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

Ned Augenblick and Aaron Bodoh-Creed*

January 2014

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 reveal information to determine the merger's profitability, but would prefer to keep the information private if the deal fails. This preference for privacy can lead agents with rare realizations of high-value traits to avoid communicating and the chance to profitably match. We explore solutions to this inefficiency, including the use of mediators as well as a universally-desired dynamic communication protocol (a conversation) that gradually screens non-matching types.

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

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

*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.

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 or adversaries. 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 face a penalty for revealing information about their traits. We first explore how these privacy concerns can lead to frictions that make equilibria with efficient matching impossible. We demonstrate that players with traits that are unusual or have a high sensitivity have the most to lose from these discussions, and we present equilibria in which players with these *taboo* traits choose to entirely avoid communicating and lose the chance to match. We then examine a variety of solutions to increase efficiency, such as communicating through a third-party mediator. Even without a mediator, the friction can be alleviated by using a dynamic communication protocol, what we call a *conversation.* The conversation allows agents to dynamically screen out non-matching types over time, thereby limiting the information revealed to these types. We prove that dynamic conversations can eliminate a large fraction of the welfare loss caused by privacy concerns, which expands the scope of possible economic and social interactions.

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 because agents simply prefer to reveal less information about their own type to other agents, either due to a personal desire for privacy or as a reduced-form representation of settings where revealed information can lead to lower payoffs in the future. We believe that the desire for privacy is important in many situations in which parties exchange information, above and beyond standard asymmetric information issues.

In the formal 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 (i.e., the agents have different realizations of one or more traits), he ends the conversation as a profitable match is impossible and mismatching is costly. At the end of the game, 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* for each trait at the end of the game. The information penalty is the product of (1) the trait's *information value*, which represents the potential welfare costs of the loss of privacy, and (2) the receiver's posterior belief of the likelihood of the sender's true trait realization, which represents how much the receiver knows about the sender's trait (i.e., how much privacy has been lost with regard to that trait). For example, if the sender actually has a realization of 0 for trait five and the receiver believes with high (low) probability that the sender has a realization of 0 for trait five, 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 much more than revealing a common realization of a low information value trait.

As a result of this setup, the sender prefers to reveal as little information as possible to the receiver. However, she can only learn if the receiver shares her type by revealing information. To sharply display the frictions caused by privacy concerns, we first analyze a simple one-stage game in which each sender-type must fully reveal her type to all receiver-types in order to discover if a profitable match is possible. When agents receive a large payoff from matching, it is possible that all sender-types enter the conversation and reveal their type, leading to fully efficient matching. However, as the payoff from matching falls, full participation is not possible as the welfare loss from the reduced privacy required by communication outweighs the benefit of matching. The welfare cost of the loss of privacy is particularly large for types with rare realizations of traits with high information value, who reveal more costly information by communicating and have a lower chance of finding a match. When full-participation is impossible, we document equilibria in which some types choose to not participate in the game, leading certain trait realizations to never be discussed in equilibrium. We call these *taboo* traits, as agents with a particular realization of any one of these traits never enter the conversation or match.

We then explore a variety of solutions to increase the efficiency of the matching. Not surprisingly, the use of a third-party mediator who acts as an "information sink" can provide first-best payoffs and fully efficient matching. The mediator privately learns each agent's type and only communicates whether the agents match or not. We show that communication through a mediator leads to smaller frictions than any other mechanism (dynamic or static).

Even without a mediator, efficiency can be increased by allowing the use of dynamic conversations. However, the dynamic structure allows for a large set of equilibria as any sender strategy can be supported in a perfect Bayesian equilibrium. We therefore aim to refine the set of equilibria to focus on more plausible outcomes. We show that subject to a mild and realistic restriction on the forms of signaling possible, all sender-types prefer the same conversation structure, which implies that any conversation that utilizes this structure achieves the unique Pareto optimal outcome. To build this conversation structure, we first show that all sender-types strictly prefer to reveal one trait at a time, which implies a pattern to the *quantity* of information revealed over time. The dynamic nature of a conversation allows the agents to stop conversing at each stage if they learn that no profitable match is possible, which encourages the sender to reveal small amounts of information in each message. Second, all types of senders prefer to disclose traits in increasing order of information value regardless of their trait realizations, which describes a pattern to the *kind* of information revealed as a conversation progresses.

The uniformity of the sender-types' most preferred equilibrium is surprising, as one might imagine that a sender would prefer to discuss her own unusual trait realizations later in the conversation as these traits are then revealed less often. If senders did prefer to reveal their rare trait realizations later in the conversation, then a sender's most preferred equilibrium would depend on her type. Interestingly, the incentive to reveal rare trait realizations later in the conversation is balanced by the dynamic benefit of eliminating more non-matching receiver-types from the conversation in the initial stages by revealing a rare trait earlier in the conversation.

We close by discussing the relative size of the economic friction under the static, dynamic and mediated conversations. We show that the advantage of a dynamic conversation relative to a static conversation is increasing in the rarity of trait realizations and the information value of the traits. Furthermore, dynamic conversations are relatively better when there is greater heterogeneity in the information value of the traits, because the high-value traits can be pushed later into the conversation and be revealed to fewer receiver-types. Finally, we study the marginal increase in the information penalty per trait caused by participating in a dynamic conversation relative to not conversing, which is a measure of the magnitude of the friction caused by a preference for privacy. The friction vanishes quickly as the amount of information that needs to be revealed grows, which implies that dynamic communication can eliminate a large fraction of the friction in conversations of moderate length.

Finally, we examine the consequences of four modifications to our model. The first extension allows for a profitable match when agents have identical realizations of some, but not all, traits. We show that the equilibrium strategies for our dynamic model continue to be optimal in this case. In our second extension, we show that our main results continue to hold when the cost of sending messages is small but positive. As costs rise, senders have an incentive to reveal multiple traits in a single message, and we partially characterize the solution to this problem. In the third extension, we show that the results above are robust to small changes in the utility from privacy. However, as the information penalty given the receiver's beliefs becomes more concave (convex), the marginal increase in the information penalty caused by revealing a rare trait early outweighs (underweighs) the dynamic benefit of screening out more receiver-types from the conversation, which means each sender prefers to discuss her common (rare) traits first. In this case, different sender-types have different preferences over the structure of the conversation, making predictions more difficult. In the final extension we allow the sender's messages to be cheap-talk. This opens up the possibilities for more deviations, and we discuss when and how these deviations may alter the equilibria discussed in the paper. We show these additional deviations are not optimal if the penalty for mismatching is sufficiently large.

We start by reviewing related literature in section 2. In section 3 we outline the theoretical model. In section 4 we discuss the equilibria with one period of messages and characterize when equilibria with full participation (and hence no taboos) exist. In section 5 we discuss the use of a mediator to alleviate frictions caused by a preference for privacy. In section 6, we characterize the equilibria of our dynamic conversation model, and in section 7 we argue that a dynamic conversation can eliminate a significant portion of the friction in conversations of moderate length. Finally, we discuss a variety of extensions to the model in section 8 and conclude in section 9. All proofs are contained in Appendix A.

2 Literature Review

Contemporaneously and independently of our work, Dziuda and Gradwohl [6] develop a model of the exchange of informative messages with a concern for privacy that focuses on screening between one productive and many unproductive partner types where the private information of each player is exogenously ordered and infinitely-divisible. The main results of Dziuda and Gradwohl [6] provide conditions under which the joint surplus of productive 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 unproductive types and the cost of revealing additional pieces of information. In 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 unproductive types as the message size grows.[1] The focus on the optimality of gradual vs. one-shot information exchange is orthogonal to our discussions of economic frictions, identifying types that are likely to avoid conversations, taboo equilibria, and the structure of the ordering of information in the conversation.

Stein [22] models conversations between competing firms to study a tension between payoff-enhancing information exchange and the firms' desire to retain ideas for private use. Stein [22] assumes a pre-determined message order and focuses on the incentives to continue the conversation, whereas our paper is focused on what is communicated and how the messages conveying the information are structured in equilibrium. Ganglmair and Tarantino [7] modify Stein [22] by including private information.

[1] In our model, the second type of equilibrium does not satisfy our equilibrium selection as sender-types (1) can reorder the information so that less important information is revealed first and (2) can jointly deviate rather than reveal all information at once (which leads to the lowest possible payoff for all sender-types).

Hörner and Skrzypacz [13] focuses on the incentives for a sender to gather information of value only to the receiver. This paper studies the dynamic information revelation scheme that maximizes the revenues of the sender, which in turn maximizes the sender's incentives to gather the costly information. The model results in rich dynamic communication, but the goals and structure of the model are unrelated to our work.

The literature on persuasion games (e.g., Milgrom and Roberts [18]) has a loose connection with the model we develop. In most persuasion models, a persuader sends verifiable messages to a receiver in order to convince the recipient to take an action. The focus of these models is (often) deriving the properties of equilibria when the persuader and the recipient have different preferences over the actions. The majority of these models are static, although some recent papers study persuasion in a dynamic setting (Honryo [12], Hörner and Skrzypacz [13], Sher [21]). Some of the recent literature also takes a mechanism design approach and attempts to derive optimal persuasion mechanisms (Glazer and Rubinstein [10], Kamenika and Gentzkow [14]). In addition there are a wealth of applied models, many of which are discussed in Milgrom [16].

Our model touches on the literature on cheap talk. Although our benchmark model uses verifiable signals, we show in section B.4 that we can achieve many of the same results in a model with nonverifiable messages. Unlike cheap talk models, our structure is based on different preferences over information disclosure rather than different preferences over the realized outcomes. Within the cheap talk literature, the closest papers to ours are those studying dynamic cheap talk (e.g. Aumann and Hart [1], Krishna and Morgan [15]).

The computer science literature has recently defined *differential privacy*, which focuses on the amount of information revealed by an algorithm with privacy defined in terms of realized outcomes (see the survey by Dwork [5]). A broad literature incorporating bounds on differential privacy into mechanism design problems and implementation theory has begun to flourish (e.g., Gradwohl [11] and references therein).

Unlike the differential-privacy literature, we assume that the sender has preferences over the amount of knowledge that the receiver possesses about her type. In effect, the sender has preferences over the beliefs of the receiver. This aspect of our work has similarities to the literature on psychological games (Geanakoplos et al. [8]), although the similarities between our paper and the literature on psychological games ends there. Bernheim [2] presents a model of conformity with payoffs in terms of the beliefs of other agents, but the model takes place in a static setting and is focused on conformity.

Our model is distinguished from these works in a number of ways. First, the agents in our model have perfectly aligned preferences over the final outcome of the information exchange — the economic friction is rooted in the sender's preferences regarding *how* to exchange information to achieve the mutually desired goal. Second, since the sender explicitly desires to limit the disclosure of information, the dynamic aspect of our model is crucial and cannot be reduced to an equivalent game with a few stages (e.g., Sher [21]). In addition, our dynamic structure means that we must take care to address the dynamic signaling aspects of our equilibria. Third, we do not attempt to design a conversation, but take the primitives and structure of the game as given. Our goal is to analyze a realistic model of conversations and determine how our non-standard preferences over the beliefs of others generate economic frictions and outcomes such as taboos. Finally, our equilibrium resembles games where agents attempt to build trust over time (Ghosh and Ray [9] and Watson [23] among many others), although the incentive structure in and the interpretation of our model is different.

3 Model

There are two players, a sender and receiver, indexed by $i \in \{S, R\}$. 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 $\Pr(\omega) = \rho_1 \cdot (1 - \rho_2) \cdot \rho_3$. 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 assume as a convention that $\rho_i \in [\frac{1}{2}, 1)$, so $\omega_j = 1$ is the common trait realization and $\omega_j = 0$ the rare trait realization. Therefore, high values of ρ_i 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 other party, but the probabilities of the trait realizations are common knowledge.

The sender reveals messages about her traits over multiple *stages,* indexed by $t \in \{1, 2, ...\}$. In sections 4 and 5 we analyze models with a single stage of information transmission. In section 6, we allow for multiple stages of information exchange and call the dynamic exchange a *conversation.* In each stage, the sender reveals a message of the form $m \in \{\varnothing, 0, 1\}^N \subset \mathcal{M}$, 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.[2,3]

In our analysis, we will often be concerned with the "participation" decision of the sender. In many of the real-life environments we are modeling, the sender could choose to not attend the conversation in the first place (which, of course, might provide indirect information to the receiver). In our model, a sender-type that does not want to participate in the conversation can choose to only reveal null messages in equilibrium. However, to make this decision more explicit, we allow the sender to issue a *Not Attend* message in the first stage, which immediately ends the conversation. Conversely, we will refer to a sender's choice to send a message (and there not choose Not Attend) as a choice to *Attend.*

For expositional purposes, we model the receiver as mechanical.[4] In particular, we assume that the receiver initiates a match if and only if he is certain that the sender shares his type. This assumption avoids the analysis of equilibria in which a sender reveals partial information because she believes that the receiver expects only partial information in equilibrium, which may lead to inefficient matches between non-matching types. We focus on how our preference privacy can eliminate the existence of fully-efficient equilibria, which makes these other equilibria of secondary concern.

Formally, after each choice by the sender, the receiver forms beliefs about the sender's type and chooses an action.[5] If the receiver believes that the sender shares his type with certainty, the receiver chooses to *Match.* In this case, the sender receives a match payoff of $M > 0$ if the sender and receiver share the same type and $-L < 0$

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

[3] 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).

[4] In a previous version of the paper, the receiver was modeled as strategic with a preference for privacy like the sender. In addition, the agents switched communication roles in each period. In this case, all of the major conclusions in the paper hold.

[5] We note that the receiver still forms beliefs following the sender's choice of Not Attend. Although there is no action for the receiver to take, the receiver's beliefs factor into the sender's payoffs directly.

otherwise. If the receiver has learned that the sender and receiver do not share the same type with certainty, then he chooses to *End* the conversation. Similarly, if the receiver is uncertain about the match and expects no further (direct or indirect) information to be revealed given the sender's equilibrium strategy, the receiver chooses to End the conversation. If the receiver chooses to End the conversation, both agents get a match payoff of 0. In all other cases, the receiver chooses to *Continue* the conversation.

A history is comprised of all of the actions of the players up to that point in the game. As the receiver has only one action (Continue) in which the game continues, we describe a history by listing the attendance decision and the messages disclosed by the sender up to that point.[6] We denote the initial null history as h_0. We denote the history in which the sender sends a Not Attend message in the first stage (and ends the game) as h_{na}. We denote the set of histories in which the sender has issued t messages as $\mathcal{H}_t$. A generic history is a sequence $h = (m_1, ..., m_t)$ with $m_i \in \mathcal{M}$.

Denoting the set of all possible histories as $\mathcal{H}$, a strategy for the sender in each stage is a mapping of the form:[7]

$$\sigma : \Omega \times \mathcal{H} \rightarrow \mathcal{M}$$

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 $\mu(\omega^S = \omega|h)$ as the equilibrium belief held by the receiver at history h that the sender's type is equal to $\omega \in \Omega$. 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. This function is used by the sender to calculate the receiver's knowledge about the sender's true type. 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 $\mu(\omega_2^S = \omega_2^{S*}) = .8$.

The novel component of our setting is that senders have direct preferences over the information that they reveal. Specifically, we assume that a sender of type ω suffers

[6] Terminal histories reflect whether the game ended with a Not Attend message by the sender or the receiver's choice to either Match or End.

[7] One can easily show that for the agents to exploit all possible opportunities to profitably interact, the equilibrium must be outcome equivalent to a pure-strategy Nash equilibrium. We discuss the effect of mixed strategies in Appendix E.

an *information penalty* of the following form if the game ends at history h:

$$\text{Information penalty at } h\text{:}\ -\textstyle\sum_{j=1}^{N}\pi(\mu(\omega_j^S=\omega_j^{S*}|h))*v_j \tag{3.1}$$

where $v_j \geq 0$ is an exogenous *information value* assigned to trait j and $\pi(\cdot)$ is a strictly increasing function that maps beliefs into an *information penalty multiplier.* We adopt the trait labeling convention that $v_{i+1} \geq v_i$. Unless otherwise noted we assume that $\pi(\mu_j) = \mu_j$, so that the information multiplier is equal to the other player's beliefs.[8] 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 much transparent and better demonstrates the intuitions underlying our result.[9]

In our setup, the information penalty to 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. We interpret

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

as the utility representation of the preference of the sender 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 (and the agent has preferences over these future outcomes). The later interpretation is more appropriate

[8] We explore different forms of $\pi(\mu_j)$ in section B.3.

[9] 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. The only change relative to the main text is that one secondary result (proposition 1, claim 2) does not always hold. We discuss these issues in detail in appendix D.

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.[10]

4 One-Stage Message Exchange: Privacy and Taboos

In this section, we demonstrate the fundamental economic friction caused by a preference for privacy through the analysis of a simple one-stage model. We first discuss the potential existence of an equilibrium with *full participation*. Equilibria with full participation represent cases where all profitable matches are executed. We then analyze how the friction from privacy preferences can destroy this equilibrium. We describe the conditions that allow for a full participation equilibrium and identify the players most tempted to deviate from the this equilibrium. We then discuss alternative equilibria in which some sender-types choose to Not Attend the conversation and therefore do not match. We demonstrate an equilibrium where sender-types with a certain realization on a specific trait choose to Not Attend, leading to a *taboo* trait realization that is never discussed in equilibrium.

4.1 Full Participation Equilibrium

In the one-stage model, the sender can only reveal a single message to the receiver. In this game, we focus on the potential existence of a *full-participation equilibrium*, which requires all sender-types choose Attend and that the matches that follow are efficient.

Definition 1. ***Full-Participation Equilibrium***: *An equilibrium in which all sender-types engage in successful matches when the receiver has a matching type.*

To discuss the conditions under which full-participation equilibria exist, we must first characterize the sender payoffs in such an equilibrium. Since the receiver must be confident that he shares the sender's type to choose Match, the sender's message must fully reveal all of his or her traits to all receiver-types in equilibrium. The payoff for a sender of type ω is:

$$M * \prod_{j=1}^{N} \rho_j^* - \sum_{j=1}^{N} v_j$$

[10] 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.

recalling that $\prod_{j=1}^{N} \rho_j^* = \Pr(\omega)$. To understand the first term in the equation, recall that the sender communicates with a receiver with the same type with probability $\prod_{j=1}^{N} \rho_j^*$, and in such an event the sender reveals her traits, the receiver chooses Match, and both agents receive a payoff of M. Otherwise, the receiver chooses Not Match and both agents receive a match payoff of 0. The second term in the equation represents the sender's information penalty, which is always $-\sum_{j=1}^{N} v_j$ since she must completely reveal all of her traits to all receiver-types. The full participation one-stage equilibrium leads to the largest possible information penalty as, with one stage, players must reveal all information in order to match.

4.2 Existence of Full Participation Equilibrium (Inference Case)

In a full participation equilibrium, no sender-types choose to Not Attend the conversation. If the match payoff is sufficiently small, senders may have incentive to deviate and Not Attend. The payoff to this deviation depends on the inference made by receivers after observing this (off-the-equilibrium-path) choice. Note that we are free to choose this belief as it cannot be determined using Bayes rule. We denote the off-path beliefs of the receiver that a given sender-type's trait j is equal to a sender-type's actual realization given the action Not Attend as $\mu(\omega_j^S = \omega_j^{S*} | h_{na})$.

A sender-type that deviates from a full participation equilibrium by choosing Not Attend receives:

$$-\sum_{j=1}^{N} \mu(\omega_j^S = \omega_j^{S*} | h_{na}) v_j$$

The sender-type will Attend a full participation equilibrium if and only if

$$M \cdot \prod_{j=1}^{N} \rho_j^* - \sum_{j=1}^{N} v_j \geq -\sum_{j=1}^{N} \mu(\omega_j^S = \omega_j^{S*} | h_{na}) v_j \tag{4.1}$$

Equation 4.1 is equivalent to:

$$M \cdot \prod_{j=1}^{N} \rho_j^* \geq \sum_{j=1}^{N} (1 - \mu(\omega_j^S = \omega_j^{S*} | h_{na})) v_j \tag{4.2}$$

The existence of a full participation equilibrium relies on setting the receiver's beliefs following a deviation to Not Attend so that this constraint is satisfied for *every* sender-type. Determining the precise parameters that lead to existence is a linear programming problem with no closed form solution. However, it is possible to establish the following results and comparative statistics.

Proposition 1. *The following are true*[11]

(1) For a given vectors $\mathbf{v}$ *and* $\boldsymbol{\rho}$*, there exists* $M^*(\mathbf{v},\boldsymbol{\rho})$ *such that there is a full participation equilibrium if and only if* $M \geq M^*(\mathbf{v},\boldsymbol{\rho})$

(2) $M^*(\widetilde{\mathbf{v}},\boldsymbol{\rho}) \geq M^*(\mathbf{v},\boldsymbol{\rho})$ *for any* $\widetilde{\mathbf{v}} \geq \mathbf{v}$

(3) $M^*(\mathbf{v},\widetilde{\boldsymbol{\rho}}) \geq M^*(\mathbf{v},\boldsymbol{\rho})$ *for any* $\widetilde{\boldsymbol{\rho}} \geq \boldsymbol{\rho}$

(4) If $M = M^*(\mathbf{v},\boldsymbol{\rho})$*, Equation 4.2 is binding for sender-type* $\omega = (0,0,...,0)$

The first part of the proposition notes that there is a critical threshold for the match values above which a full participation equilibrium exists. The second and third parts establish that the critical value rises (effectively making full participation harder to sustain) as any trait becomes more sensitive (higher v_j) or more homogenous (higher ρ_j) in the population. The fourth part notes that the participation constraint of the rarest type is hardest to satisfy. Intuitively, the rarest type reveals the most information in a full participation equilibrium and has the lowest chance of finding a match and therefore is most likely to deviate.

We note that the full-participation equilibrium might require complicated beliefs to incentivize each sender-type to attend the conversation. One might be concerned that bounded rationality limits the ability of the receivers to form such beliefs. Alternatively, it is possible that, in real life, there are many exogenous reasons that a player might not attend a conversation, leading to less extreme receiver updating. To capture these situations, we define the *No Inference case*:

Definition 2. ***No Inference case***: *If the sender chooses Not Attend, the receiver's beliefs equal the prior belief:*

$$\mu(\omega_j^S = 1 | h_{NA}) = \rho_j$$

In the No Inference case, the receiver does not make any additional inferences when a sender chooses to Not Attend a conversation. Reassuringly, all of the statements of proposition 1 continue to hold in the No Inference case. Additionally, it is possible to obtain a closed form solution for $M^*(\mathbf{v},\boldsymbol{\rho})$:

Proposition 2. *In the No Inference Case,* $M^*(\mathbf{v},\boldsymbol{\rho}) = \frac{\sum_{j=1}^{N} \rho_j v_j}{\prod_{j=1}^{N}(1-\rho_j)}$

[11] Claim 2 may not hold (i.e., $M^*(\mathbf{v},\boldsymbol{\rho})$ responds ambiguously to changes in v) if the sender only suffers the information penalty if the receiver does not choose Match. See appendix D for details.

The minimum match value M required for a full-participation equilibrium to exist must be (weakly) lower in the Inference case than in the No Inference case. To see this, note that by setting beliefs in the Inference case to the prior beliefs, we mimic the No Inference case. The freedom to further modify these beliefs can only make it easier to satisfy equation 4.1.

4.3 An Example Equilibrium: Taboos

There are many potential equilibria that can occur if the full participation equilibrium does not exist. We close this section with a simple two trait example where agents with the rare realization of the first trait choose Not Attend. In other words, ω_1 is a *taboo* trait in the sense that no one with a specific realization of that trait attends the conversation. Although our example is computed under the Inference case, the non-existence of the full-participation equilibrium and the existence of the taboo equilibrium continue to hold in the No Inference case as well.

Example 1 (Taboos under the Inference case):

Assume N=2, $\rho_1 = 0.75$, $\rho_2 = 0.5$, $v_1 = 1$ and $v_2 = 2$

If $M \geq 8$, then a full-participation equilibrium exists with the off-path beliefs $\mu(\omega_1^S = 1|h_{NA}) = 0$ and $\mu(\omega_2^S = 1|h_{NA}) = 0.5$. The payoffs are

	Attend	Not Attend
[1 1]	$0.375 * M - 3$	-1
[1 0]	$0.375 * M - 3$	-1
[0 1]	$0.125 * M - 3$	-2
[0 0]	$0.125 * M - 3$	-2

If $M < 8$, then full-participation cannot be sustained using any off-path beliefs. There is an equilibrium in which [0 1] and [0 0] do not attend the conversation. The payoffs for this equilibrium match those in the above table.

Checking the existence of a taboo equilibrium is usually more complicated than in our simple example. For sender types that are (in equilibrium) expected to attend the conversation, we must check that deviating to either Not Attend (consequently receiving a benefit from “tricking” some receivers about her type) or sending an unexpected off-path message is not profitable. For sender types that are (in equilibrium) expected choose Not Attend, we must check that deviating to sending a message is not profitable. In Example 2, setting beliefs $\mu(\omega_1^S = 1|h) = 0$ and $\mu(\omega_2^S = 1|h) = 0.5$ at any off-path history eliminates all of these potential profitable deviations.

5 Mediated Conversation

The primary conclusion of section 4 is that the presence of high information value and rare trait realizations can eliminate full participation equilibria that allow for efficient matching. In this section, we discuss how the addition of a third party mediator can re-establish full participation. In the real-world, communication is often mediated. Common examples include investment banks that attempt to match a client with possible acquisition targets and match-makers who pair couples with a high potential for marriage.

In the one-stage game, each sender-type suffers a large information penalty because she must reveal all of her information to all receiver-types. Optimally, a given sender-type would only reveal information to the receiver if the receiver has a matching type. This is, of course, impossible because the sender cannot determine the receiver's type without revealing a large amount of information to the receiver.

In the mediated conversation, we assume that the mediator observes the sender's type ω^S and receiver's type ω^R, but that this knowledge does not lead to information penalties. After observing the types, the mediator sends a message directly to the receiver from some finite space of messages $\mathcal{M}' = \{m_1^{'}, m_2^{'}, m_3^{'}...\}$ using the probability distribution $f(\mathcal{M}'|\omega^S, \omega^R)$. After observing the message, the receiver makes a decision to Match or End the game. Focusing on full participation, a mediator mechanism that maximizes ex-ante social welfare is surprisingly simple:

Proposition 3. *A socially optimal mediator mechanism with optimal matching uses two messages. If the types of the sender and the receiver match, the receiver observes the first message. Otherwise, the receiver observes the second message. The expected information penalty of a sender of type* ω^S *is:*

$$-\Pr(\omega^S)\textstyle\sum_{j=1}^N v_j - \sum_{\omega' \in (\Omega \backslash \omega^S)} Pr(\omega') \left[\sum_{j=1}^N v_j \mu(\omega_j^S = \omega_j^{S*} | \omega^S \neq \omega')\right] \qquad (5.1)$$

In this optimal mechanism, the mediator simply tells the players if they match or not. Given any two distinct messages issued with positive probability when the agents' types do not match, it is always socially beneficial for the mediator to combine these messages into a single joint message. The joint message releases less information than the two distinct messages, reducing the expected information penalty.

Intuitively, in any mechanism with full participation, players must at least learn if they have matching types. In the optimal mechanism, the mediator acts as an "information sink" by revealing exactly this information. This low level of information revelation is impossible in the non-mediated game because more specific information must be revealed before players can determine if there is a match.

6 Dynamic Message Exchange: Conversations

In section 4 we discussed how privacy preferences can lead some players to not participate in conversations, leading to inefficient matching. In section 5, we proposed the use of a mediator to restore the existence of a full participation equilibrium. However, this requires a third party to whom the sender is willing to reveal all of her information. In this section, we discuss a dynamic conversation protocol that increases the potential for full participation but does not require a third-party. In this conversation, the sender dynamically screens out non-matching receivers over time, which limits the expected information penalty of all sender-types.

We first define some useful theoretical objects in our dynamic model. Then we prove that the full set of PBEs is extremely large, which is not surprising for a multi-stage communication model with signaling. Therefore, we restrict our attention to ex-post Pareto optimal equilibria. This restriction removes equilibria in which there is another equilibrium that makes some player-type strictly better off, while leaving the other player-types at least weakly better off.

We then discuss a example of a seemingly unrealistic equilibrium that is ex-post Pareto optimal. In this equilibrium, senders exploit the fact there is no information penalty if receivers learn information about the sender's *type* ω^S without learning any information about the sender's *traits* ω_j^S because our model does not penalize for revealing information about correlations between traits. While this example is mathematically interesting, it presents behavior that seems very unlikely in the real world.

Therefore, we focus on equilibria that satisfy *block inference*, which requires any information revealed about *type* to be contained in the information revealed about *traits*, but still allows for signaling. Somewhat surprisingly, within this set of equilibria there is one equilibrium that (weakly) ex-post Pareto dominates all other equilibria. This is a very strong sense of efficiency as *all types of both players* (weakly) prefer this equilibrium to all other equilibria. Therefore, the equilibrium also (weakly) ex-ante Pareto dominates all other equilibria as both the sender and receiver receive (weakly) higher *expected* payoffs from this equilibrium before learning their respective types.

6.1 Setup and Examples

The sender's messaging strategy defines the set of traits that will be verifiably revealed in each stage of the game. Specifically, given a sender strategy σ, define $\widetilde{m}_1$ as $\sigma(h_0)$ and recursively define $\widetilde{m}_t$ as $\sigma(\widetilde{m}_1, \widetilde{m}_2, ...\widetilde{m}_{t-1})$. A sender-type's *grammar* g is 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. A grammar that contains every trait is referred to as a *complete grammar.*

The sender's message has two effects on the receiver's beliefs. First, the message conveys verifiable information about traits. Second, the choice of message can signal the value of traits that are not directly revealed by the message. These effects are demonstrated in the next two examples.

Example 2 (Equilibrium Inference):

Assume that N is a multiple of 2.
Consider the following potential equilibrium behavior in stage t :

Types where $\omega^S_{2*t-1} = \omega^S_{2*t}$	Stage t: Reveal trait ω^S_{2*t-1}
Types where $\omega^S_{2*t-1} \neq \omega^S_{2*t}$	Stage t: Reveal trait ω^S_{2*t}

Then after the first stage:

If $m_1 = \{\omega_1 = 0\}$	then $\mu(\omega^S_1 = 1 \vert m_1) = 0$	and $\mu(\omega^S_2 = 1 \vert m_1) = 1$
If $m_1 = \{\omega_1 = 1\}$	then $\mu(\omega^S_1 = 1 \vert m_1) = 1$	and $\mu(\omega^S_2 = 1 \vert m_1) = 0$
If $m_1 = \{\omega_2 = 0\}$	then $\mu(\omega^S_1 = 1 \vert m_1) = 0$	and $\mu(\omega^S_2 = 1 \vert m_1) = 0$
If $m_1 = \{\omega_2 = 1\}$	then $\mu(\omega^S_1 = 1 \vert m_1) = 1$	and $\mu(\omega^S_2 = 1 \vert m_1) = 1$

Example 3 (Equilibrium Inference):

Assume that $N = 2$ and that $\rho_1 = .8$ and $\rho_2 = .6$.

Consider the following potential equilibrium behavior:

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

Then after the first stage:

If $m_1 = \{\omega_1 = 0\}$ then $\mu(\omega_1^S = 1|m_1) = 0$ and $\mu(\omega_2^S = 1|m_1) = .6$
(as [0 1] and [0 0] are the players that send that message in equilibrium.)
If $m_1 = \{\omega_1 = 1\}$ then $\mu(\omega_1^S = 1|m_1) = 1$ and $\mu(\omega_2^S = 1|m_1) = 1$
(as [1 1] is the only player who sends that message in equilibrium.)
If $m_1 = \{\omega_2 = 0\}$ then $\mu(\omega_1^S = 1|m_1) = 1$ and $\mu(\omega_2^S = 1|m_1) = 0$
(as [1 0] is the only player who sends that message in equilibrium.)
If $m_1 = \{\omega_2 = 1\}$ then $\mu(\omega_1^S = 1|m_1) \in [0, 1]$ and $\mu(\omega_2^S = 1|m_1) \in [0, 1]$
(as this action is off the equilibrium path.)

6.2 The (Large) Set of Equilibria

In this section we show that there are a large number of potential equilibria given a sufficiently high match value M. In fact, any sender strategy which assigns a complete grammar to each sender-type can be an equilibrium.

Proposition 4. *Any sender strategy using a complete grammar for every sender-type can be supported in a perfect Bayesian equilibrium with full participation for sufficiently large M.*

To construct each equilibrium, 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 End following a sender's deviation. As long as M is large enough to off-set the information penalties, all sender-types prefer to follow their assigned grammar (and have a chance at matching) than deviate (and have no chance at matching).

As with most dynamic signaling models, some of these equilibria involve unrealistic behavior and elaborate signaling schemes. In many cases, all agents obtain strictly higher payoffs in some other equilibrium, but the sender-types do not attempt to deviate to such an equilibrium because receivers would choose End in response. Therefore, we focus on equilibria which are ex-post Pareto optimal. However, as we demonstrate in the following section, this leaves a set of equilibria that, while mathematically interesting, are unrealistic. Therefore, we further refine the set of equilibria in section 6.4.

6.3 Undominated but Unrealistic Equilibria

Example 4 describes an equilibrium that is ex-post Pareto optimal, but is seemingly unrealistic. In this example, each initial 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 about the probabilities of traits one, three, and four.

Example 4 (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 comprises an ex-post 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 vast majority of receivers (in this case, $\frac{7}{8}$ of receivers). In general, other ex-post 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.

While this result is mathematically interesting, there are many arguments against using these equilibria for behavioral predictions. Most clearly, the optimality of these strategies is an artifact of our parameterization of the penalty function. In example 5, a sender of type [0 1 0 0] only suffers an information penalty for revealing the second trait since, although the other traits are perfectly correlated, the marginal probability the receiver's posterior places on the individual realizations of the other traits is unchanged.

In order to create a tractable model of privacy with different information values for different traits, our model assumes privacy is solely a function of marginal beliefs about individual traits and not about correlation between traits, although these correlations may be an interesting source of privacy loss in some settings.[12]

6.4 A Refinement on Beliefs: Block Inference

In order to remove the unrealistic behavior noted in Example 5, we introduce an equilibrium refinement. Prior work restricts the inferences that can be made by, for example, exogenously requiring agents to reveal information in a prespecified order (Stein [22], Dziuda and Gradwohl [6]). Our refinement admits a richer variety of signaling phenomena than prior works.

Our principal goal is to illustrate how dynamic message exchange can reduce the expected information penalties from choosing Attend and alleviate the frictions caused by a preference for privacy. At a minimum, our use of a refinement implies that we are providing a lower bound for the benefit of dynamic conversations. Although equilibria that do not satisfy our refinement can be counter-intuitive, considering these additional equilibria can only imply greater benefits to dynamic conversations.

Specifically, we choose to examine equilibria that exhibit a property we call *block inference.* Loosely speaking, in an equilibrium that satisfies block inference the information revealed about the sender's type can be described solely in terms of the marginal probability of each of the sender's traits. That is, if a trait realization has not been revealed (either through a verifiable message or signaling), the receiver must believe its realization is uncorrelated with the realization of any other trait.[13]

Definition 3. *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 $\mu_R(\omega_j^S = 1|h) = \rho_j$.*

[12] An alternative model could have instead based preferences over the posterior likelihood of an agent's true type (as opposed to the individual traits). This would eliminate the issue with correlations. However, many of the signature predictions of our model are based on our formulation of types as a set of traits that can have different variances and information values. Without a types-as-traits framework, it is unclear if analogs of our results could even be stated (much less proven).

[13] Since we are focusing on equilibria in pure strategies, under generic values of ρ_i complete revelation of a trait is required for beliefs about traits to be uncorrelated

We note that examples 2 and 3, both of which exhibit nonverifiable signaling, satisfy block inference. Block inference requires that all signaled information be fully resolved within the stage the signal is issued. Block inference allows for the bulk of traits to be revealed by signaling instead of by verifiable messages, which we show formally in Appendix C.

6.5 Optimal Dynamic Equilibrium

There are still many equilibria of our game that satisfy block inference. In the following sections, we demonstrate that one particular equilibrium (weakly) ex-post Pareto-dominates all others within this set. As all types of both players (weakly) prefer this equilibrium to all other equilibria, we call this the *Pareto Optimal Dynamic Equilibrium (PODE).* In the PODE, all senders follow the same grammar. In this grammar, one trait is revealed in each stage, and traits are revealed in order of increasing information value. The universal preference for the PODE implies that the equilibrium is coalition-proof. To the extent that senders can break grammars by credibly (and collectively) insisting on an interpretation of their messages, our theorem shows that such a pre-conversation message is credible.

The section proceeds in two parts. First, we define the strategies in the PODE. Stated informally, at any history in the equilibrium, all types of senders reveal the lowest information value trait that has not been revealed thus far. Second, we briefly outline the proof showing that the PODE is preferred by all types of both players, which provides the intuition behind the uniform optimality of this equilibrium.

To formally define the PODE, we need to define the behavior of each sender-type at each history, $\sigma^*(\omega, h)$, and the beliefs of receivers at each history, $\mu^*(h)$. In the PODE, all senders reveal the one trait with the lowest information value that has not been revealed yet. Recalling that traits are indexed by information value, we define $m(h)$ as a function that maps a history to the message revealing the lowest index trait not previously revealed in h

$$\sigma^*(\omega, h) = m(h)$$

Since this strategy generates a grammar that is independent of the sender-type, receivers only update based on the verifiable information contained in the message. That is, when the sender sends the message $\{\omega_3 = 1\}$, the receiver's only inference is that $\omega_3^S = 1$. We call this *straightforward inference*, which we define explicitly for use later in the section:

Definition 4. *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.*

The following proposition states that this equilibrium exists as long as M is large enough and is (weakly) preferred by every type of both players to any other equilibrium.

Proposition 5. *$\sigma^*(\omega, h)$ and $\mu^*(h)$ is a PBE for a sufficiently large M. If this equilibrium exists, it provides a weakly higher payoff to all sender-types and receiver-types than any equilibrium that satisfies block inference. Furthermore, this equilibrium provides a strictly higher payoff for some type of sender than any equilibrium that satisfies block inference and is not outcome equivalent to $\sigma^*(\omega, h)$ and $\mu^*(h)$.*

This uniformity of preferences is surprising given that the senders' preferences do not obey the conditions (e.g., single crossing) usually required to well-order the actions taken in a signaling game. Furthermore, it is interesting that the ordering depends solely on the information value, v_j, and not the rarity of a sender-type's realization of trait j. A natural (and incorrect) conjecture is that players would prefer to reveal traits in an order that minimizes the expected increase in their information penalty, which would mean that the preferred order of trait revelation would depend both on v and the probability of an agent's particular trait realization. For example, suppose $\rho_1 = .6$ and $\rho_2 = .8$, and v_1 is only slightly smaller than v_2. Our (false) conjecture would suggest that player [1 1] would rather reveal $\omega_2 = 1$ if forced to reveal one trait. Since receivers are already confident about the realization of ω_2 for the sender, in the event of a failure to match this would cause a smaller decrease in the information penalty (i.e., a less negative information penalty) than revealing ω_1, a trait about which the receivers have less information.

If this conjecture were correct, senders of different types would generally not agree on the optimal grammar for all senders to employ. However, this logic ignores the dynamic benefit of revealing rare trait realizations: 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 revealing that $\omega_1 = 1$ immediately ends 40% of the conversations (rather than 20% in the case of $\omega_2 = 1$), which means that no more information will be given about the sender's type to 40% of the receivers. The proof of proposition 5 demonstrates that since these forces balance, the sender's sole concern is the information value of the trait.[14] As information about a trait becomes more sensitive (v rises), the sender receives a higher information penalty from revealing that trait. However, there is no dynamic benefit of revealing high value traits early. As a result, players prefer to reveal more sensitive information later in the conversation.

We now provide a sketch of the proof of proposition 5 in order to build intuition about sender preferences over potential equilibria and establish the uniform preference of our equilibrium for all sender-types. We start with an arbitrary equilibrium, which we denote (σ, μ). We then modify the equilibrium strategies and beliefs in three steps, leading to the PODE (σ^*, μ^*). In each step, we show that all sender-types are weakly better off. The sender strategy and receiver belief pairs constructed at each step may not be equilibria until the final step where we reach (σ^*, μ^*).

The basic intuition we harness throughout the proof is that senders wish to reveal as little information as possible in each message since this minimizes the information penalty faced in the event that the sender and receiver fail to match at that stage. In the first step, we modify the actions such that senders at terminal nodes explicitly reveal all traits that were revealed to the receiver through signaling and insist the receivers use straightforward inferences. This leads to (σ', μ'), and senders are indifferent between (σ', μ') and (σ, μ) because both lead to the same matching opportunities and yield the same information penalty in the event that a match does not occur. In the second step, we modify (σ', μ') by replacing any messages that reveal multiple traits with a set of sequential one-trait messages that reveal the same set of traits. Again, forcing straightforward inferences for receivers, this yields (σ'', μ''). Senders prefer (σ'', μ'') to (σ', μ') because revealing one trait at a time prevents non-matching receiver-types from learning about the other traits revealed in the original message, which reduces the information penalty. Finally, we modify the ordering of the revealed traits such that traits with lower information value are revealed first, which leads to (σ''', μ''') = (σ^*, μ^*). Senders prefer (σ^*, μ^*) to (σ'', μ'') because traits with higher information value are revealed later when fewer receiver-types remain in the conversation.

[14] While this feature is a result of our linear parameterization, we show the result holds with slightly non-linear π functions in section B.3 and discuss what can occur when π is significantly concave or convex.

6.6 Examples of PODEs

To further build intuition about the logic of the universal preference of all player-types for the PODE, we present an example of a specific senders-type's preferences over grammars given straightforward inference. The payoffs to sequential revelation of traits is always greater than simultaneous revelation, and the preferred ordering depends only on v_1 and v_2.

Example 5 (Preferences over grammars):

Assume $N = 2$, $\rho_1 = .8$ and $\rho_2 = .6$. Focus on sender-type [1 1].
Consider payoff under straightforward inference:
(Probabilities of facing receiver-types [1 1], [1 0], [0 1], and [0 0] are .48,.32,.12, and .08)

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

All receiver-types learn the sender's type.
For these types, $\mu(\omega_1^S = \omega_1^{S*}) = 1$ and $\mu(\omega_2^S = \omega_2^{S*}) = 1$
E[info penalty]: $= -.48(v_1 + v_2) - .32(v_1 + v_2) - .12(v_1 + v_2) - .08(v_1 + v_2)$
$= -v_1 - v_2$

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

Receiver-types [0 1] and [0 0] drop out after first stage, learn sender's 1st trait
For these types, $\mu(\omega_1^S = \omega_1^{S*}) = 1$ and $\mu(\omega_2^S = \omega_2^{S*}) = .6$
Receiver-types [1 1] and [1 0] learn the sender's type.
For these types, $\mu(\omega_1^S = \omega_1^{S*}) = 1$ and $\mu(\omega_2^S = \omega_2^{S*}) = 1$
E[info penalty]: $= -.48(v_1 + v_2) - .32(v_1 + v_2) - .12(v_1 + .6v_2) - .08(v_1 + .6v_2)$
$= -v_1 - .92v_2$

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

Receiver types [1 0] and [0 0] drop out after first period, learn sender's 2nd trait
For these types, $\mu_R(\omega_1^S = \omega_1^{S*}) = .8$ and $\mu_R(\omega_2^S = \omega_2^{S*}) = 1$
Receiver types [1 1] and [0 1] learn the sender's type.
For these types, $\mu_R(\omega_1^S = \omega_1^{S*}) = 1$ and $\mu_R(\omega_2^S = \omega_2^{S*}) = 1$
E[info penalty]: $=-.48(v_1 + v_2) - .32(.8v_1 + v_2) - .12(v_1 + v_2) - .08(.8v_1 + v_2)$
$= -.92v_1 - v_2$

Payoffs from Grammars 2 and 3 are larger than Grammar 1 regardless of v_1 or v_2
Payoff from Grammar 2 is larger than Grammar 3 $\Leftrightarrow v_2 > v_1$

Finally, we present an example of the PODE with three traits in order to demonstrate a longer conversation, using a format designed to make the path-of-play explicit.

Example 6: (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: $.144M - v_1 - v_2 - v_3$

7 Benefit of Dynamic Conversation

In this section, we discuss the benefits of dynamic screening in achieving full participation relative to the one-stage and mediator games. Recall that proposition 1 establishes the existence of a full-participation equilibrium in our static game if the match payoff is at least some minimal value $M^*_{Static}(\boldsymbol{\rho}, \mathbf{v})$ that is a function of $\boldsymbol{\rho}$ and $\mathbf{v}$. It is simple to show that there exist similar parameters $M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v})$ and $M^*_{Mediator}(\boldsymbol{\rho}, \mathbf{v})$ for the dynamic and the mediator game. We evaluate the size of the privacy friction in each game by the respective value of M^*. Our previous results imply

$$M^*_{Static}(\boldsymbol{\rho}, \mathbf{v}) \geq M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v}) \geq M^*_{Mediator}(\boldsymbol{\rho}, \mathbf{v})$$

In this section we provide some results on the relative size of these frictions. Our principal results show that dynamic communication does significantly better than static communication, particularly when rare traits are very rare, some traits are very valuable, or a large number of traits need to be revealed. Moreover, dynamic communication reduces the welfare loss from privacy quickly as the length of the conversation grows, which means the frictions caused by a preference for privacy can be thought of as small under dynamic communication both relative to situations with a mediator and to a no-privacy-loss benchmark.

The next proposition compares how the difference between $M^*_{Static}(\boldsymbol{\rho}, \mathbf{v})$ and $M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v})$ changes with $\boldsymbol{\rho}$ and $\mathbf{v}$. Throughout this section we study the No Inference case in order to obtain closed form solutions.

Proposition 6. *The following comparative statics hold*

(1) $M^*_{Static}(\boldsymbol{\rho}, \mathbf{v}) - M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v})$ is increasing in $\boldsymbol{\rho}$ and v

(2) For an arbitrary v, any trait i, and $\varepsilon > 0$ define $\widetilde{v}$ where

$$\begin{aligned} \widetilde{v}_j &= v_j \text{ if } j \neq i, i+1 \\ \widetilde{v}_i &= v_i - \varepsilon \\ \widetilde{v}_{i+1} &= v_{i+1} + \varepsilon \end{aligned}$$

and assume $\widetilde{v}_i$ increases in i. Then we have $M^*_{Dynamic}(\boldsymbol{\rho}, \widetilde{\mathbf{v}}) \leq M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v})$ and $M^*_{Static}(\boldsymbol{\rho}, \widetilde{\mathbf{v}}) = M^*_{Static}(\boldsymbol{\rho}, \mathbf{v})$.

Our first claim implies that the benefit of dynamic conversation is particularly large when ρ or v is high. If ρ is high, then the screening power of revealing a 0 realization is high, which means agents with rare trait realizations are likely to avoid revealing high value information in a dynamic conversation. When v is large, large welfare gains are caused by only revealing high information value traits to those select few receiver with whom a sender is most likely to match.

The second claim of proposition 6 is stated in a limited fashion for expositional ease. This result implies that a transfer of information value from any low value trait to a higher information value trait lowers $M^*_{Dynamic}$ (making full participation more viable), but leaves M^*_{Static} is unchanged (as the total information value of the traits is constant). Therefore dynamic conversations are particularly useful (relative to static information exchange) when there is greater heterogeneity in the information value of the traits, because the high-value traits can be discussed later in the conversation, at which point many non-matching receiver-types have already been screened out.

Finally, we discuss the speed at which dynamic conversations can reduce the information penalty and alleviate the economic frictions caused by a preference for privacy. Note that even if the rarest sender-type (who determines $M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v})$) chooses to Not Attend, she will suffer an average information penalty per trait[15] equal to

$$-\frac{1}{N}\textstyle\sum_{i=1}^{N}(1-\rho_i)v_i$$

We show that the average information penalty per trait in a dynamic conversation approaches this level at a rate $O(\frac{1}{N})$. Recollect that the rarest sender-type's expected information penalty from Attending the dynamic conversation is

$$-\textstyle\sum_{i=1}^{N} E\left[\mu_R\left(\omega_i^S=0|h\right)\right]v_i$$

where h, the point at which the conversation stops, is a random variable dependant on the receiver's type, which is unknown to the sender when she makes her attendance choice.

Proposition 7. *Consider a sequence of traits with associated parameters (ρ_i, v_i) and a conversation with N traits includes the first N traits in this sequence. Assume that v_i is bounded.*[16] *Then* $\frac{1}{N}\sum_{i=1}^{N}\left\{E\left[\mu_R\left(\omega_i^S=0|h\right)\right]-(1-\rho_i)\right\}v_i = O\left(\frac{1}{N}\right)$

[15] We use the average information penalty per trait to ease comparison of the welfare loss from privacy in conversations of different lengths.

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

Our bound is essentially tight. Consider a case where $(\rho_i, v_i) = \left(\frac{1}{2}, 1\right)$. In a dynamic 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.

Our result implies that the information frictions in dynamic conversations are small for conversations of moderate length. In other words, if N is of moderate size, then switching from a static to a dynamic conversation allows the sender to recover most of the previously lost opportunities to profitably interact with the receiver. Finally, our result implies that the additional advantage of having a mediator relative to a dynamic conversation is small for N of moderate size.

8 Extensions

We now discuss how weakening some of our assumptions affects the model outcomes. We provide overviews of our extensions here, while the full discussion is contained in Appendix B.

Our first extension considers situations in which successful matches require matching on some (but not necessarily all) traits. We show that the PODE discussed above continues to provide the unique Pareto Optimal outcome. We use our results to describe the full set of PODEs.

Our second extension provides a partial characterization of our model when sending messages is costly. When the cost of sending messages is relatively 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 that 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 of growing length as the conversation proceeds.

Our third extension focuses on alternative information penalty functions: $\pi(\cdot)$. 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 that sender-type's 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.

9 Conclusion

Our paper analyzes situation in which agents must exchange information to discover whether they can productively match, but the agents have a preference for privacy preferring 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 when the preference for privacy must eliminate some profitable opportunities to interact, study the structure of communication, and identify both the quantity and the kind of information disclosed as the conversation progresses.

We first show that in a setting where a single round of communication is possible, the economic frictions caused by a preference for privacy can be large. The rarest type of agent is the most tempted to Not Attend a conversation, as this agent has the most to lose in terms of lost privacy and the least as gain as finding a match is unlikely. This can destroy the existence of an equilibrium with full participation. Alternative equilibria can include taboos, which are traits that are never discussed in equilibrium because individuals with the taboo trait never choose to Attend the conversation.

We next show how the use of a mediator can mitigate the welfare loss from a concern for privacy. We show that a mediator is the first-best mechanism, but that dynamic communication can do almost as well for conversations of moderate length. Moreover in conversations of moderate length, dynamic communication can eliminate most of the frictions present in the strategic case.

To study the effects of dynamic communication, we first provide a description of the sender-optimal grammar employed to structure the conversation. This grammar involves the delayed revelation of information with more sensitive data disclosed once the agents have become more confident that a profitable match is possible. The order in which information is revealed depends only on the information value of the traits and not on the rarity of the traits. This surprising result 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 end the conversation with high probability, which reduces the expected information penalties for traits revealed in later stages.

We provide a number of extensions of our benchmark model. First, we argue that our results extend to the case where the sender and receiver need only match on a subset of the traits to profitably match. Second, we describe the outcome when messages are costly, which creates a tension between the incentive to dynamically screen out non-matching receivers and the desire to reduce message costs. Third, we argue that our results are robust to moderate changes in the utility functions of the agents and discuss the complexity of analyzing our model under general preference structures. Finally, we analyze the incentive issues that arise when the messages are not verifiable, and we show that if L is sufficiently large that truthfulness can be preserved.

While our principal goal is to study how preferences for privacy influence the structure and timing of information exchange, in future work we hope to incorporate our analysis into more general models that would allow us to endogenize the match value. 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 agents.

We believe that our reduced-form characterization of privacy concerns is portable to a wide variety of economic settings that do not involve explicit information exchange. For example, privacy is often a concern in auction settings where the bidders may interact later. Spectrum auctions usually involve a small number of nation-wide bidders that are engaged in competitive interactions that can span decades, and the information about corporate strategy revealed in an auction could be critically important. Our model can present clues as to how the bidders' preference for privacy can modify the structure of the optimal auction.

A preference for privacy may also make it difficult to truthfully solicit the opinions of voters or consumers through surveys. This problem has been tackled in the computer science literature, but only through the lens of differential privacy. Our model may allow to provide useful analyses of the tension between the welfare costs of privacy and survey accuracy, which may provide insights into survey techniques.

References

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

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

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

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

[5] 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.

[6] Dziuda, W. and R. Gradwohl (2012) “Achieving Coordination Under Privacy Concerns,” *mimeo.*

[7] Ganglmair, B. and Tarantino, E. (2013) “Conversation with Secrets,” *mimeo.*

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

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

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

[11] Gradwohl, R. (2012) “Privacy in Implementation,” *mimeo.*

[12] Honryo, T. (2011) “Dynamic Persuasion,” *mimeo.*

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

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

[15] 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.

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

[17] 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.

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

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

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

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

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

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

APPENDIX: For Online Publication Only

A Proofs

Proposition 1. *The following are true*

(1) For a given vectors $\mathbf{v}$ *and* $\boldsymbol{\rho}$*, there exists* $M^*(\mathbf{v},\boldsymbol{\rho})$ *such that there is a full participation equilibrium if and only if* $M \geq M^*(\mathbf{v},\boldsymbol{\rho})$

(2) $M^*(\widetilde{\mathbf{v}},\boldsymbol{\rho}) \geq M^*(\mathbf{v},\boldsymbol{\rho})$ *for any* $\widetilde{\mathbf{v}} \geq \mathbf{v}$

(3) $M^*(\mathbf{v},\widetilde{\boldsymbol{\rho}}) \geq M^*(\mathbf{v},\boldsymbol{\rho})$ *for any* $\widetilde{\boldsymbol{\rho}} \geq \boldsymbol{\rho}$

(4) If $M = M^*(\mathbf{v},\boldsymbol{\rho})$*, Equation 4.2 is binding for sender-type* $\omega = (0, 0, ..., 0)$

Proof. First we prove claim (1), that for each $(\mathbf{v},\boldsymbol{\rho})$ there is some minimal M for which the set of constraints given by equation 4.2 can be satisfied. For a particular value of $(\mathbf{v},\boldsymbol{\rho})$, let $\mathcal{M}$ be the set of M such these constraints can be satisfied. Let $M^*(\mathbf{v},\boldsymbol{\rho}) = \inf \mathcal{M}$. Since the equations are weak inequalities, it is clear that $M^*(\mathbf{v},\boldsymbol{\rho}) \in \mathcal{M}$. Finally, if $\mu(\circ|h_{na})$ are off-path beliefs such that the constraints are satisfied for all types at $M^*(\mathbf{v},\boldsymbol{\rho})$, then $\mu(\circ|h_{na})$ also satisfies the set of equations 4.2 for any $M \geq M^*(\mathbf{v},\boldsymbol{\rho})$.

Second we prove that if an equilibrium exists, then we can describe an equilibrium where type $\omega = (0, 0, ..., 0)$ has the lowest payoff, which implies claim (4). We then use this observation to show that as v or ρ_j increases, the constraint for this lowest type binds less tightly, which establishes claims (2) and (3).

Suppose we have a full participation equilibrium with efficient matching with off-path beliefs $\mu(\circ|h_{na})$. Suppose for some $\widehat{j}$ we have $\mu(\omega_{\widehat{j}} = 1|h_{na}) > \frac{1}{2}$. In other words, the off-path-beliefs suggest types with a common trait realization are more likely to deviate to Not Attend than agents with the rare trait realization. Consider two sender-types ω^S and $\widetilde{\omega}^S$ that differ only on trait $\widehat{j}$. Formally let $\omega^S_j = \widetilde{\omega}^S_j$ for $j \neq \widehat{j}$, $\omega^S_{\widehat{j}} = 1$, and $\widetilde{\omega}^S_{\widehat{j}} = 0$. We define $\rho^*_j = \rho_j$ if $\omega^S_j = 1$ and $\rho^*_j = 1 - \rho_j$ if $\omega^S_j = 1$.

With this notation out of the way, the existence of the equilibrium implies for types ω^S and $\widetilde{\omega}^S$ respectively

$$\begin{array}{rcl} M * \rho_{\widehat{j}} * \prod_{j=\widehat{j}} \rho^*_j & \geq & (1 - \mu(\omega^S_{\widehat{j}} = 1|h_{na}))v_{\widehat{j}} + \sum_{j=\widehat{j}}(1 - \mu(\omega^S_j = \omega^{S*}_j|h_{na}))v_j \\ M * (1 - \rho_{\widehat{j}}) * \prod_{j=\widehat{j}} \rho^*_j & \geq & \mu(\omega^S_{\widehat{j}} = 1|h_{na})v_{\widehat{j}} + \sum_{j=\widehat{j}}(1 - \mu(\omega^S_j = \omega^{S*}_j|h_{na}))v_j \end{array}$$

By assumption $M * \rho_{\widehat{j}} * \prod_{j=\widehat{j}} \rho^*_j \geq M * (1 - \rho_{\widehat{j}}) * \prod_{j=\widehat{j}} \rho^*_j$ and $(1 - \mu(\omega^S_{\widehat{j}} = 1|h_{na}))v_{\widehat{j}} \leq$

$\mu(\omega^S_{\hat{j}} = 1|h_{na})v_{\hat{j}}$. But then this implies that

$$\begin{aligned} M * \rho_{\hat{j}} * \textstyle\prod_{j=\hat{j}} \rho^*_j &\geq \frac{1}{2}v_{\hat{j}} + \textstyle\sum_{j=\hat{j}}(1-\mu(\omega^S_j = \omega^{S*}_j|h_{na}))v_j \\ M * (1-\rho_{\hat{j}}) * \textstyle\prod_{j=\hat{j}} \rho^*_j &\geq \frac{1}{2}v_{\hat{j}} + \textstyle\sum_{j=\hat{j}}(1-\mu(\omega^S_j = \omega^{S*}_j|h_{na}))v_j \end{aligned}$$

which would be the participation for an equilibrium where the off-path belief was identical except for the change $\mu(\omega^S_{\hat{j}} = 1|h_{na}) = \frac{1}{2}$. Since the number of traits is finite, induction implies we can construct an equilibrium where for each j we have $\mu(\omega^S_{\hat{j}} = 1|h_{na}) \leq \frac{1}{2}$.

This first step allows us to focus on equilibria where for each j we have $\mu(\omega^S_{\hat{j}} = 1|h_{na}) \leq \frac{1}{2}$. If such a full participation equilibrium does not exist, then there can be no full participation equilibrium. In any such equilibrium, the tightest participation constraint is for a type $(0, 0, ..., 0)$, which can be written

$$M * \textstyle\prod_{j=1}^N (1-\rho_j) \geq \textstyle\sum_{j=1}^N (1-\mu(\omega^S_{\hat{j}} = 1|h_{na}))v_j$$

Of all of the possible types, the left side of the equation takes the lowest and the right side the highest value. For any choice of $\mu(\omega^S_{\hat{j}} = 1|h_{na})$ this equation becomes harder to satisfy (for any off-path-belief) as v or ρ increases. This implies that the smallest M such that equilibrium can be supported (i.e., $M^*(\mathbf{v}, \boldsymbol{\rho})$) must be weakly increasing in v and ρ, which yields claims (2) and (3). □

Proposition 3. *A socially optimal mediator mechanism with optimal matching uses two messages. If the types of the sender and the receiver match, the receiver observes the first message. Otherwise, the receiver observes the second message. The expected information penalty of a sender of type* ω^S *is:*

$$-\Pr(\omega^S) \textstyle\sum_{j=1}^N v_j - \textstyle\sum_{\omega' \in (\Omega \backslash \omega^S)} Pr(\omega') \left[\textstyle\sum_{j=1}^N v_j \mu(\omega^S_j = \omega^{S*}_j | \omega^S \neq \omega') \right] \qquad \text{(A.1)}$$

Proof. Assume that $\mathcal{M}$ has more than one element sent with positive probability to the receiver in the event the agent types do not match, and consider two arbitrary nonidentical such messages m_1 and m_2. We argue that by issuing a single message, denoted m_{12}, every time the mediator would have sent either m_1 and m_2 results in a mediated communication equilibrium with weakly lower information penalties. Therefore issuing only a single message in the event $\omega^R \neq \omega^S$ is weakly optimal.[17]

Since the information penalties are additive across traits, it suffices to consider the information penalty associated with trait j. Call $\Omega(\omega_j = 1)$ the set of all types with $\omega_j = 1$. Fixing the receiver type ω^R, the average over sender types of the information penalty from trait j when message m_1 is observed by the receiver is

$$\sum_{\omega^S \in \Omega(\omega_j=1)} \Pr(\omega^S) f(m_1|\omega^R, \omega^S) \left(v_j \frac{\sum\limits_{\omega' \in \Omega(\omega_j=1)} \Pr(\omega') f(m_1|\omega^R, \omega^S)}{\sum\limits_{\omega' \in \Omega} \Pr(\omega') f(m_1|\omega^R, \omega^S)} \right) \tag{A.2}$$
$$+ \sum_{\omega^S \in \Omega(\omega_j=0)} \Pr(\omega^S) f(m_1|\omega^R, \omega^S) \left(v_j \frac{\sum\limits_{\omega' \in \Omega(\omega_j=0)} \Pr(\omega') f(m_1|\omega^R, \omega^S)}{\sum\limits_{\omega' \in \Omega} \Pr(\omega') f(m_1|\omega^R, \omega^S)} \right)$$

Note the term within parentheses is the same for all types with the same realization of ω_j. Combining these terms together we find

$$v_j \frac{\left[\sum\limits_{\omega^S \in \Omega(\omega_j=1)} \Pr(\omega^S) f(m_1|\omega^R, \omega^S) \right]^2 + \left[\sum\limits_{\omega^S \in \Omega(\omega_j=0)} \Pr(\omega^S) f(m_1|\omega^R, \omega^S) \right]^2}{\sum\limits_{\omega^S \in \Omega} \Pr(\omega^S) f(m_1|\omega^R, \omega^S)} \tag{A.3}$$

Correspondingly for message m_2 we have

$$v_j \frac{\left[\sum\limits_{\omega^S \in \Omega(\omega_j=1)} \Pr(\omega^S) f(m_2|\omega^R, \omega^S) \right]^2 + \left[\sum\limits_{\omega^S \in \Omega(\omega_j=0)} \Pr(\omega^S) f(m_2|\omega^R, \omega^S) \right]^2}{\sum\limits_{\omega^S \in \Omega} \Pr(\omega^S) f(m_2|\omega^R, \omega^S)} \tag{A.4}$$

Note that the total information penalty induced by sending m_1 and m_2 is the sum of equations A.3 and A.4.[18]

[17] Any elements of $\mathcal{M}$ sent with 0 probability can be discarded from the mechanism without affecting the player's information penalties.

[18] We need not account for the relative probabilities of the messages as these are built into the terms $\Pr(\omega^i)\Pr(m_1|\omega^i)$ of equation A.2.

Finally we find for the combined message m_{12} that

$$v_j \frac{\left[\sum_{\omega^S\in\Omega(\omega_j=1)} \Pr(\omega^S) f(m_{12}|\omega^R,\omega^S)\right]^2 + \left[\sum_{\omega^S\in\Omega(\omega_j=0)} \Pr(\omega^S) f(m_{12}|\omega^R,\omega^S)\right]^2}{\sum_{\omega^S} \Pr(\omega^S) f(m_{12}|\omega^R,\omega^S)} \tag{A.5}$$

Note $f(m_{12}|\omega^R,\omega^S) = f(m_1|\omega^R,\omega^S) + f(m_2|\omega^R,\omega^S)$, which implies

$$\sum_{\omega^S\in\Omega'} \Pr(\omega^S) f(m_{12}|\omega^R,\omega^S) = \sum_{\omega^S\in\Omega'} \Pr(\omega^S) f(m_1|\omega^R,\omega^S) + \sum_{\omega^S\in\Omega'} \Pr(\omega^S) f(m_2|\omega^R,\omega^S)$$

for any $\Omega' \subseteq \Omega$. Therefore we can write equation A.5 as follows

$$v_j \frac{\begin{array}{r}\left[\sum_{\omega^S\in\Omega(\omega_j=1)} \Pr(\omega^S) f(m_1|\omega^R,\omega^S) + \sum_{\omega^S\in\Omega(\omega_j=1)} \Pr(\omega^S) f(m_2|\omega^R,\omega^S)\right]^2 + \\ \left[\sum_{\omega^S\in\Omega(\omega_j=0)} \Pr(\omega^S) f(m_1|\omega^R,\omega^S) + \sum_{\omega^S\in\Omega(\omega_j=0)} \Pr(\omega^S) f(m_2|\omega^R,\omega^S)\right]^2\end{array}}{\sum_{\omega^S\in\Omega} \Pr(\omega^S) f(m_1|\omega^R,\omega^S) + \sum_{\omega^S\in\Omega} f(\omega^S) \Pr(m_2|\omega^R,\omega^S)} \tag{A.6}$$

To find the difference in information penalties caused by combining messages m_1 and m_2, we subtract equation A.6 from the sum of equations A.3 and A.4. This difference is equal to

$$2 * v_j \frac{\left[\begin{array}{r}\sum_{\omega^S\in\Omega(\omega_j=1)} \Pr(\omega^S) f(m_1|\omega^R,\omega^S) * \sum_{\omega^S\in\Omega(\omega_j=0)} \Pr(\omega^S) f(m_2|\omega^R,\omega^S) - \\ \sum_{\omega^S\in\Omega(\omega_j=0)} \Pr(\omega^S) f(m_1|\omega^R,\omega^S) * \sum_{\omega^S\in\Omega(\omega_j=1)} \Pr(\omega^S) f(m_2|\omega^R,\omega^S)\end{array}\right]^2}{\begin{array}{r}\left[\sum_{\omega^S\in\Omega} \Pr(\omega^S) f(m_1|\omega^R,\omega^S)\right] * \left[\sum_{\omega^S\in\Omega} \Pr(\omega^S) f(m_2|\omega^R,\omega^S)\right] * \\ \left[\sum_{\omega^S\in\Omega} \Pr(\omega^S) f(m_1|\omega^R,\omega^S) + \sum_{\omega^S\in\Omega} \Pr(\omega^S) f(m_2|\omega^R,\omega^S)\right]\end{array}}$$

which is weakly positive. □

Proposition 4. *Any sender strategy that includes a complete grammar along every path of play can be supported in a perfect Bayesian equilibrium for sufficiently large M.*

Proof. Along the equilibrium path, assume that the sender-types follow the complete grammars 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 5. *$\sigma^*(\omega, h)$ and $\mu^*(h)$ is a PBE for a sufficiently large M. If this equilibrium exists, it provides a weakly higher payoff to all sender-types than any equilibrium that satisfies block inference. Furthermore, this equilibrium provides a strictly higher payoff for some type of sender than any equilibrium that satisfies block inference and is not outcome equivalent to $\sigma^*(\omega, h)$ and $\mu^*(h)$.*

Proof. First note that for large M, it must be that in any Pareto optimal equilibrium that all types of sender reveal all of their traits so that the receiver chooses to Match if and only if the sender and receiver types match. Any equilibrium that violates this claim would not reap the full benefit of the large value of M (and hence not be Pareto efficient) if at some history the receiver chooses End when the sender and receiver might share the same type.

Consider an arbitrary alternative equilibrium (σ, μ) that utilizes a complete grammar for all types. We proceed with our proof through a series of lemmas. Each step alters the strategy-belief pair from the previous step in a fashion that improves the welfare of all of the agents. These intermediate steps may not be PBEs since the beliefs of the receiver and the strategy of the sender may not be consistent, but the strategy-belief pair we construct in the final step is a PBE.

Our first step is to modify $(\sigma,\ \mu)$ to create $(\sigma',\ \mu')$, so that all traits fully revealed by senders $(\sigma,\ \mu)$ are revealed through messages at the terminal histories of $(\sigma',\ \mu')$.[19]

[19] This step requires the assumption of block inference. Without block inference, μ might not be replicated by straightforward inference from any set of messages.

Specifically, for terminal histories in which some traits have been revealed through signaling, we delay the termination of the game for one stage and append a stage in which the sender reveals all of the signaled traits in a message. Next, we force the beliefs of the receiver to follow straightforward inference, in which beliefs are based only on information revealed in the messages. Note that the addition of these messages and the use of straightforward inference does not change the terminal beliefs of the receiver and therefore does not change the sender's payoffs. Importantly, straightforward inferences might not be an equilibrium inference at this point.

Lemma 1. *(σ', μ') leads to payoffs equal to those under (σ, μ) for all sender-types.*

Proof. Block inference insures that after any history traits have either been revealed verifiably, signaled in such a way that the receiver knows the true realization, or the receiver's belief about the trait's realization has not been updated. Once we append the message verifiably revealing the realizations of the signaled traits, the receiver's straightforward inferences at the terminal histories are the same as under (σ, μ). Therefore the payoffs in the event of a failure to match are the same under (σ', μ') land (σ, μ). □

The second step is to modify (σ', μ') so that traits are revealed one at a time, rather than revealing multiple traits in the same message (and still requiring straightforward inference), leading to (σ'', μ''). Under straightforward inference, the sender is weakly better off. For the purposes of stating our result, we use the notation $g \oplus g'$ to describe a truncation of grammars g and g' where g is followed by g'.

Lemma 2. *Consider a history in which a sender-type has followed grammar g and traits i and j have not yet been revealed. Let m_{ij} denote a message that reveals traits i and j simultaneously, while m_i and m_j are messages that reveal i and j separately. Under straightforward inference by receivers, the sender-type has higher utility if she follows the grammar $g \oplus (m_i, m_j) \oplus g'$ relative to the grammar $g \oplus (m_{ij}) \oplus g'$ where g' is any grammar that completes $g \oplus (m_i, m_j)$.*

Proof. The payoff to grammars $g \oplus (m_i) \oplus (m_j) \oplus g'$ and $g \oplus (m_{ij}) \oplus g'$ differ only in the event that the sender and receiver differ in traits i or j and none of the traits revealed in g. Assume that grammar g reveals traits $\mathcal{T} \subset \{1, .., N\}$. Note that following g the expected payoff to the sender in the event that the sender and receiver differ on trait

i or j under grammar $g \oplus (m_i) \oplus (m_j) \oplus g'$ is

$$\begin{aligned} -(1-\rho_i^*)* &\left(\textstyle\sum_{t \in \mathcal{T}} v_t + v_i + \sum_{t \notin \mathcal{T} \cup \{i\}} \rho_t^* * v_t\right) \\ &- \rho_i^*(1-\rho_j^*)\left(\textstyle\sum_{t \in \mathcal{T}} v_t + v_i + v_j + \sum_{t \notin \mathcal{T} \cup \{i,j\}} \rho_t^* * v_t\right) \end{aligned} \tag{A.7}$$

Grammar $g \oplus (m_{ij}) \oplus g'$ has an expected payoff following g conditional on differing on either trait i or j is equal to

$$-(1-\rho_i^*\rho_j^*)\left(\textstyle\sum_{t \in \mathcal{T}} v_t + v_i + v_j + \sum_{t \notin \mathcal{T} \cup \{i,j\}} \rho_t^* * v_t\right) \tag{A.8}$$

Subtracting equation A.8 from A.7 yields

$$(1-\rho_i^*) * (1-\rho_j^*) * v_j > 0$$

The sender thus has a strict preference for grammar $g \oplus (m_i) \oplus (m_j) \oplus g'$. □

By choosing to break up a message that reveals multiple traits, the sender can retain the option to avoid releasing some traits revealed by the larger message if the sender and receiver fail to match on the initial traits revealed by the broken-up messages. Therefore, senders that were previously sending messages with multiple traits are strictly better off. Although the lemma is stated in terms of splitting of messages that reveal two traits, it is obvious (although algebraically intensive to show) that it holds for messages revealing more traits.

The third step reorders the messages in order of increasing information value, which yields the strategy-belief pair (σ^*, μ^*). Note that straightforward inference is now consistent with the sender strategy since all sender-types use the same grammar. Under straightforward inference, the sender prefers to reveal the traits in order of increasing v_j in order to screen out non-matching types before revealing high information value traits. (σ^*, μ^*) is again weakly preferred by all senders and strictly preferred by those that were not previously following this ordering.

Lemma 3. *Let $\mathcal{G}^*$ denote the set of grammars that reveal one trait per message. Given straightforward inference, all agents prefer the grammar $g \in \mathcal{G}^*$ that reveals traits in order of increasing v_i.*

Proof. Suppose that the receiver's beliefs satisfy straightforward inferences and 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_i reveals trait $\beta(i)$. Suppose for some $i \in \{1, .., N-1\}$ we have $v_{\beta(i)} > v_{\beta(i+1)}$ contradicting our claim for senders of type ω^S. We show that senders of type ω^S prefer the grammar $g' = (m_1, .., m_{i-1}, m_{i+1}, m_i, m_{i+1}, ., m_N)$ (again if the receiver's beliefs satisfy straightforward inferences) to g which contradicts our assumption that g is the most preferred grammar of type ω^S and establishes our claim.

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

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

$$\begin{aligned} -\left(1-\rho^*_{\beta(i)}\right) * \left(\textstyle\sum_{j=1}^{i} v_{\beta(j)} + \sum_{j=i+1}^{N} \rho^*_{\beta(j)} v_{\beta(j)}\right) \\ -\left(1-\rho^*_{\beta(i+1)}\right)\rho^*_{\beta(i)} * \left(\textstyle\sum_{j=1}^{i+1} v_{\beta(j)} + \sum_{j=i+2}^{N} \rho^*_{\beta(j)} v_{\beta(j)}\right) \end{aligned} \tag{A.9}$$

Conditional on the sender and receiver having different values of trait i or j, the sender has an expected utility under grammar g' equal to

$$\begin{aligned} -\left(1-\rho^*_{\beta(i)}\right)\rho^*_{\beta(i+1)} * \left(\textstyle\sum_{j=1}^{i+1} v_{\beta(j)} + \sum_{j=i+2}^{N} \rho^*_{\beta(j)} v_{\beta(j)}\right) \\ -\left(1-\rho^*_{\beta(i+1)}\right) * \left(\textstyle\sum_{j=1}^{i-1} v_{\beta(j)} + v_{\beta(i+1)} + \rho^*_{\beta(i)} v_{\beta(i)} + \sum_{j=i+2}^{N} \rho^*_{\beta(j)} v_{\beta(j)}\right) \end{aligned} \tag{A.10}$$

Subtracting equation A.9 from equation A.10 yields

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

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

Together lemmas 1, 2, and 3 imply that the candidate PODE yields a higher utility for all sender types than any other equilibrium. To show that the equilibrium path of play described by the candidate PODE grammar can be supported in a PBE, we appeal to proposition 4. □

Proposition 6. *The following comparative statics hold*

(1) $M^*_{Static}(\boldsymbol{\rho}, \mathbf{v}) - M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v})$ is increasing in $\boldsymbol{\rho}$ and v

(2) For an arbitrary v, any trait i, and $\varepsilon > 0$ define $\widetilde{v}$ where

$$\begin{aligned} \widetilde{v}_j &= v_j \text{ if } j \neq i, i+1 \\ \widetilde{v}_i &= v_i - \varepsilon \\ \widetilde{v}_{i+1} &= v_{i+1} + \varepsilon \end{aligned}$$

and assume $\widetilde{v}_i$ increases in i. Then we have $M^*_{Dynamic}(\boldsymbol{\rho}, \widetilde{\mathbf{v}}) \leq M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v})$ and $M^*_{Static}(\boldsymbol{\rho},\widetilde{\mathbf{v}}) = M^*_{Static}(\boldsymbol{\rho}, \mathbf{v})$.

Proof. $M^*_{Static}(\boldsymbol{\rho}, \mathbf{v})$ and $M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v})$ satisfy the following equations

$$M^*_{Static}(\boldsymbol{\rho}, \mathbf{v}) * \prod_{j=1}^N (1-\rho_j) - \sum_{j=1}^N v_j = -\sum_{j=1}^N (1-\rho_j) v_j$$

$$M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v}) * \prod_{j=1}^N (1-\rho_j) - \sum_{j=1}^N \left[\left\{ \prod_{i=1}^{j-1} (1-\rho_i) \right\} \rho_j + (1-\rho_j) \right] v_j = -\sum_{j=1}^N (1-\rho_j) v_j$$

Combining these results we find

$$M^*_{Static}(\boldsymbol{\rho}, \mathbf{v}) - M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v}) = \frac{\sum_{j=1}^N v_j \rho_j \left(1 - \left\{ \prod_{i=1}^{j-1} (1-\rho_i) \right\} \right)}{\prod_{j=1}^N (1-\rho_j)} \tag{A.11}$$

Claim 1: Equation A.11 clearly is increasing in v_j. The numerator is increasing in ρ_j and the denominator is decreasing in ρ_j, which implies that equation A.11 is increasing in ρ_j.

Claim 2: Consider the two vectors v and $\widetilde{v}$. $M^*_{Static}(\boldsymbol{\rho}, \mathbf{v}) = M^*_{Static}(\boldsymbol{\rho},\widetilde{\mathbf{v}})$ since the total information value of the traits is the same in both cases. Now fix the strategy of the sender to be the PODE strategy for vector v. Consider the sender's payoff if the same sender strategy and receiver beliefs are adopted given vector $\widetilde{v}$. Note that the PODE under $\widetilde{v}$ can only do better than this (i.e., yield a lower $M^*_{Dynamic}(\boldsymbol{\rho},\widetilde{\mathbf{v}})$). Simplifying our formula for $M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v})$ we have

$$\begin{aligned} M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v}) * \prod_{j=1}^N (1-\rho_j) &= \sum_{j=1}^N \left[\left\{ \prod_{k=1}^{j-1} (1-\rho_k) \right\} \rho_j + (1-\rho_j) \right] v_j - \sum_{j=1}^N (1-\rho_j) v_j \\ &= \sum_{j=1}^N \left\{ \prod_{k=1}^{j-1} (1-\rho_k) \right\} \rho_j v_j \end{aligned}$$

Noting that the trait probabilities are same for both $\mathbf{v}$ and $\widetilde{\mathbf{v}}$, $M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v})$ must be higher than $M^*_{Dynamic}(\boldsymbol{\rho},\widetilde{\mathbf{v}})$ if

$$\textstyle\sum_{j=1}^N \left\{\prod_{k=1}^{j-1}(1-\rho_k)\right\}\rho_j v_j \geq \sum_{j=1}^N \left\{\prod_{k=1}^{j-1}(1-\rho_k)\right\}\rho_j \widetilde{v}_j$$

Canceling out common terms yields

$$\begin{aligned}&\left\{\textstyle\prod_{k=1}^{i-1}(1-\rho_k)\right\}\rho_i v_i + \left\{\textstyle\prod_{k=1}^{i}(1-\rho_k)\right\}\rho_{i+1}v_{i+1} \geq \\ &\qquad\qquad \left\{\textstyle\prod_{k=1}^{i-1}(1-\rho_k)\right\}\rho_i \widetilde{v}_i + \left\{\textstyle\prod_{k=1}^{i}(1-\rho_k)\right\}\rho_{i+1}\widetilde{v}_{i+1}\end{aligned}$$

Simplifying this yields

$$\begin{aligned}\rho_i v_i + (1-\rho_i)\rho_{i+1}v_{i+1} &\geq \rho_i\widetilde{v}_i + (1-\rho_i)\rho_{i+1}\widetilde{v}_{i+1} \\ \rho_i(v_i - \widetilde{v}_i) &\geq (1-\rho_i)\rho_{i+1}\left(\widetilde{v}_{i+1} - v_{i+1}\right)\end{aligned}$$

Noting that $(v_i - \widetilde{v}_i) = (\widetilde{v}_{i+1} - v_{i+1}) = \varepsilon$ gives us

$$\frac{\rho_i}{1-\rho_i} \geq \rho_{i+1}$$

Since $\frac{\rho_i}{1-\rho_i} \geq 1 \geq \rho_{i+1}$ this relation must hold (note that $\rho_i \geq \frac{1}{2}$) and we are done. □

Proposition 7. *Consider a sequence of traits with associated parameters (ρ_i, v_i) and a conversation with N traits includes the first N traits in this sequence. Assume that v_i is bounded. Then $\frac{1}{N}\sum_{i=1}^N \left\{E\left[\mu_R\left(\omega_i^S = 0|h\right)\right] - (1-\rho_i)\right\} v_i = O\left(\frac{1}{N}\right)$*

Proof. In the dynamic case we have

$$E\left[\mu_R\left(\omega_i^S = 0|h\right)\right] = \textstyle\sum_{j=1}^N \left[\left\{\prod_{k=1}^{j-1}(1-\rho_k)\right\}\rho_j + (1-\rho_j)\right]$$

which implies that

$$\begin{aligned}\textstyle\sum_{i=1}^N \left\{E\left[\mu_R\left(\omega_i^S = 0|h\right)\right] - (1-\rho_i)\right\} v_i &= \textstyle\sum_{j=1}^N \left\{\prod_{k=1}^{j-1}(1-\rho_k)\right\}\rho_j v_j \\ &< \textstyle\sum_{j=1}^N \left(\frac{1}{2}\right)^{j-1} v_j\end{aligned}$$

Since v_j is bounded, there is some $\overline{v}$ such that $v_j \leq \overline{v}$. We then have

$$\textstyle\sum_{i=1}^N \left\{E\left[\mu_R\left(\omega_i^S = 0|h\right)\right] - (1-\rho_i)\right\} v_i < \sum_{j=1}^N \left(\frac{1}{2}\right)^{j-1}\overline{v} = 2\overline{v}\left(1 - \frac{1}{2^N}\right)$$

Dividing by N yields

$$\frac{1}{N}\sum_{i=1}^{N}\left\{E\left[\mu_R\left(\omega_i^S=0|h\right)\right]-(1-\rho_i)\right\}v_i<\frac{2}{N}\overline{v}\left(1-\frac{1}{2^N}\right)=\Theta\left(\frac{1}{N}\right)$$

□

B Extensions

B.1 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 $Q \leq N$ trait realizations. We show that sender optimal equilibria that satisfies block inference continue to have the same structure as described in section 6.5.[20]

Our candidate PODE is one in which traits are revealed by all sender-types at all histories in order of increasing information value and receivers make straightforward inferences at all histories. Since choosing Match is incentive compatible as soon 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 a PODE.

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

[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 $N_h - Q_h$ traits.

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 dynamic conversation game analyzed in the main text.

These two observations imply that when assessing whether (σ^*, μ_R^*) is a PODE, we can restrict attention to improvements along paths of play that terminate at last chance histories. As argued in section 6.5, 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 a PODE.

The PODE is one of several payoff equivalent equilibria of our game. Consider some history h, and let B_h denote the traits that must to be revealed before every 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 $Q = N - 1$ then in our equilibrium B_{h_0} contains the two lowest value traits. Since B_{h_0} are surely revealed in the PODE, the sender-types are indifferent between the order in which these traits are revealed. The information value of these traits are 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.2 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. However, there are often real costs 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 PODE.[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

[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

call this 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 lemma 3 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 $v_i = 1$ and $\rho_i = \frac{1}{2}$. 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 $p_i \geq 1$ describes an integer number of traits to reveal in stage i.

The packing problem involves complicated combinatorics, which means it is not amenable to closed form analysis. However, there are a few results we can show 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

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

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 have 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 traits, 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.3 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 5). This result is independent of ρ and the sender-type's trait realizations due to the trade-off mentioned in the original analysis: 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 grammars.

The notion of preferences over grammars is not a well-defined concept in that a sender-type's preferences over any grammar depends on the endogenous inferences made by the receiver. To make the notion of preferences over grammars concrete, we assume that the sender's preferences over grammars are formed as if the receiver's beliefs satisfy straightforward inference (section 6.5). In effect, the sender evaluates grammars as if the receiver only updates based on the verifiable content of the messages she sends. Another interpretation is that we are studying sender preferences over equilibria where all sender-types use the same grammar, in which case the equilibrium beliefs of the receiver must satisfy straightforward inference.

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 8. *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_i reveals trait $\beta(i)$. Suppose for some $i \in \{1, .., N-1\}$ we have $v_{\beta(i)} > v_{\beta(i+1)}$ contradicting our claim for senders of type ω^S. We show that senders of type ω^S prefer the grammar $g' = (m_1, .., m_{i-1}, m_{i+1}, m_i, m_{i+1}, ., m_N)$ 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 $\beta(i)$ is revealed before $\beta(i+1)$, whereas under g' trait $\beta(i+1)$ is revealed before trait $\beta(i)$. Note that the sender's payoff only differs between the grammars on the event where the sender and receiver have different realizations of either trait $\beta(i)$ or $\beta(i+1)$.

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

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

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

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

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

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

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

The non-linearity of π combined with the fact that different types of sender have different values of $\rho_j^*(\omega)$ drives the senders to have different most preferred 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 1. *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 2. *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

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

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 7 (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 problem for future work.

[24] Principally, this difficulty is due to the lack of a clear criteria for choosing between the equilibria. Standard signaling equilibrium refinements are of little use in our model.

B.4 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 we can assume that the messages are cheap-talk and the senders will continue to use truthful messages. Cheap-talk messages can eliminate truthful equilibria, and in particular equilibria with efficient matching, when payoffs from participation are low.

Suppose some sender-type ω^S obtains a negative payoff from participating, but 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.

We can eliminate the need for verifiable messages if we are willing to assume that L is large. However, the frictions caused by a preference for privacy are most important when M is small. Our intuition suggests that in most applications the gains from an appropriate match (M) ought to be of the same scale as the losses from an inappropriate match (L), so it feels unnatural to at the same time consider cases where M is small and L is large.

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

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

In any equilibrium that satisfies block inference, we must have that those 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

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

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.

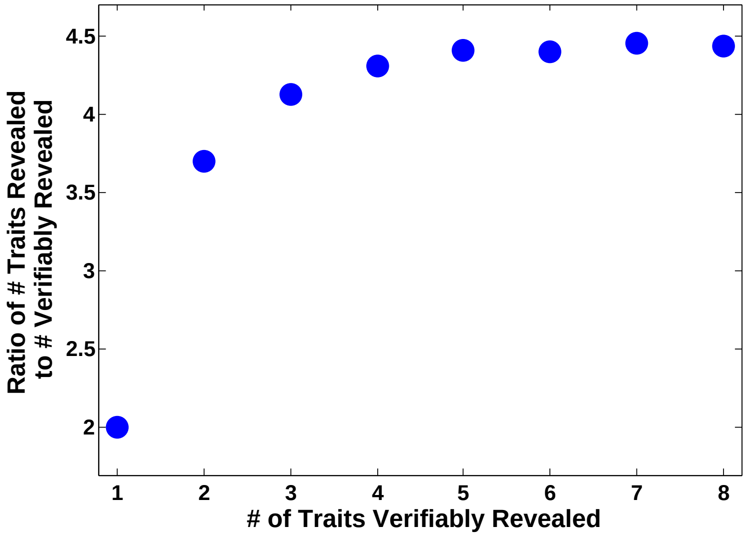

Figure 1: Numerical Results

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

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

Note that this term does not depend on the actions of any agent, so our results on the structure of the optimal static, mediated, and dynamic conversations do not change. The only results that may change are comparative statics regarding the choice to Attend (e.g., the value of M^*_{Static}). One can show that adding this term to both sides of the participation constraint yields

$$M \cdot \frac{\prod_{j=1}^{N} \rho_j^*}{1 - \prod_{j=1}^{N} \rho_j^*} \geq \sum_{j=1}^{N} (1 - \mu(\omega_j^S = \omega_j^{S*} | h_{na}))v_j$$

It is algebraically intensive, but straightforward, to show that all of our comparative statics results regarding the values of M^*_{Static}, $M^*_{Mediated}$, and $M^*_{Dynamic}$ continue to hold in this alternative model.

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.

The only points at which this renormalization could cause our results to change would be when we consider comparative statics that involve changes to the total information value, $\sum_{j=1}^{N} v_j$. In our main model, changing the total information value had no impact on M. Under our renormalization the change to the total information value would also (implicitly) involve a change to M. We now return to each such comparative static with respect to the information values and consider whether and how the results would change.

First consider proposition 6, claim 1. This claim studies the difference $M^*_{Static}(\boldsymbol{\rho}, \mathbf{v}) - M^*_{Dynamic}(\boldsymbol{\rho}, \mathbf{v})$, which means our renormalization would simply cancel and our claim remains true. Proposition 6, claim (2) is unchanged since we are considering redistributing the information value assigned to each trait without altering $\sum_{j=1}^{N} v_j$. Proposition 7 would need to be redefined to be a sequence of (M_i, ρ_i, v_i), but the claim regarding the average information penalty would remain unchanged.

Finally, consider proposition 1, claim 2, which is the only claim in the paper that may fail to hold. To understand why, note that the version of equation 4.1 that holds now is

$$M \cdot \prod_{j=1}^{N} \rho_j^* - (1 - \prod_{j=1}^{N} \rho_j^*) \sum_{j=1}^{N} v_j \geq - \sum_{j=1}^{N} \mu(\omega_j^S = \omega_j^{S*} | h_{na}) v_j$$

Rewriting this we find the following version of equation 4.2

$$M \cdot \prod_{j=1}^{N} \rho_j^* \geq \sum_{j=1}^{N} ((1 - \prod_{j=1}^{N} \rho_j^*) - \mu(\omega_j^S = \omega_j^{S*} | h_{na})) v_j$$

It is not obvious that the above equation increases with v since it could be the case that $(1 - \prod_{j=1}^{N} \rho_j^*) < \mu(\omega_j^S = \omega_j^{S*} | h_{na})$. In fact, if we return to example 1 under our new assumption we find the following results:

Example 1 Redux (Taboos under the Inference case):

Assume N=2, $\rho_1 = 0.75$, $\rho_2 = 0.5$, $v_1 = 1$ **and** $v_2 = 2$

If $M \geq 5$, then full-participation can be sustained with the off-path beliefs $\mu(\omega_1^S = 1 | h_{NA}) = 0$ and $\mu(\omega_2^S = 1 | h_{NA}) = 0.5$. The payoffs are

	Attend	Not Attend
[1 1]	$0.375 * M - 1.875$	-1
[1 0]	$0.375 * M - 1.875$	-1
[0 1]	$0.125 * M - 2.625$	-2
[0 0]	$0.125 * M - 2.625$	-2

If $M < 5$, then full-participation cannot be sustained using any off-path beliefs. However, if we let v_1=1.5, then we find the following

	Attend	Not Attend
[1 1]	$0.375 * M - 2.188$	-1
[1 0]	$0.375 * M - 2.188$	-1
[0 1]	$0.125 * M - 3.063$	-2.5
[0 0]	$0.125 * M - 3.063$	-2.5

Now equilibria with full participation exist if $M \geq 4.5$.

Our new assumption highlights the dual role of information values. First, high information values disincentivize senders from choosing Attend since they imply a high welfare loss when information is revealed. Second (and more counter-intuitively), a high information value, when combined with our freedom to choose off-path beliefs, can provide us a tool to disincentivize choosing Not Attend.

In the main text, the participation payoff decreases one-for-one with increases in v_i, whereas the payoff from Not Attend can decrease no more than one-for-one with increases in v_i. In other words, the first effect always dominates the second, which implies proposition 1, claim 2. Under our alternative assumption the participation payoff decreases strictly less than one-for-one with increases in v_i, so whether the first or second effect is stronger is ambiguous - it is easy to generate examples for either result.

Interestingly, under the No Inference case we recover proposition 1, claim 2. To the extent that this claim is compelling, it suggests that the No Inference case may be more intuitive than the sometimes very strong inferences allowed by the (more game-theoretically standard) Inference case.

Proposition 10. *In the No Inference case, $M^*_{Static}(\widetilde{\mathbf{v}}, \boldsymbol{\rho}) \geq M^*_{Static}(\mathbf{v}, \boldsymbol{\rho})$ for any $\widetilde{\mathbf{v}} \geq \mathbf{v}$*

Proof. In the No Inference, it remains true that the critical value $M^*_{Static}(\mathbf{v}, \boldsymbol{\rho})$ is determined by the rarest sender-type. That type's participation constraint can be written

$$M \cdot \textstyle\prod_{j=1}^{N}(1-\rho_j) \geq \sum_{j=1}^{N}((1-\prod_{j=1}^{N}(1-\rho_j)-(1-\rho_j))v_j$$

Simplifying this yields

$$M \cdot \textstyle\prod_{j=1}^{N}(1-\rho_j) \geq \sum_{j=1}^{N}(\rho_j-\prod_{j=1}^{N}(1-\rho_j))v_j$$

Since $\rho_j \geq \prod_{j=1}^{N}(1-\rho_j))$, this equation clearly is satisfied by fewer M as v increases. Therefore $M^*_{Static}(\mathbf{v}, \boldsymbol{\rho})$ must increase in v. □

E Mixed Strategies

Our focus is on determining when there is a tension between a preference for privacy and efficient matches. Any equilibrium with efficient matches is outcome equivalent to the use of truthful pure strategies. 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 losses from failing to match when it would be efficient must be 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 payoffs of the sender is negative, which is precisely when full-participation equilibria may fail to exist.

To illustrate these effects, we have constructed the following simple example. Although the parameter values are highly symmetric for computational ease, the general character of the equilibria is general.

Example 8. *Consider an example where there are three traits ($N = 3$). Each realization of each trait is equally likely ($\rho_j = \frac{1}{2}$), and we assume that the information value of each trait is 1. In such a setting, a sender of type ω has a symmetric partner type $\overline{\omega}$ that has the opposite realization of each trait (i.e., $\overline{\omega} = 1 - \omega$).*

We now construct an equilibrium where a sender of type ω and her symmetric partner type partially pool together by mixing over two messages, m_ω and $m_{\overline{\omega}}$, that are not used by any other type of sender. In our equilibrium a receiver of type ω chooses Match if only if he receives the message m_ω. Obviously this equilibrium will not feature full-participation (and so frictions exist), but we show that the utility of the receivers is improved by reducing the information penalty relative to a static full participation equilibrium.

Denote the probability that a sender of type ω sends message m_ω be p. For mixing to be an equilibrium, it must be that

$$\frac{M}{8} - 3p = -\frac{L}{8} - 3(1-p)$$

Solving for p we find

$$p = \frac{M+L}{48} + \frac{1}{2}$$

Obviously this places an upper bound on M and L after which this equilibria cannot exist. Let us now consider whether it is optimal for a receiver of type ω to match only when he observes the message m_ω. This requires

$$pM - (1-p)L > 0 > (1-p)M - pL$$

where the first inequality expresses that the receiver is willing to choose Match if he receives m_ω, while the second inequality implies the receiver will not choose Match if he receives $m_{\overline{\omega}}$. If $M > L$, then this condition is satisfied.

To close our discussion, let us compare the payoffs for senders in this equilibrium relative to an equilibrium with efficient matching. In the equilibrium with mixing, the expected payoff for the sender is

$$p\left(\frac{1}{8}M - 3\right) \leq 0$$

whereas the payoff in the equilibria with efficient matching is

$$\frac{1}{8}M - 3$$

The benefit to mixing is decreasing in p*, which implies that the greatest benefit to such an equilibrium arises when* M *and* L *are small. In other words, when the information penalties are large in magnitude relative to the payoffs generated by a successful match.*