}{2}$, and a truthful equilibrium exists. Consider receiver-type [0 1] and an initial sender message $\{\omega_1 = 1\}$. Given this message, the receiver knows that his type does not match the sender's. If the receiver confirms that he does not match the sender's first trait, he earns a payoff of $-1 * v_1 - \frac{1}{2} * v_2 = -2$. However, if the receiver deviates (and lies) by confirming the sender's message and then leaves after the next message (by claiming to not match the receiver's second trait), the receiver expects at the time of deviation to earn a payoff of $0 * v_1 - \frac{1}{2}(0 * v_2) - \frac{1}{2}(1 * v_2) = -1$.[27] Since the deviation is profitable for the receiver, truthfulness cannot be an equilibrium. The following proposition implies that this issue holds more generally in the context of the sender-optimal grammar.

Proposition 7. *For $N \geq 2$, there is no truthful equilibrium.*

Proof. Suppose there is a truthful equilibrium when $N \geq 2$, and consider any history h where the sender has revealed a trait that does not match the receiver's type at stage $j < N$. If the receiver truthfully reveals that his type does not match the sender's by

[27]The expectation includes the probability $\frac{1}{2}$ event that the sender lies in the second period by claiming $\omega_2^R = 0$ and the probability $\frac{1}{2}$ that the sender truthfully reveals $\omega_2^R = 1$ to end the game.

choosing End he obtains a payoff equal to

$$-\sum_{k=1}^{j} v_k - \sum_{k=j+1}^{N} \rho_k^* v_k \tag{B.5}$$

where $\rho_k^* = \mu_S(\omega_k^R = \omega_k^{R*})$. Consider what occurs if the receiver instead chooses Continue, in effect claiming that he matches trait j of the sender. The receiver follows this deviation by (falsely) verifying the messages of the sender so long as his type does not match the traits revealed by the sender. In the event the sender reveals a trait that does match the receiver's, the receiver chooses End and in effect claims the traits do not match and ends the conversation. In the event that the conversation ends with such a message at stage $\widetilde{j}$ the receiver's information penalty is

$$-\sum_{k=1}^{j-1} v_k - \sum_{k=\widetilde{j}+1}^{N} \rho_k^* v_k$$

which is an improvement over Equation B.5.

The remaining event is where the receiver deviated from truthfulness and his deviation required him to either choose Match and suffer loss $-L$ and reveal all of his traits, or reveal (truthfully) that the trait N of the sender and receiver do not match by choosing End. When the receiver reveals that he does not have the same realization of trait N as the sender by choosing End, the receiver gets a payoff of

$$-\sum_{k=1}^{N-1} v_k - v_N$$

Since the probability of this event is less than ρ_N^*, the payoff from deviating as described is greater than the payoff from issuing truthful messages. □

To prove this proposition, we describe a general deviation from truthfulness that is welfare improving for any type of receiver. Consider any history prior to the final stage of the conversation where the sender reveals a trait that does not match the receiver's type. Instead of leaving the conversation, the receiver (non-truthfully) confirms the sender's message. The receiver then confirms every message from the sender that does not match his own type until a message revealing a trait that does match the receiver's type is issued. The receiver then causes the sender to believe the two agents have dif-

ferent realizations of the final trait revealed by taking action End. In the event that no such matching trait is revealed by the sender before the final stage of the conversation, the receiver chooses End at the last stage, in effect truthfully disconfirming the final trait revealed by the sender. Given a proposed truthful, full-participation equilibrium, the expected utility of every receiver-type is strictly improved by following such a deviation since each lie told reduces the receiver's information penalty. In the event that the receiver exits without ever truthfully revealing information about his type, then the information penalty of the receiver is strictly lower than if he had failed to deviate. In the event that the receiver must truthfully reveal his N^{th} trait in the final stage of the conversation, the receiver does receive an information penalty for revealing his N^{th} trait. The expected information penalty of the receiver is reduced, however, when such a deviation is followed.

The use of costly messages can mitigate the nonverifiability problem and allow for truthful information exchange. By making a payment that is only cost-effective if there is still a chance of receiving the match payoff M, the receiver can credibly demonstrate that the agents match on the previously revealed traits. We assume this cost is paid (by receivers) at the beginning of the relevant stage (e.g., cost c_2 is paid by the receiver during stage 2 before the sender has sent the message for that stage). Note that the payments need not increase monotonically as the conversation progresses since the information value and the expected information penalty need not increase together.

Proposition 8. *Suppose in stage $t \in \{1, .., N-1\}$ the receiver incurs a cost, denoted c_t, where $c_t \geq v_t + (2\rho_{t+1} - 1)v_{t+1}$. If M and L are sufficiently large, a truthful equilibrium exists.*

Proof. Note that for any set of per-stage payments, if M is sufficiently large both sender and receiver find it optimal to be truthful at any history where the receiver knows a profitable match is possible. The sender's truthfulness is a non-issue since any lie on the part of the sender leads to a failure to match with a desired partner. For sufficiently large M, the loss of the opportunity to match caused by a lie outweighs the possible benefit of any lie (even one that would set information penalties to 0). Following similar logic, for sufficiently large M, truthful messages are incentive compatible for the receiver along any path where the receiver believes that his type may match the type of the sender.

Once the sender conveys a message that reveals to the receiver that a Match is not optimal, the receiver has an incentive to lie to distort the sender's beliefs about the

receiver's type and reduce the receiver's information penalty. Notice that at stage N truthfulness is incentive compatible - if the sender and receiver types are revealed to not match at this stage, it is optimal for the receiver to reveal the failure to match rather than face a penalty of $-L > 0$ for matching with a sender of another type as well as suffering the information penalties.

Consider a receiver who first realizes his type does not match the sender's when trait $N-1$ is revealed in stage $N-1$. Let the probability of the receiver's realization of trait N be ρ_N^*. In this event, a truthful reply (revealing that the sender and receiver do not match on trait $N-1$) and a lie (claiming a match) have respective payoffs

$$\begin{aligned} Truth &: \quad -v_{N-1} - \rho_N^* v_N \\ Lie &: \quad 0 - (1-\rho_N^*)v_N - c_N \end{aligned}$$

To make truth-telling optimal, we must have

$$c_N \geq v_{N-1} + (2\rho_N^* - 1)v_N$$

Noting that $\rho_N^* \in \{\rho_N, 1-\rho_N\}$, we have that

$$c_N \geq v_{N-1} + (2\rho_N - 1)v_N$$

is required.

Turning to stage $N-2$, assume the receiver lies in stage $N-2$. We now discuss the dynamic concerns of the receiver in stage $N-1$, at which point the receiver may have to choose between telling the truth at stage $N-1$ and ending the conversation, or lying a second time and allowing the conversation to proceed to stage N. Our inequality on c_N implies that the latter form of lie is suboptimal, so (if the receiver lies) he will choose End the conversation in stage $N-1$ even if this necessitates truthfully revealing trait $N-1$.[28] Comparing the payoff from truthfully ending the conversation at stage $N-2$ to lying and ending the conversation at stage $N-1$ we find

$$\begin{aligned} Truth &: \quad -v_{N-2} - \rho_{N-1}^* v_{N-1} - \rho_N^* v_N \\ Lie &: \quad 0 - (1-\rho_{N-1}^*)v_{N-1} - \rho_N^* v_N - c_{N-1} \end{aligned}$$

[28]Note that this also implies that if the receiver can end the conversation with a lie at stage $N-1$, he will clearly choose to do so.

To make truth-telling optimal for all types of receivers, we must have

$$c_{N-1} \geq v_{N-2} + (2\rho_{N-1} - 1)v_{N-1}$$

Similar logic holds when considering any prior round, which generates the sequence of costs c_i described in the proposition. □

The intuition behind this proposition is composed of two parts. First, senders will never find it optimal to send nontruthful messages if the receiver assumes the messages are true and the sender places a sufficiently high value on matching. A lie by the sender would foreclose the opportunity to match for the relatively small benefit of improving the payoff in the event of a failure to match. Similarly, if the receiver believes that a profitable match is possible, he will respond truthfully to maintain the possibility of a match. As discussed above, the incentive problems arise when the sender conveys a message to the receiver that reveals to the receiver that a profitable match is not possible. Our choice of costs for the receiver renders all non-truthful deviations suboptimal.

Interestingly, the receiver with the most common type has the strongest incentive to deviate and lie if the sender reveals a non-matching trait. Relative to truthful behavior, all receivers gain v_j at stage j if they lie and convince the sender the receiver's trait matches the sender's message when in fact it does not. Ideally the receiver would cause the conversation to end by lying a second time and claiming a mismatch with the traits revealed by the sender in a later stage. Receivers with common trait realizations are more likely to reap this second benefit in a future stage, which makes lying in the current stage more tempting.

In lieu of a cost, the receiver could provide a transfer to the sender. Note that this transfer cannot be paid up front and must either be provided dynamically as the conversation progresses or as a (previously contracted) final payment at the close of a conversation where no match occurs. When a firm is being acquired, the seller often requires that the buyer pay a sequence of fees as the due-diligence process progresses and pay a break-up fee in the event of failed merger negotiations.

C Signaling Under Block Inference

By focusing on equilibria that satisfy block inference, we have not eliminated the potential use of signaling by sender-types through the use of type-specific grammars. Unfortunately we cannot provide analytic bounds on the number of traits that can be signaled except in special cases. The following proposition characterizes the limits of signaling when at some history all sender-types verifiably reveal a subset of traits from a set V of previously unrevealed traits. Given this restriction, the sender can signal up to roughly 50% more traits than are verifiably revealed.

To state the following proposition, note that $floor(x)$ refers to the largest integer smaller than x.

Proposition 9. *Consider history h consistent with an equilibrium that satisfies block inference. Suppose there exists a set $V \subseteq U(h)$ where $|V| = k > 0$ and all senders at history h verifiably reveal traits from V using messages of length less than or equal to k. Then at a successor history h' we can have $|K(h')| \geq |K(h)| + floor(\log_2(3^k - 1))$.*

Proof. Consider an arbitrary history h of an equilibrium that satisfies block inference. At any successor history h' it must be the case that $K(h) \subset K(h')$. Let $n = |K(h')| - |K(h)|$ denote the number of traits disclosed by the messages sent at history h. For n traits to be disclosed, we must distinguish between 2^n types of senders that are present at history h. The set of messages that verifiably reveal up to k traits within V is of size

$$\sum_{m=1}^{k} \left(\begin{array}{c} k \\ m \end{array} \right) 2^m$$

where the combinatorial term accounts for the different sets of m traits from V that can be verifiably revealed, and 2^m refers to the possible realizations of these traits. This summation is equal to

$$3^k - 1$$

In order to fully reveal n traits, we must have

$$3^k - 1 \geq 2^n$$

Solving for n we have

$$n \geq floor(\log_2(3^k - 1))$$

□

We can numerically compute the maximum number of traits that can be revealed using messages of length less than or equal to k at a history h consistent with an equilibrium that satisfies block inference. The number of length k messages that can be formed from the possible realization of n traits is

$$\sum_{m=1}^{k} \left(\begin{array}{c} n \\ m \end{array} \right) 2^m$$

In any equilibrium that satisfies block inference, we must have that traits verifiably revealed and those that are signaled must fall within the same set of n traits. In other words, the messages must be sufficient to reveal all 2^n of the possible realizations of the n traits in $K(h') \setminus K(h)$. Formally, this means we must have

$$\sum_{m=1}^{k} \left(\begin{array}{c} n \\ m \end{array} \right) 2^m \geq 2^n$$

Although there do not exist closed forms for this partial sum, figure 1 demonstrates the largest number of traits that can be revealed (n) as a function of the message length (k). After $k = 10$ the plot asymptotes to roughly $n = 4.5k$. For example if up to 4 traits are revealed verifiably, up to 18 additional traits may be revealed through signaling. Therefore even under block inference the bulk of the information conveyed by a message can be carried by signaling as opposed to verifiable information.

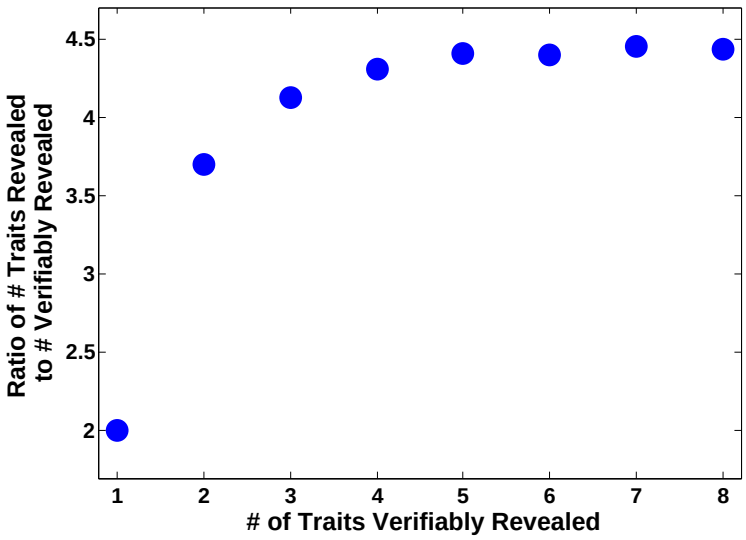

Figure 1: Numerical Results

D When Choosing Match Preserves Privacy

The model developed in the main body presumes that the information penalty suffered by a receiver who chooses to Attend does not depend on the receiver's type or whether the receiver chooses Match or End. There are two obvious alternatives to this. First, we could assume that the sender only suffers an information penalty if the receiver is of a different type (regardless of the outcome of the conversation). Second, we could assume that the information penalty is only suffered by the sender in the event that the receiver does not choose Match.

Consider the first alternative, where the sender only suffers an information penalty if the receiver is of a different type. This requires us to add the following term to the payoff from Attend and Not Attend stated in the main body

$$\left(\Pi_{j=1}^{N} \rho_j^*\right) \left(\sum_{j=1}^{N} v_j\right)$$

Note that this term does not depend on the actions of any agent, so our results on the structure of the optimal conversation do not change.

Now consider the second alternative. Suppose we adjust the match value by adding $\sum_{j=1}^{N} v_j$ to M. In the event a match occurs, the sender is "refunded" the information penalty he would have suffered in the model from the main text. Given this renormalization, all of the equations in the main text hold as stated, although we would get slightly different numerical values in our examples.

E Simple Microfoundation of Information Penalty

In order to argue that our particular reduced form information penalty is plausible, we now sketch out a microfoundation for it. We have in mind a setting where if the receiver and sender fail to match, the receiver can take actions that have (type-dependent) effects on the sender and receiver utilities. For example, the sender could be an entrepreneur whose type reflects information about a valuable new product, and the receiver is a company interested in either acquiring the entrepreneur's company (a Match outcome) or being first to market with the introduction of a competing product (an End outcome). We think of the product introduction phase as a subgame following the conversation in the event End is the outcome of the conversation, and the information penalty function is the sender's value function from this subgame.

Since a technology company can beat an entrepreneur to market, we assume the fraction of the market captured by the technology firm is increasing in the substitutability of the entrepreneur's and the technology company's products. Following these lines, we assume that the entrepreneur's payoff is decreasing in the substitutability of the products. This fits settings where the technology company is incorporating the new product into a broader platform and users face switching costs for using products outside of the platform. To overcome platform inertia, the entrepreneur's product must be markedly more appealing than alternatives (to at least some consumers).

To reflect the receiver's desire to maximize substitutability by matching the sender's type, we use the following receiver utility function in the subgame

$$u_R(a^R) = -\sum_{j=1}^{N} E\left(a_j^R - \omega_j^S\right)^2 \tag{E.1}$$

where $a_R \in [0,1]^N$ is the receiver's action and the expectation is taken with respect to the receiver's beliefs about the sender's type. Equation E.1 implies that the receiver's payoff increases with the closeness of the match between the receiver's action and the sender's type. Given Equation E.1, the optimal choice for the receiver is $a^R = E\left(\omega^S\right)$. If the sender's utility from this subgame is

$$u_S(a^R, \omega^S) = \sum_{j=1}^{N} \left\|a_j^R - \omega_j^S\right\| * v_j$$

we obtain the reduced form of Equation 2.1 as the value function for the sender in the subgame (i.e., $u_S(a^R = E\left(\omega^S\right), \omega^S)$).

F Mixed Strategies

Our focus is on determining when there is a tension between a preference for privacy and efficient matches. One might conjecture that if we are willing to abandon efficient matching outcomes then we might trade-off reduced information penalties against reduced efficiency of the matches. One way to accomplish this is for (some) of the sender-types to employ the same messages, potentially in mixed strategy equilibria.

Our goal in this section is not to conduct a comprehensive analysis of these possibilities, but to illustrate a few of the trade-offs and discuss the real-world plausibility of

the resulting equilibria. First, let us consider a simple case where multiple sender-types reveal the same message in a pure strategy equilibrium. For moderate levels of L, it is likely that the receiver will choose to not match with any of these types. These equilibria seem particularly perverse in that these agents would have achieved the same outcome if these senders (as a group) made the more intuitive choice of refusing to converse.

Now consider a sender who mixes over multiple messages, and suppose that some of these messages are also revealed by other sender-types. Suppose that each of these messages provokes a choice to Match from a single type of receiver. For simplicity, assume that message m_1 induces a matching receiver type to choose Match, while message m_2 induces a non-matching receiver type to choose Match. Let $\prod_{j=1}^{N} \rho_j^*$ denote the probability of a matching receiver type, which is relevant if message m_1 is sent. Let P denote the probability that m_2 induces a non-matching receiver to choose Match. If the sender is willing to mix over messages m_1 and m_2 it must be that the following indifference condition holds

$$M * \prod_{j=1}^{N} \rho_j^* - \sum_{j=1}^{N} E\left[\mu(\omega_j^S = \omega_j^{S*}|m_1)\right] v_j = -L * P - \sum_{j=1}^{N} E\left[\mu(\omega_j^S = \omega_j^{S*}|m_2)\right] v_j$$

Obviously if M is large, this condition cannot be satisfied - mixing is only possible if the expected games from a match are of the same order as the potential gains from reducing the information penalty. Second, if L is large then the possibility of matching with an incompatible receiver will render mixing impossible. Third, when the sender sends m_2 he must receive a lower information penalty than when he chooses to send m_1. This suggests that the sender sends m_2 only rarely and that the other agents who send m_2 must have very different trait realizations. Finally, and most interestingly, these equilibria can only exist when the payoff of the sender is negative, which is precisely when full-participation equilibria may fail to exist.