

Conversations About Type: Privacy, Grammars and Taboos

Ned Augenblick and Aaron Bodoh-Creed

Abstract

We provide a model in which two strategic agents exchange verifiable information about their traits in order to determine if they have matching types, but have a preference to reveal as little information as possible if they do not match. Examples include firms exchanging information to determine if a joint venture would be profitable, a firm negotiating a potential merger with a competitor, communication in criminal enterprises, and the formation of dissident political groups in hostile regimes. Our main model focuses on multi-stage conversations about type in which agents use the dynamic nature of the conversation to screen out non-matching types over time, thereby reducing the amount of information revealed to these types. While there are many equilibria of this game, we identify a unique sender-optimal equilibrium in which all senders prefer to reveal one trait at-a-time in increasing order of importance. We show that if preferences for privacy are sufficiently strong, full-participation equilibria can fail to exist. In this case, there are equilibria in which certain traits are taboo and not discussed in any equilibrium. We provide examples of equilibria with taboos and argue that traits that are rare and sensitive are more likely to be subject to taboos. Finally, we discuss extensions to the model, such as allowing non-verifiable messages, mediated conversations, alternate privacy preferences, and unstructured conversations.

1 Introduction

In many strategic situations, people want to find partners with whom they can profitably interact but are also concerned about revealing information to parties with whom they cannot match. For example, a political dissident might want to determine if a new acquaintance shares the same political beliefs but does not want to reveal subversive views to a member of the ruling party. In economic situations agents need to determine if profitable contracts (e.g. mergers, joint ventures) exist but are concerned with the economic losses caused by revealing information to potential future trading partners or adversaries. Criminal organizations have an interest in acquiring new members or planning illegal actions without revealing information to either law enforcement agents posing as criminals or other criminals who might take advantage of the information. Finally, in social situations a person with unusual preferences might like to know if another person shares those preferences, but the person worries about revealing his type to someone with different tastes. The challenge facing the agents in these examples is to reveal enough information to determine if a match is optimal while revealing as little information as possible if the agents choose not to match.

In this paper, we present a novel theoretical model of information exchange between two agents, a sender (she) and a receiver (he), that desire to determine if they can profitably match but have a preference for privacy. Each player has private information about his or her type, a set of traits with realizations of either 0 or 1. Each trait is characterized by *rarity*, the probability 0 is realized, and an *information value* representing the utility penalty due to the loss of privacy if a trait value is revealed. In each stage, the sender issues a verifiable message about her trait realizations to the receiver, who can verifiably confirm that he shares the trait realizations in the message. After each message, the players update their beliefs about the other player's type based on the (implicit and explicit) content of the messages. At the end of the game, players receive a positive match value if the other player shares the same trait realizations and they choose to match, but each agent also receives an information penalty depending on the amount of information revealed to the other player regardless of whether a match occurs. The penalty for revealing information about a given trait is calculated by multiplying the information value of that trait with the other player's posterior belief of the likelihood of the player's true trait realization. For example, if the sender actually has a 0 realization on trait five and the receiver holds a high (low) belief that the sender has a 0 realization on trait five, then the receiver receives a relatively high (low) information penalty for that trait.

For expositional purposes, we first discuss the static game in which players exchange one message about their type. In this case, the sender must fully reveal her type in order to

match, which leads to a high information penalty as the receiver is then certain about the sender's type. As a result of this high penalty, a high match value is required for full-participation. The relatively low payoffs in unmediated, static communication provide a rationale for dynamic communication over many stages, which we refer to as a conversation.

When communication occurs over multiple stages, different sender-types reveal traits using a *grammar*, a sequence of sets of traits revealed in successive messages. Note that a grammar refers only to the traits and not particular realizations of traits, which implies that agents of different types using the same grammar generate different histories in equilibrium. If sender-types use different grammars, receivers update beliefs based on the information signaled by the choices of message as well as the verifiable information in the message. The potential for dynamic signaling results in a large set of equilibria - in fact, any sender strategy can be supported in some perfect Bayesian equilibrium.

Our first goal is to refine our set of equilibria to eliminate implausible predictions. We show that subject to a mild and realistic restriction on the forms of signaling possible in equilibrium all sender-types prefer the same grammar. To construct the sender-optimal grammar we first show that all sender-types strictly prefer to reveal one trait at a time. This is because the dynamic nature of a conversation allows the agents to stop conversing at each stage if they learn that they are not a match, which encourages sender to reveal small amounts of information in each message. The incentive to delay information revelation in order to remove non-matching types from the conversation provides a novel explanation for the empirical observation of delays in bargaining. Second, all types prefer to order the conversation from least important to most important traits, regardless of their trait realizations. This uniformity is surprising as one might imagine that a sender would prefer to discuss her own idiosyncratic, unusual trait realizations later in the conversation (as these traits are then revealed less often). Interestingly, this effect is balanced by the dynamic benefit of eliminating more non-matching types from the conversation in the initial stages by revealing a rare trait earlier in the conversation. With this uniformity of preference in mind, we refine our equilibrium set by focusing on the conversational structure that is optimal for all sender-types.

Several features of conversation follow from our refinement. First, senders reveal the minimal amount of information possible with each message, one trait at a time, which implies a pattern to the *quantity* of information revealed over time. Second, the senders reveal the lowest information value, least sensitive information first. This suggests a general (and intuitive) pattern to the *nature* of information revealed as a conversation progresses. Both aspects of our characterization of conversation follow from the dynamic incentives facing the agents and reveal the importance of these intertemporal concerns for understanding the

process of communication.

Note that this conversation structure still requires players to reveal information to non-matching types. If the match value is not particularly high, the loss of privacy might not be worth the benefit of matching for some types, and these types might choose to not enter the conversation. We define a trait as *taboo* if every type that shares a realization of that trait does not enter the conversation. Taboo traits are not discussed in equilibrium because it is (correctly) assumed that people with rare realizations of those traits do not wish to converse.

We model taboo traits in two environments, which differ in the inference that players make about agents that choose not to participate. We first discuss equilibria in which agent’s beliefs about non-participating agents are constrained to match prior beliefs about types (the No Inference Case). This model is appropriate if we believe that agents are unsophisticated in their counterfactual reasoning. We show there is a minimal match value required to induce all agents to participate in the conversation, provide an ordering over the willingness of different types to participate, and argue that taboo traits are more likely to occur when trait realizations are very rare or agents face a high-cost from revealing a trait to others.

We then discuss equilibria in which agent’s inference about non-participating players is only constrained by the standard rules of perfect Bayesian equilibrium (the Inference Case). In this case, the information penalty of non-participation is endogenous, which significantly complicates the analysis. We are able to describe the set of beliefs that support full-participation using a set of linear constraints and show that full-participation can be sustained with a lower match value than in the No Inference Case. By skewing beliefs following an observation of nonparticipation towards those agents most tempted to take that choice, we “punish” the failure to participate by these agents and provide an incentive to take part in the conversation.

We provide a simple example where no full participation equilibrium exists and there is an equilibrium with taboo traits. When the payoff from successfully matching is sufficiently high we show, as in the No Inference Case, that rare trait realizations make full participation harder to support in equilibrium. More interestingly, increasing the information value of a trait both reduces the payoff from participation, which makes participation harder to support, and makes the off-path beliefs the agents hold upon observing non-participation more punishing, which enhances participation. In contrast to our results in the No Inference Case, in the Inference Case examples can be generated where either effect of raising the penalty for revealing a trait dominates.

Finally, we show the effect of deviating from the model by (1) removing the verifiability

of messages, (2) adding a mediator, (3) changing the information penalty function, and (4) loosening the conversation structure to make it more symmetric.

In the first extension, we argue that if cheap-talk messages are employed, only the receiver has an incentive to “lie.” Interestingly, the lies the receivers are tempted to use involve non-truthful behavior over several stages, and (counter-intuitively) it is the receivers with the most common types that have the most to gain in the event of a lie. The incentive to lie can be mitigated by having the receiver burn money at each stage of the conversation (or provide payments to the sender). We characterize the required payment schemes and show that they may not be monotone in equilibrium.

In the second extension, we show that social welfare can be strictly improved by a uninterested informed mediator that acts as an “information sink.” The mediator’s optimal action is to simply tell the participants if they match or not, which significantly reduces the information penalty when agents do not match.

In the third extension, we show that the results above are robust to small changes in the utility from privacy. However, as our utility for privacy becomes more concave (convex), the current informational penalty from revealing a rare trait early overweights (underweights) the dynamic benefit of removing more agents from the conversation, and the agent prefers to discuss his common (rare) traits first. In this case, different types have different preferences over grammars. We show that with a stronger refinement in which agents are naive about future stages, there is a unique equilibrium in which players reveal traits in an order determined by the value of the trait, convexity of π , and rarity of each trait. For a concave π function and equal trait values v_j , the intuition of the refinement is that all agents with the individual trait realization most common in the population at that stage of the game prefer grammars in which that trait is discussed first. Assuming the agents are naive, these agents can credibly reveal the preferred trait first and argue that the sender cannot infer anything about their other traits.

In the final extension, we argue that loosening the form of the conversation will not lead to significantly different results, although it does lead to a more complicated analysis and many more equilibria. For example, consider a situation in which two players alternate holding the role of sender and receiver. In this case, all types of the first player to serve as sender prefer to reveal the trait with the lowest information value. Given this, all types of the second player to serve as sender prefer to reveal the trait with the second lowest information value, and so on. We leave considering more elaborate structures with three or more players or endogenous sender-receiver roles for future work, but believe the basic insights of our model continue to hold in these other settings.

We start by reviewing related literature in Section 2. In Section 3, we outline the

theoretical model, which we apply to static and dynamic conversations in Section 4 and Section 5, respectively. We discuss the emergence of taboo traits in Section 6. Finally, we discuss a variety of extensions to the model in Section 7 and conclude in Section 8.

2 Literature Review

Three papers in particular generate results akin to ours. Contemporaneously and independently of our work, Dziuda and Gradwohl [10] develop a model of information exchange with a concern for privacy that focuses on screening between productive and unproductive partner types through the exchange of information. In their model the information exchanged is both infinitely divisible and undifferentiated, which removes the signaling possibilities present in our model and forces them to focus on the volume of information revealed in equilibrium. We employ our differentiated information structure to study both the volume and kind of information exchanged over the course of conversation and to discuss the existence of endogenous taboos.

Stein [22] models conversations between competing firms, and these conversations are rendered incentive compatible over time due to the back-and-forth of information exchange between the firms. 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.

The third paper, Hörner and Skrzypacz [12], focuses on the incentives for a sender to gather information of value only to the receiver. The paper studies the dynamic information revelation scheme that maximizes the revenues obtainable by the sender, which in turn maximizes the sender's incentives to gather the costly information initially. Although the model results in rich dynamic communication, 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 close 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 occur in static settings, although some recent papers study persuasion in a dynamic setting (Honryo [11], Hörner and Skrzypacz [12], Sher [21]). Some of the recent literature also takes a mechanism design approach and attempts to derive optimal persuasion mechanisms (Glazer and Rubinstein [9], Kamenika and Gentzkow [13]). In addition there are a wealth of applied models, many of which are discussed in Milgrom [16].

Our model is distinguished from these works in a number of ways. First, the agents in our model have perfectly aligned preferences over the outcome at the end of the information exchange - the strategic tension is rooted in the agents' different preferences regarding *how* to exchange information to achieve the mutually desired goal. Second, since our agents explicitly desire to limit the revelation 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 an optimal 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 outcomes such as taboos. Our equilibrium resembles games where agents attempt to build trust over time (Ghosh and Ray [8] and Watson [23] amongst many others), although the incentive structure in our model is different.

The model also touches tangentially on the literature on cheap-talk. Although our benchmark model uses verifiable signals, we show in section 7.1 that we can achieve many of the same results in a model with cheap-talk messages and transfers. Unlike cheap-talk models, our structure is based on different preferences over information revelation 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 [14]).

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]). This literature discusses the design of algorithms that adhere to bounds on privacy, where privacy refers to the change in the probability of different outcomes as a function of changes to the inputs to the algorithm.

Unlike the differential privacy literature, we assume that each agent has preferences over the amount of knowledge that other agents possess about his or her type. In effect, the agents have preferences over the beliefs of the other agents. This aspect of our work has links to the literature on psychological games (Geanakoplos et al. [7]), although the purpose of psychological games is to model preferences over mental states (e.g. fairness, anger) rather than privacy. 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.

3 Model

There are two players, a sender and receiver, indexed by $i \in \{S, R\}$.¹ The payoff-relevant characteristics of the players are defined by N binary traits, so agent types are drawn from the set $\Omega = \{0, 1\}^N$. A generic agent type is denoted $\omega \in \Omega$ with j^{th} trait denoted $\omega_j \in \{0, 1\}$. The probability that trait j has a realized value of 1 is denoted ρ_j and the realization of each trait is stochastically independent. For example, the probability that $\omega = [1 \ 0 \ 1 \ 1]$ is realized is $\rho_1 * (1 - \rho_2) * \rho_3 * \rho_4$ and denoted $\Pr(\omega)$. 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 $\omega_j = 0$ and increase 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 probability of the trait realizations are common knowledge.

The players attempt to determine if they have the same type and should choose to match. To determine this, the players exchange messages that reveal information about their traits over multiple *stages*, indexed by $\{1, 2, \dots, N, \dots\}$. In Section 4, we analyze the game with a single stage of information transmission. In Section 5, we allow for multiple stages of information exchange and call the dynamic exchange a *conversation*.

In each stage senders reveal a message of the form $m \in \{\emptyset, 0, 1\}^N \subset \mathcal{M}$, where (for example) $m = (\emptyset, \emptyset, 1, \emptyset, 0)$ 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 $\{\omega_3 = 1, \omega_5 = 0\}$. To reduce arbitrary signaling equilibria, we require that m_t only include previously unrevealed traits.² For the majority of the paper, we assume that these messages are verifiable and cannot contain false information.

After the sender issues a message m_t in stage t , the receiver is forced to take one of three actions: *Confirm* to continue the game or end the game by choosing to *Leave* or *Match* with the other player. If the receiver chooses Confirm or Leave after stage t , the receiver must issue a message that verifiably reveals the traits the sender revealed in m_t . If the receiver chooses Match, both of the players' types are immediately revealed.³ If the players' types are identical ($\omega^S = \omega^R$), then the match is successful and both players receive a match payoff $M > 0$. If a match is initiated and the players' types are not identical ($\omega^S \neq \omega^R$), both

¹In Section 7.5, we show that the general results hold when players alternate roles.

²Without this requirement a sender who previously revealed $\omega_1 = 0$ could again reveal that $\omega_1 = 0$ to signal that $\omega_2 = 0$.

³One might imagine a similar setup in which the receivers request to initiate a match must be first approved by the sender. This setup does not change the qualitative results of this paper, but it does allow for more complicated equilibria in which the players use the decision to match to convey information. This added complexity is orthogonal to our investigation, so we focus on this simplified version.

agents receive match payoff $-L < 0$. If the game ends with a Leave decision, each agent receives a match payoff of 0. For the majority of the paper we assume that the receiver's message is verifiable.⁴

Until our discussion of endogenous taboos in section 6, we do not employ a formal participation decision or an associated participation constraint for either senders or receivers. However, receivers can de facto fail to participate by choosing Leave before learning whether or not they are a match for the sender's type. An equilibrium has *full participation* if no receiver chooses Leave prior to learning that he cannot match with the sender with probability 1, and our analysis prior to section 6 focuses on full-participation equilibria.⁵

A history is comprised of the verifiable messages exchanged by the agents. As the receiver has only one action (Confirm) in which the game continues, we describe a history by only listing the sender's trait revelations.⁶ We denote the set of histories of trait revelation up to stage n as \mathcal{H}_n with the initial null history at the first stage denoted h_0 . We denote the set of all possible histories as $\mathcal{H} = \cup_{n=0}^N \mathcal{H}_n$. A generic history of an ongoing conversation at stage t is then a sequence of the form $h = (m_1, \dots, m_t)$.

An equilibrium strategy for the sender in each stage is a function of the form:

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

An equilibrium strategy for a receiver is a function of the form:

$$\chi : \Omega \times \mathcal{H} \rightarrow \{\text{Confirm, Leave, Match}\}$$

We use perfect Bayesian equilibria (PBE) as our solution concept. In every PBE, each history h is associated with beliefs for each player about the other agent's traits. $\mu_{-i}(\omega^i = \omega|h)$ denotes the equilibrium beliefs held by the other player (denoted $-i$) at history h that player i 's type, ω^i , is equal to $\omega \in \Omega$. Let $\mu_{-i}(\omega_j^i = 1|h)$ be the equilibrium beliefs held by the other player at history h that player i 's realization of trait j , ω_j^i , is equal to 1. $\mu_{-i}(\omega_j^i = 1|h)$ can be derived on the equilibrium path from $\mu_{-i}(\omega^i = \omega|h)$ using Bayes' rule. We often use this alternative expression of the agents' beliefs for expositional ease.

⁴One might believe that the receiver plays a purely mechanical role of confirming after observing a message that demonstrates shared traits and leaving after a message demonstrates a mismatch of traits, but this is not entirely true. As we discuss below, senders can use verifiable messages about some traits to signal realizations of other traits in a nonverifiable fashion. When this occurs, the receiver forms posteriors based on equilibrium actions and chooses to confirm or leave based on these posteriors. Note that the sender does not verifiably reveal traits the sender conveys nonverifiably.

⁵In section 6 we allow both the sender and receiver to make participation decisions prior to participating in the conversation (but after learning their types).

⁶Terminal histories must reflect whether the game ended with the receiver choosing to Match or Leave.

We now define compressed notation for several of the commonly used beliefs that appear in our analysis. We define $\mu_{-i}(\omega_j^i = \omega_j^{-i*}|h)$ as the belief of player $-i$ that player i 's realization of trait j equals player $-i$'s, where we use the notation ω_j^{-i*} to emphasize that ω_j^{-i*} is the true realization known to player $-i$. This belief is used by player $-i$ to calculate the probability that player i shares the same type as $-i$. Finally, we define $\mu_{-i}(\omega_j^i = \omega_j^{i*}|h)$ as the probability that player $-i$'s beliefs place on i 's true realization of trait j , where we use the notation ω_j^{i*} to emphasize that ω_j^{i*} is the true realization known to player i . This function is used by player i to calculate player $-i$'s knowledge about player's i 's true type.

Example 1 (Belief Functions):

Let $\omega^S = [1 \ 0]$ and $\omega^R = [0 \ 1]$.

Imagine that after history h :

The receiver's belief that the sender's first trait realization equals 1 is 1

The receiver's belief that the sender's second trait realization equals 1 is .2

Then:

$$\begin{aligned} \mu_R(\omega_1^S = 1) &= 1 & \mu_R(\omega_2^S = 1) &= .2 \\ \mu_R(\omega_1^S = \omega_2^{R*}) &= 0 & \mu_R(\omega_2^S = \omega_2^{R*}) &= .2 \\ \mu_R(\omega_1^S = \omega_2^{S*}) &= 1 & \mu_R(\omega_2^S = \omega_2^{S*}) &= .8 \end{aligned}$$

Beliefs on the equilibrium path are defined according to Bayes' rule given the strategies and an on-path history h . Recall that $\Pr(\omega)$ is the ex-ante probability of a realization of type ω . Let $\Omega(h)$ denote the set of sender-types that have a path of equilibrium actions compatible with the messages in history h and $1\{\omega_j = 1\}$ be an indicator variable for the event that type ω has a realization of 1 for trait j . The conditional probability of sender trait j having realization 1 is then

$$\mu_R(\omega_j^S = 1|h) = \frac{\sum_{\omega \in \Omega(h)} 1\{\omega_j = 1\} \Pr(\omega)}{\sum_{\omega \in \Omega(h)} \Pr(\omega)}$$

The novel component of our setting is that agents have direct preferences over the information that they reveal to other players. Specifically, we assume that player i of type ω suffers an *information penalty* of the following form if the game ends at history h :

$$\text{Player } i\text{'s information penalty at } h: - \sum_{j=1}^N \pi(\mu_{-i}(\omega_j^i = \omega_j^{i*}|h)) * v_j \quad (3.1)$$

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 $v_{j_1} \geq v_{j_2}$ if $j_1 \geq j_2$. For the balance of the work we assume that $\pi(\mu_j) = \mu_j$, so that the information multiplier is equal to the other player's beliefs.⁷ Information penalties enter utility additively with match payoffs (M , $-L$, or 0).

In our setup, the information penalty of player i increases as player $-i$ places more probability on player i 's realized traits. We interpret

$$\mu_{-i}(\omega_j^i = \omega_j^{i*} | h) - \mu_{-i}(\omega_j^i = \omega_j^{i*} | h_0)$$

as the amount of *information* revealed (i.e. privacy lost) through the conversation at history h about player i 's trait j . We interpret

$$- (\mu_{-i}(\omega_j^i = \omega_j^{i*} | h) - \mu_{-i}(\omega_j^i = \omega_j^{i*} | h_0)) * v_j$$

as the utility representation of the preference of player i to avoid this privacy loss.

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 in the context of bargaining over mergers or sales between firms that requires 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.⁸

For ease of analysis, we assume that there is a high cost to a mismatch:

Assumption 1. Let $\phi = \frac{\prod_{j=1}^N \rho_j}{\prod_{j=1}^N \rho_j + \prod_{j=1}^N (1-\rho_j)}$. We assume: $L > \frac{\phi}{(1-\phi)} M$

Note that ϕ is the probability of the most common type conditional on an agent being of either the most or least common type. Therefore ϕ is the least amount of uncertainty that a receiver can have regarding the sender's type when the sender's type is not known with certainty. As a result of this assumption, agents will only initiate a match when they are certain that the other player shares the same type. Given that we assume that players must fully reveal their type once a match is initiated, this assumption eliminates equilibria in the dynamic game in which players match before all traits have been revealed, even though

⁷We explore different forms of $\pi(\mu_j)$ in section 7.3.

⁸We do not include the advantage from learning about the other agent's type in our model explicitly. Our results would be qualitatively similar in such a model. An alternate justification for our structure is that the partner can only partially internalize the advantageous information and agents that are not party to the conversation capture a large fraction of the benefits.

delaying the match to reveal more information would be Pareto improving. If we do not make assumption 1, the agents in our model would need to choose which traits to reveal along any path of play.^{9,10}

4 Static Message Exchange

In this section, we discuss the game with only one period in order to provide a lower bound for sender-payoffs and provide motivation for the dynamic game. In this static game, given the sender's message m , the receiver determines the probability that the sender shares the receiver's type, $\mu_R(\omega^S = \omega^{R^*} | m)$. If this probability is equal to 1, then the receiver strictly prefers to initiate a match. If it is not equal to 1 (i.e., the sender did not reveal all of his traits), the expected match payoff from matching is bounded above by $\phi M - (1 - \phi)L$, which is negative by Assumption 1. Therefore, the receiver will only initiate a match if all traits are believed to match with probability 1. We illustrate the equilibrium outcomes (with full-participation) below.

In Example 2 we compute the ex ante expected payoffs for one of the sender-types. This example serves as a useful prelude to our discussion of expected payoffs from dynamic conversations in section 5.

In the static game, the sender must reveal all N traits to *all* receiver-types. If the sender participates, she suffers the highest information penalty possible - either a match or the revelation of all of his traits to a non-matching receiver.

Lemma 1. *Consider any PBE where a sender of type ω^* matches with positive probability. The information penalty of the sender is*

$$\sum_{j=1}^N v_j \tag{4.1}$$

Proof. Assumption 1 implies that agents must completely reveal their type to secure a match. This implies ex ante payoffs for any equilibrium that involves participation are described by equation 4.1. □

⁹To understand the degree of added complexity in such a model, note that even if the receiver does not make inferences beyond the verifiable information revealed, choosing the traits to reveal in a way that minimizes the expected information penalty of the sender is an NP-Complete problem. Since these computational concerns are orthogonal to the questions of dynamic information exchange we wish to study, we prefer to avoid them using assumption 1.

¹⁰Alternately, we could recover the game analyzed below if we assume that not all traits need to be revealed and interpret assumption 1 as a choice to focus on equilibria where all agents reveal the same set of traits prior to matching as a social convention.

Example 2: (Static Communication)

Assume that $N=3$ and that $\rho_1=.8$, $\rho_2=.6$, and $\rho_3=.7$.

Focus on sender-type [110]. Consider static payoff:

Stage 1

Sender	Message	Receiver (Potential Types)	Receiver Inference (About Sender)	Receiver Response	Sender Payoff
110	"I am a 110"	000	110	Leave	$-v_1 -v_2 -v_3$
	"I am a 110"	001	110	Leave	$-v_1 -v_2 -v_3$
	"I am a 110"	010	110	Leave	$-v_1 -v_2 -v_3$
	"I am a 110"	011	110	Leave	$-v_1 -v_2 -v_3$
	"I am a 110"	100	110	Leave	$-v_1 -v_2 -v_3$
	"I am a 110"	101	110	Leave	$-v_1 -v_2 -v_3$
	"I am a 110"	110	110	Match	$M -v_1 -v_2 -v_3$
	"I am a 110"	111	110	Leave	$-v_1 -v_2 -v_3$
Expected payoff to sender-type [110]:				$.144M - v_1 - v_2 - v_3$	

Note that if the expected cost of revealing this amount of information exceeds the expected benefit of matching, then no messages are sent in equilibrium and no matches can be formed. Lemma 1 implies that a necessary condition for participation by all types of agents is that for all types the following holds:

$$M \prod_{j=1}^N \rho_j^* - \sum_{j=1}^N v_j \geq - \sum_{j=1}^N \rho_j^* v_j \tag{4.2}$$

where we let $\rho_j^* = \rho_j$ if $\omega_j = 1$ and $\rho_j^* = 1 - \rho_j$ if $\omega_j = 0$. It is easy to show that the receiver's payoff is maximized by this communication protocol and is weakly larger than the sender's payoff. Therefore, equation 4.2 also insures all types of receivers participate.

5 Dynamic Message Exchange: Conversations

In the previous section, we demonstrated that static information exchange leads to the maximum possible amount of sender information revelation when matching does not occur and therefore the lowest payoff for the sender. In this section, we argue that senders can do better when information is revealed slowly in a dynamic setting.

First, we define some useful theoretical objects in this dynamic model. Then we note that the full set of PBE is extremely large, which is not surprising for a multi-stage model with signalling. We discuss some of these equilibria, including a set of equilibria in which

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 . While mathematically interesting, the equilibria are unrealistic, require heavy sender coordination that evolves over several periods, and involve complicated inference by receivers. 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*. Within this (potentially more realistic) set of equilibria, we identify the unique equilibrium that is uniformly preferred by all types of senders to any other equilibrium. We argue that this sender-optimal equilibrium is a plausible refinement of the equilibrium set, and for the remainder of the paper we focus on this equilibrium.

5.1 Setup and Examples

In this section, we define a concept called a *grammar* and then provide examples of updating in the dynamic context.

As the receiver has only one action (Confirm) in which the game continues, the sender’s strategy defines the set of traits that will be verifiably revealed in each stage of the game. Specifically, given a sender strategy σ , define \tilde{m}_1 as $\sigma(h_0)$ and recursively define \tilde{m}_t as $\sigma(\tilde{m}_1, \tilde{m}_2, \dots, \tilde{m}_{t-1})$. A strategy-type’s *grammar* g is the sequence of sets of *traits* revealed in each stage in $\tilde{m}_1, \tilde{m}_2, \dots$. For example, if type $\omega = [1 \ 1 \ 0 \ 1]$ issues messages $\tilde{m}_1 = \{\omega_1 = 0, \omega_4 = 1\}$ and $\tilde{m}_2 = \{\omega_2 = 1, \omega_3 = 0\}$, 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*.

In terms of updating, the choice of a verifiable message in a conversation has two effects. First, the message conveys verifiable information about the agent’s type. The second effect is that the choice of message can signal the value of traits that are not directly revealed by the message. These effects are demonstrated in the following two examples:

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

Example 4 (Equilibrium Inference):

Assume that N is a multiple of 2

Consider the following potential equilibrium behavior in stage t :

Types where $\omega_{2^{*t-1}}^S = \omega_{2^{*t}}^S$	Stage t : Reveal trait $\omega_{2^{*t-1}}^S$
Types where $\omega_{2^{*t-1}}^S \neq \omega_{2^{*t}}^S$	Stage t : Reveal trait $\omega_{2^{*t}}^S$

Then after the first stage:

If $m_1 = \{\omega_1 = 0\}$	then $\mu_R(\omega_1^S = 1 m_1) = 0$	and $\mu_R(\omega_2^S = 1 m_1) = 1$
If $m_1 = \{\omega_1 = 1\}$	then $\mu_R(\omega_1^S = 1 m_1) = 1$	and $\mu_R(\omega_2^S = 1 m_1) = 0$
If $m_1 = \{\omega_2 = 0\}$	then $\mu_R(\omega_1^S = 1 m_1) = 0$	and $\mu_R(\omega_2^S = 1 m_1) = 0$
If $m_1 = \{\omega_2 = 1\}$	then $\mu_R(\omega_1^S = 1 m_1) = 1$	and $\mu_R(\omega_2^S = 1 m_1) = 1$

5.2 The (Large) Set of Equilibria

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

Proposition 1. *Any sender strategy that uses a complete grammar for every sender-type can be supported in a perfect Bayesian equilibrium with full participation for sufficiently large M .*

The multitude of equilibria is not surprising given that we are analyzing a multi-stage signaling environment. To construct each equilibrium, receivers choose to “Match” if they share all traits with the sender with certainty. Off-the-path receiver beliefs are assigned

such that receivers never match with a sender that deviates from the proposed grammar. Therefore, 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 signaling models, some of these equilibria involve unrealistic behavior and elaborate signaling systems. Given the richness of the equilibrium set, the crucial step for the remainder of this section is refining our equilibrium to a subset that exhibits plausible behavior and can be fruitfully characterized.

5.3 A Refinement on Beliefs: Block Inference

The goal of our refinement choice is to focus on plausible equilibria, while still allowing a rich set of potential conversational structures and resulting inferences. The multiplicity of equilibria is a common feature of dynamic information revelation models. Prior works restrict the inferences that can be made by, for example, exogenously requiring agents to reveal information in a prespecified order (e.g. Stein [22]) or by assuming agents make limited inferences from the information revealed (Dziuda and Gradwohl [10]).¹¹ Our refinement admits a richer variety of signalling phenomena than either of these prior works.

To motivate our refinement we provide an example of a set of equilibria that are mathematically interesting, but potentially unrealistic.

Example 5 (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 $\mu_R(\omega_1^S = 1|m_1) = 0$ and $\mu_R(\omega_j^S = 1|m_1) = .5$ 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 $\mu_R(\omega_1^S = 1|m_1) = 1$ and $\mu_R(\omega_j^S = 1|m_1) = .5$ 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 $m_1 = \{\omega_k = x\}$ then $\mu_R(\omega_k^S = x|m_1) = 1$ and $\mu_R(\omega_j^S = 1|m_1) = .5$ for $j \neq k$

¹¹Dziuda and Gradwohl [10] assume that information is “undifferentiated” in that agents cannot discern between infinitesimal pieces of information, which implies that inferences are based only on the *volume* rather than the *content* of the information revealed.

In this example, each initial message reveals a large amount of information about the type of the sender, but leads to relatively little 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, two, and four.

The strategies in the above example comprise the sender-optimal equilibrium for the given parameters because they lead to a small information penalty for the sender (in this case, revealing one trait realization) when facing the vast majority of receivers (in this case, $\frac{7}{8}$ of receivers). In general, the sender-optimal equilibrium for other parameters shares the same basic characteristic: senders group in a way such that the receiver potentially learns much 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. First, the optimality of these strategies is an artefact of our parameterization of the penalty function. In example 5 the sender 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 simple model of privacy with different information values on different traits, we do not penalize for correlational trait information, which is presumably still important for an agent with privacy concerns. We discuss this issue in Section 7.4 .

A second reason to view example 5 as an unlikely prediction is that the strategies involve significant coordination between senders and require complicated, multi-stage inference on the part of the receivers even under our simple parameterization. Although signaling in an important real-world phenomenon, the intricate inferences required in our simple example appear unlikely to occur in reality.

For these reasons, 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 in solely terms of the marginal probability of each of the sender’s traits. That is, if a trait realization has not been completely revealed (either through a verifiable message or signaling), the receiver must believe its realization is uncorrelated with the realization of any other trait. We now state our condition formally as follows:

Definition 1. *At any history h along the path of play of an equilibrium satisfying block inference, we can define $K(h), U(h) \subset \{1, \dots, N\}$ that denote the sets of known and unknown traits at history h . We require that $K(h) \cup U(h) = \{1, \dots, N\}$ and $K(h) \cap U(h) = \emptyset$. 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 $\mu_R(\omega_j^S = 1|h) = \rho_j$.

We note that examples 3 and 4 satisfy block inference. As in Examples 3 and 4, senders can still signal information about type through the use of type-specific grammars. Block inference requires that all signaled information be fully resolved within the same stage of the signal. The restriction of block inference still allows for the bulk of traits to be revealed by signalling instead of by verifiable messages in equilibrium, which we show formally in Appendix A.

5.4 Sender Optimal Grammars

There are many equilibria of our game that satisfy block inference. In the following sections, we will focus on one particular equilibrium in which 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. We justify this choice by showing that an equilibrium where all senders employ this grammar is preferred by *all* types of senders to any other equilibrium that satisfies block inference. We argue that this is the appropriate equilibrium for analysis for two reasons. First, it is natural in many settings to assume that the agent revealing information has bargaining power regarding which traits to reveal. Second, the sender-optimal equilibrium is sender-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 sender-optimal equilibrium. Stated informally, at any history in the sender-optimal PBE, all types of senders reveal the lowest information value trait that has not been revealed thus far. Second, we outline the proof showing that the equilibrium is sender-optimal, which is useful in understanding the intuition behind the uniform optimality (for senders) of this strategy.

To formally define the sender-optimal PBE, we need to define the behavior of each sender-type at each history, $\sigma^*(\omega, h)$, the behavior of each receiver-type at each history, $\chi^*(\omega, h)$, and the beliefs of senders and receivers at each history, $\mu_S^*(h)$ and $\mu_R^*(h)$ (which, following the rest of the paper, we state as beliefs about trait realizations rather than full type realizations). In the sender-optimal PBE, 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

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

As this strategy generates a grammar that is independent of 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 $\omega_3^S = 1$. Therefore, we set the beliefs of the receiver about the sender's trait j at any history, $\mu_R^*(\omega_j^S = 1|h)$, to match the verifiable message sent about that trait or be equal to the prior beliefs ρ_j if no message about that trait has been sent. We call this *straightforward inference*.

Definition 2. *An agent's beliefs satisfy **straightforward inference** if the agent's beliefs following any history of play are conditioned only on the verifiable information contained in the messages received.*

Straightforward inferences eliminate any signaling that could be conducted through type-dependent grammars and is, as a result, a strong refinement of the block inference concept defined in section 5.3. Note, however, if all senders use the same grammar (as in the sender-optimal equilibrium), the equilibrium beliefs satisfy straightforward inference.

Given these beliefs, the receiver's optimal action at each history, $\chi^*(\omega, h)$, is to Match if $\mu_R^*(\omega_j^S = \omega_j^{R*}|h) = 1$ for all traits, Leave if $\mu_R^*(\omega_j^S = \omega_j^{R*}|h) = 0$ for any trait, and Confirm otherwise. Finally, given these strategies, we set the sender's beliefs about the receiver's trait j at any history using straightforward inference.

The following Proposition states that this equilibrium is preferred by every sender-type to any other equilibrium with full participation and that the equilibrium exists if M is large enough. We focus on the sender-optimal equilibrium throughout the rest of the paper.

Proposition 2. *$\sigma^*(\omega, h), \chi^*(\omega, h), \mu_S^*(h), \mu_R^*(h)$ 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 $\sigma^*(\omega, h), \chi^*(\omega, h), \mu_S^*(h), \mu_R^*(h)$.¹²*

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 penalty value, v_j , and not the relative rareness of a specific 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

¹²Our theorem is a statement regarding the uniqueness of on-path actions. Multiple equilibria may exist with the same equilibrium outcomes supported by different off-path beliefs.

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 exactly one trait. Since receivers are already confident in 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 more negative information penalty) than revealing ω_1 , a trait about which the receivers have less confidence.

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 2 demonstrates that since these forces balance, the sender’s sole concern is the information value of the trait.¹³ 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 2 in order to build intuition about sender preferences over potential equilibria and establish the uniform sender-optimality of our equilibrium. Before we begin, we summarize the proof strategy. We start with an arbitrary equilibrium, which we denote (σ, μ_R) .¹⁴ We then modify the equilibrium actions and beliefs in three steps, leading to the sender-optimal equilibrium (σ^*, μ_R^*) . In each step, we show that all sender-types are weakly better off. We note that the sender strategy and receiver belief pairs constructed at each step may not be equilibria until the final step where we reach (σ^*, μ_R^*) .

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 fully revealed¹⁵ to the receiver through signaling (i.e. nonverifiably) and insist the receivers use straightforward inferences. This leads to (σ', μ'_R) , which might not constitute an equilibrium. Senders are indifferent between (σ', μ'_R) and (σ, μ_R) because both lead to

¹³While this feature is a result of our linear parameterization, we show the result holds with slightly non-linear π functions in Section 7.3 and discuss what can occur when π is significantly concave or convex.

¹⁴Optimal receiver actions and the resulting sender beliefs can be constructed by backward induction.

¹⁵Trait i is “fully revealed” following history h if $\mu_R(\omega_i = 1|h) \in \{0, 1\}$.

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'_R) 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, we create (σ'', μ''_R) . Senders prefer (σ'', μ''_R) to (σ', μ'_R) because revealing one trait at a time potentially prevents non-matching receiver 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 $(\sigma''', \mu'''_R) = (\sigma^*, \mu^*_R)$. Senders prefer (σ^*, μ^*_R) over (σ'', μ''_R) because traits with more information value are revealed later when fewer receiver-types are in the conversation. We now detail this proof through a series of lemmas.

Our first step is to modify (σ, μ_R) to create (σ', μ'_R) , so that all traits fully revealed by senders (σ, μ_R) are revealed through verifiable messages at the terminal histories of (σ', μ'_R) . To do this, we consider each terminal history and potentially append another stage of messaging. Specifically, for terminal histories in which all traits have not been verifiably revealed in messages, we delay the termination of the game (either a Leave or a Match action by the receiver) 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 only based on information verifiably revealed in the messages. Note that the addition of these messages and the use of straightforward inference does not change the optimal actions or terminal beliefs of the receiver and therefore does not change the sender's payoffs. Importantly, straightforward inferences might not be an appropriate equilibrium inference and therefore (σ', μ'_R) may not constitute an equilibrium.

Lemma 2. *(σ', μ'_R) leads to payoffs equal to those under (σ, μ_R) for all sender-types.*

The second step is to modify (σ', μ'_R) 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''_R) . Again, (σ'', μ''_R) is potentially not an equilibrium as the receiver is forced to follow straightforward inference. 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 grammar that follows g and is then followed by g' .

Lemma 3. *Consider a situation in which a sender-type uses an arbitrary grammar g along which 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 prefers to use the grammar*

$g \oplus (m_i, m_j) \oplus g'$ to the grammar $g \oplus (m_{ij}) \oplus g'$ where g' is an arbitrary grammar that completes $g \oplus (m_i, m_j)$.

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 three or more traits.

The third step reorders the messages in order of increasing information value to generate (σ''', μ_R''') . 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_R''') is again weakly preferred by the senders (and strictly preferred by those that were not previously following this ordering).

Lemma 4. *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 .*

Straightforward inference is now the equilibrium inference as all sender-types use the same grammar. Therefore, all sender-types must weakly prefer (σ^*, μ_R^*) to (σ, μ_R) (with strict preference for type ω^S if the outcomes differ for that type) as stated in Proposition 2.

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

Example 6 (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.

$$\begin{aligned} \text{For these types, } \mu_R(\omega_1^S = \omega_1^{S*}) &= 1 \text{ and } \mu_R(\omega_2^S = \omega_2^{S*}) = 1 \\ \text{Expected info penalty: } &= -.48(v_1 + v_2) - .32(v_1 + v_2) - .12(v_1 + v_2) - .08(v_1 + v_2) \\ &= -v_1 - v_2 \end{aligned}$$

Example 6 (Preferences over grammars) Continued:

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, $\mu_R(\omega_1^S = \omega_1^{S*}) = 1$ and $\mu_R(\omega_2^S = \omega_2^{S*}) = .6$

Receiver-types $[1\ 1]$ and $[1\ 0]$ learn the sender's type.

For these types, $\mu_R(\omega_1^S = \omega_1^{S*}) = 1$ and $\mu_R(\omega_2^S = \omega_2^{S*}) = 1$

Expected 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 $\mu_R(\omega_2^S = \omega_2^{S*}) = 1$

Expected info penalty: $= -.48(v_1 + v_2) - .32(.8v_1 + v_2) - .12(v_1 + v_2) - .08(.8v_1 + v_2)$
 $= -.92v_1 - v_2$

Payoff from Grammars 2 and 3 are better than Grammar 1
regardless of v_1 or v_2

Payoff from Grammar 2 is better than Grammar 3 $\Leftrightarrow v_2 > v_1$

Finally, we present an the outcome of the sender-optimal equilibrium with three traits in order to demonstrate a longer conversation. We use the format of Example 2 to make the path-of-play explicit.

Example 7: (Three Trait Conversation)









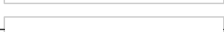

Assume that $N=3$ and that $\rho_1=.8$, $\rho_2=.6$, and $\rho_3=.7$.

Focus on sender-type $[110]$. Consider sender-optimal conversation:

Stage 1

Sender	Message	Receiver (Potential Types)	Receiver Inference (About Sender)	Receiver Response	Sender Payoff
110	"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]
	"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]

Example 7: (Three Trait Conversation) Continued:

Stage 2					
Sender	Message	Receiver (Potential Types)	Receiver Inference (About Sender)	Receiver Response	Sender Payoff
110		→ 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	$-v_1 - v_2-.3v_3$
	"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	Receiver (Potential Types)	Receiver Inference (About Sender)	Receiver Response	Sender Payoff
110		→ 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		$-v_1 - v_2-.3v_3$
	"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:				$.144M - v_1 - v_2 - v_3$	

6 A Theory of Taboos

In the previous sections, we have assumed that the match payoff is high enough to satisfy the agents' participation constraints and focused on full-participation equilibria. However, in many social and economic situations it is interesting that some people decide to not participate in a conversation at all. In this section, we examine equilibrium behavior focusing on the decision to avoid conversation.

Consequently, we modify the game by adding a participation choice at the beginning of the game:

1. Sender and receiver simultaneously choose to *Attend* or *Not Attend* the conversation.
- 2a. If both sender and receiver choose to *Attend*, the dynamic conversation game of section 5 is played.
- 2b. If either sender or receiver chooses to *Not Attend*, the decision is observed, inferences are made, and information penalties are realized.

This model makes a subtle implicit assumption that the players are paired together in the game *prior* to the conversation. For example, in the context of a joint venture, this maps to a situation in which two firms are aware of each other and the possibility of the venture *prior* to a potential conversation about the venture. In the context of social interaction, this maps to a situation in which two people know of each other *prior* to choosing to attend a party and converse.

As an alternative model, one could assume that the pairing is dependent on the participation decision. For example, the model could be altered such that senders and receivers first simultaneously decide to *Attend*, and only those agents who *Attend* are paired into conversations. Another alternative would be to assume a sequential decision: senders choose whether to *Attend* and receivers, observing the senders' decisions, choose whether to *Attend*. Both of these alternative models are qualitatively the same as the model we study in that the analytical results of the propositions and their proofs are unchanged, but the numerical examples need to be altered to reflect the different structure of the game. We focus on the model outlined above because an agent's posterior beliefs given that the other agent makes a different attendance decision can be derived more transparently.

In the participation model, we denote the history where an agent chooses *Not Attend* as h_{NA} and beliefs about an agent that chooses *Not Attend* are conditioned on this history. In the event the agent wishes to participate by choosing *Attend*, we denote the history following this choice as $h = \emptyset$ as before

We show that there is a cutoff value of M^* such that there exists an equilibrium in which all players participate if and only if $M \geq M^*$. The full participation equilibrium cannot exist when $M < M^*$ because players with rare and high information value trait realizations prefer to avoid the conversation because they have a low chance of matching and must reveal large amounts of information during the conversation. When all the players with a certain trait realization do not participate in equilibrium, we call that trait realization *taboo* as it is never discussed in conversation and players (correctly) believe that no one with that trait realization participates in conversations. Similarly, specific types of players that do not attend are denoted taboo types.

We break our discussion of taboos into two cases based on the inferences agents make when the other agent chooses *Not Attend*. One possibility compatible with our PBE solution concept is to determine beliefs using Bayes' Rule if this action is on-the-equilibrium path and leave beliefs unconstrained otherwise, which we will refer to as the *Inference Case*. The Inference Case assumes a high degree of rationality on the part of the agents and, as we see below, the equilibria in the event of taboos require sophisticated off-path beliefs in equilibrium.

The Inference Case may not be a reasonable model if the agents lack the sophistication to make inferences from non-events such as a partner's choice to Not Attend. In order to capture these situations, we define the *No Inference* case:

Definition 3. *No Inference Case:* *Beliefs about the type of a player the chooses Not Attend are restricted to the ex-ante prior probability of the realization of that type*

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

Although we motivate the No Inference case principally from bounded rationality, we believe that it accurately reflects the behavior of agents with a realistic level of sophistication. For example, if a person does not attend a party, party-goers will likely not make strong inferences about her type as there are many exogenous reasons to miss the party. To reinforce this point, we find below that many intuitive comparative statics regarding conditions under which taboos must form hold generally in the No Inference case but generally do not hold in the Inference case. To the extent these intuitions are compelling (as we believe they are), it suggests everyday reasoning falls more in line with the No Inference model. We begin by discussing the No Inference Case since it leads to more intuitive results.

6.1 The No Inference Case

As a preliminary comment, we note that it is often possible to choose parameter values where a full-participation equilibrium and an equilibrium with taboo traits coexist.¹⁶ We focus our analysis on finding conditions where full-participation equilibria do not exist, which implies that taboo equilibria are the only possible outcome.

We now state our first result on taboos, which proves that there exists a threshold match value for insuring a full-participation equilibrium exists.

¹⁶Unlike in many two-sided markets, these “no trade” equilibria with taboos cannot be broken if we insist a small number of each type of agent participate (i.e. liquidity conversants). This is because the mere act of participating requires the costly revelation of information that will dissuade entry if the probability of a match is small. However, these no-trade equilibria might be broken if we allow coalitional deviations by agents with taboo traits or types.

Proposition 3 (No Inference Case). *Given vectors ρ and v , there is some M^* such that a full participation equilibrium exists if and only if $M \geq M^*$.*

To understand the proposition, consider the payoff from participating in the full participation equilibrium for sender type ω^{S^*} given the agents follow the full participation, sender-optimal equilibrium:

$$P_N \left(M - \sum_{j=1}^N v_j \right) - \sum_{j=1}^N \left(P_{j-1} (1 - \rho_j^*) \left[\sum_{j'=1}^j v_{j'} + \sum_{j'=j+1}^N \rho_{j'}^* v_{j'} \right] \right)$$

where we let $P_J \equiv \prod_{j=1}^J \rho_j^*$, $P_0 \equiv 1$ and $\rho_j^* = \rho_j$ if $\omega_j^{S^*} = 1$ and $\rho_j^* = 1 - \rho_j$ otherwise. The payoff from not attending the conversation in the No Inference case is

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

The information penalty from attending the conversation is higher than not attending because one must reveal the true realizations of at least some traits over the course of the conversation. This cost is offset by the benefit of potentially finding a match. For every type, there is some cutoff matching value, $\overline{M}(\omega)$, that compensates for the privacy loss

$$\overline{M}(\omega) = \frac{1}{P_N} \sum_{j=1}^N \left(P_{j-1} (1 - \rho_j^*) \left[\sum_{j'=1}^j (1 - \rho_{j'}^*) v_{j'} \right] \right)$$

If $M \geq \max\{\overline{M}(\omega)\}$ is above all of these cutoffs, there exists a full participation equilibrium.

Interestingly, the cutoff values $\overline{M}(\omega)$ can be partially ordered. For example, the benefit of entering the conversation is lowest for the rarest type - $[0 \ 0 \ \dots \ 0]$ - because this type reveals the most information in the conversation and receives the lowest expected benefit from conversing (as there is a very low probability of a match). Therefore, the rarest type has the highest cutoff match value. The proposition below demonstrates that the partial ordering of the cutoff values follows the product ordering of the types.

Proposition 4 (No Inference Case). *Assuming that all agents participate, we have $\omega \geq \omega'$ implies $\overline{M}(\omega) \leq \overline{M}(\omega')$ where $\omega \geq \omega'$ only if $\omega_j \geq \omega'_j$ for all $j \in \{1, \dots, N\}$.*

Proposition 4 follows from the fact that in the No Inference case, the difference between the payoff from attending and not attending is increasing in agent type. While this difference is increasing, it is interesting to note that the payoff for participating in the conversation is not necessarily increasing in the player's type (consider Example 8 with a small M). A straightforward Corollary of our result is that the incentives of the least likely type drives the match value required to sustain full participation.

Corollary 1 (No Inference Case). $M^* = \overline{M}(0, 0, \dots, 0)$

We now provide comparative statics for the minimal match value with respect to the parameters of the model, ρ and v , respectively. Let $M^*(\rho, v)$ denote the minimal match payoff capable of sustaining full cooperation given parameters ρ, v . If $\rho_j \geq \rho'_j \geq \frac{1}{2}$ then the population has less variance in trait j under ρ_j than ρ'_j . Since the low variance makes unlikely types even rarer in the overall population, these rare types require a higher match value to be willing to bear the information penalties of engaging in conversations. Proposition 5 notes that, as ρ rises and the rarest type becomes rarer, it becomes more difficult to sustain a full participation equilibrium. If we consider the realization of each trait as a binary random variable, this suggests that we will see non-full-participation equilibria in situations where there is lower variance in the population, particularly for traits with higher information values.

Proposition 5 (No Inference Case). *Given $v \in \mathbb{R}^N$, for $\rho, \rho' \in \mathbb{R}^N$ where $\rho \geq \rho'$ we have $M^*(\rho, v) \geq M^*(\rho', v)$.*

Proposition 6 notes that the same effect as v rises. Since agents must fully reveal traits when they participate, the marginal increase in the expected information penalty for attending decreases (i.e. becomes more negative) more rapidly than the the payoff for not attending. Therefore full-participation become harder to sustain as information value rises.

Proposition 6 (No Inference Case). *Given $\rho \in \mathbb{R}^N$, for $v, v' \in \mathbb{R}^N$ where $v \geq v'$ we have $M^*(\rho, v) \geq M^*(\rho, v')$*

The participation decision is illustrated in the following example. The comparative statics with respect to variance in the population are demonstrated in the last line.

Example 8: (Participation decision under no-inference)

Assume that $N = 2$, $\rho_1 = .8$, $\rho_2 = .6$, $v_1 = 1$ and $v_2 = 2$.

Consider sender-type payoffs given full-participation equilibrium:

(Probabilities of facing receiver-types [1 1], [1 0], [0 1], and [0 0] are .48,.32,.12, and .08)

Participate with $g = [\{1\}, \{2\}]$

Sender-type [1 1]:	.48(M-1-2)	+.32(-1-2)	+.12(-1-1.2)	+.08(-1-1.2)	=-2.84+.48M
Sender-type [1 0]:	.48(-1-2)	+.32(M-1-2)	+.12(-1-0.8)	+.08(-1-0.8)	=-2.76+.32M
Sender-type [0 1]:	.48(-1-1.2)	+.32(-1-1.2)	+.12(M-1-2)	+.08(-1-2)	=-2.36+.12M
Sender-type [0 0]:	.48(-1-0.8)	+.32(-1-0.8)	+.12(-1-2)	+.08(M-1-2)	=-2.04+.08M

Example 8: (Participation decision under no-inference) Continued

Do Not Participate [Deviation]					
Sender-type [1 1]:	.48(-0.8-1.2)	+.32(-0.8-1.2)	+.12(-0.8-1.2)	+.08(-0.8-1.2)	=-2
Sender-type [1 0]:	.48(-0.8-0.8)	+.32(-0.8-0.8)	+.12(-0.8-0.8)	+.08(-0.8-0.8)	=-1.6
Sender-type [0 1]:	.48(-0.2-1.2)	+.32(-0.2-1.2)	+.12(-0.2-1.2)	+.08(-0.2-1.2)	=-1.4
Sender-type [0 0]:	.48(-0.2-0.8)	+.32(-0.2-0.84)	+.12(-0.2-0.8)	+.08(-0.2-0.8)	=-1
Requirements on M for Participation					
Sender-type [1 1]:	$M \geq$	1.75	$=\bar{M}[1 1]$		
Sender-type [1 0]:	$M \geq$	3.625	$=\bar{M}[1 0]$		
Sender-type [0 1]:	$M \geq$	8	$=\bar{M}[0 1]$		
Sender-type [0 0]:	$M \geq$	13	$=\bar{M}[0 0]$		
Then, $M^* = \bar{M}[0 0] = 13$					
If $M \geq 13$, full-participation equilibrium is possible.					
If $\rho_1 = .9$, similar calculations show that $M^* = \bar{M}[0 0] = 25.5$					

Given that there is no full participation equilibrium, what will happen? The next proposition notes that, in this case, there is always an equilibrium in which a subset of players attend the conversation.

Proposition 7 (No Inference Case). *There is always an equilibrium in which some subset of types choose Attend.*

Although the focus of our propositions is when full-participation equilibria cannot occur, the following example demonstrates that when full-participation equilibria fail to exist there is an equilibrium with taboo traits.

Example 9 (Taboos under no-inference):

Consider Example 8. If $M = 7$, sender-types [0 0] and [0 1] prefer to deviate from the full-participation equilibrium. Furthermore, there is an equilibrium in which these types do not attend the conversation. In this equilibrium, the 0 realization on trait 1 is a taboo trait.

Finally, we note that it is not necessarily the case that the utility of types that choose Attend rises when fewer types Attend in equilibrium. This is because taboo traits are (implicitly) revealed in the first period of the conversation by the choice to Attend. If these traits are associated with very high information values and are rarely revealed in the full

participation equilibrium, it is possible that more taboo types can lead to a lower payoff for conversation participants.

6.2 The Inference Case

In the Inference Case the beliefs following an agent’s (off-equilibrium path) choice of *Not Attend* are set in accord with the requirements of a perfect Bayesian equilibrium - in effect we are free to choose any such belief. First, we note that participation is easier to support in the Inference case than in the No Inference case. This result is straightforward once one realizes that the No Inference Case beliefs are just a special case of the off-path beliefs possible in a PBE for the Inference case.

Proposition 8 (Inference Case). *For a given ρ, v , the cutoff value $M^*(\rho, v)$ in the Inference Case is weakly lower than in the No Inference Case*

Proposition 8 implies that it is easier to sustain a full-participation equilibrium in the Inference Case. Intuitively, allowing off-the-path beliefs following *Not Attend* to place a higher probability on rarer types (who have a higher incentive to deviate and choose Not Attend) lowers the deviation payoff for these types as not attending then reveals more information. These “punishments” make it easier to maintain the full-participation equilibrium. This suggests that taboos are more likely to occur if players can choose to not participate without expecting strong inferences about their type.

We now provide a formal description of the set of beliefs that support a full-participation equilibrium in the Inference case. To this end, let $\Pi(\omega, M)$ denote the expected payoff from choosing Attend for a type ω given match value M if all other agents choose Attend. Note that $\Pi(\omega, M)$ is independent of the off-path beliefs of the agents. Then, given a value of M , full-participation requires that all types prefer $\Pi(\omega, M)$ to the payoff from choosing Not Attend, which depends on the off-path beliefs following a choice of Not Attend. Formally we require the following inequality hold for all ω .

$$\Pi(\omega^i, M) \geq - \sum_{j=1}^N v_j [\mu_{-i}(\omega_j^i = 1|h_{NA})1\{\omega_j^i = 1\} + (1 - \mu_{-i}(\omega_j^i = 1|h_{NA}))1\{\omega_j^i = 0\}] \quad (6.1)$$

where $1\{\omega_j^i = x\}$ is an indicator for the event $\{\omega_j^i = x\}$ and recall that $\mu_{-i}(\omega_j^i = 1|h_{NA})$ represents the other player’s beliefs about the j^{th} trait realization.

Even with inference, there is still some M such that off-the-path beliefs cannot be set in a way that satisfies each type’s constraint, as demonstrated in the following example.

Example 10 (Participation decision in Inference case):

Following Example 8:

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

Assume $v_1 = 1$ and $v_2 = 2$.

In the Inference Case, non-participation payoffs change:

Do Not Participate [Deviation]

Sender [1 1]:	$-\mu_R(\omega_1^S = 1 h_{NA})$	$-2\mu_R(\omega_2^S = 1 h_{NA})$
Sender [1 0]:	$-\mu_R(\omega_1^S = 1 h_{NA})$	$-2(1 - \mu_R(\omega_2^S = 1 h_{NA}))$
Sender [0 1]:	$-(1 - \mu_R(\omega_1^S = 1 h_{NA}))$	$-2\mu_R(\omega_2^S = 1 h_{NA})$
Sender [0 0]:	$-(1 - \mu_R(\omega_1^S = 1 h_{NA}))$	$-2(1 - \mu_R(\omega_2^S = 1 h_{NA}))$

Requirements on Beliefs for Participation

Sender [1 1]:	$= -2.84 + .48M$	$\geq -\mu_R(\omega_1^S = 1 h_{NA})$	$-2\mu_R(\omega_2^S = 1 h_{NA})$
Sender [1 0]:	$= -2.76 + .32M$	$\geq -\mu_R(\omega_1^S = 1 h_{NA})$	$-2(1 - \mu_R(\omega_2^S = 1 h_{NA}))$
Sender [0 1]:	$= -2.36 + .12M$	$\geq -(1 - \mu_R(\omega_1^S = 1 h_{NA}))$	$-2\mu_R(\omega_2^S = 1 h_{NA})$
Sender [0 0]:	$= -2.04 + .08M$	$\geq -(1 - \mu_R(\omega_1^S = 1 h_{NA}))$	$-2(1 - \mu_R(\omega_2^S = 1 h_{NA}))$

If $M = 7$, a full-participation equilibrium is possible (unlike the No Inference case). For example, if $\mu_R(\omega_1^S = 1|h_{NA}) = 0$ and $\mu_R(\omega_2^S = 1|h_{NA}) = 0.26$

If $M = 4$, a full-participation equilibrium is NOT possible.

Constraints cannot be satisfied for $\mu_R(\omega_1^S = 1|h_{NA}), \mu_R(\omega_2^S = 1|h_{NA}) \in [0, 1]$

The $M = 7$ case of Example 10 suggests the sort of equilibria that can be expected when off-path beliefs can be tuned to encourage participation. In this case the agent entertains the belief that agents who choose to Not Attend are more likely to have a low payoff from participation.

When full participation cannot be supported in equilibrium, then there must be some types who choose Not Attend despite the negative inferences that may be made regarding nonparticipants. As a preliminary observation, note that if under the parameters (v, ρ) there is an equilibrium with a single taboo trait, then there must also exist a full-participation equilibrium. Therefore, for endogenous taboos to be a necessary equilibrium outcome, there must be multiple taboo traits so that the nonparticipating agents can pool together and reduce their respective information penalties.

Proposition 9 (Inference Case). *Suppose for parameters (v, ρ) there exists an equilibrium with a single taboo trait realization. Then there also exists an equilibrium with full participation.*

The basic intuition of our proof is that in any equilibrium with a single taboo trait we can consider an arbitrary type ω' who chooses Not Attend and identify an agent with the opposite

realization for the taboo trait (and all other traits with the same realization), denoted ω , who chooses Attend. We show formally that if agents of types in all such (ω, ω') pairs participate, both types earn the same payoff as the type ω in the original equilibrium. Since the payoff as the type ω in the original equilibrium is sufficient to insure participation, there must exist an equilibrium where all agents choose Attend.

We now demonstrate endogenous taboos with an example which supports an equilibrium with two taboo traits under parameters where there is no equilibrium where all types participate. The underlying intuition is that in the presence of two taboo traits, the information penalty following the choice to Not Attend is diluted since there is uncertainty regarding which of the taboo traits (or both) each nonparticipating agent possesses.

Example 11: (Taboos under inference):

Assume that $N = 3$ and that $\rho_1 = .9$, $\rho_2 = .8$ and $\rho_3 = .5$.

Assume $v_1 = 3$, $v_2 = 4$, $v_3 = 5$ and $M = 16$.

Let $\mu_1 = \mu_R(\omega_1^S = 1 | h_{NA})$, $\mu_2 = \mu_R(\omega_2^S = 1 | h_{NA})$, and $\mu_3 = \mu_R(\omega_3^S = 1 | h_{NA})$

As in Example 9, requirements on beliefs for full participation:

Type [1 1 1]:	$-6.39 \geq$	$-3\mu_1$	$-4\mu_2$	$-5\mu_3$
Type [1 1 0]:	$-6.39 \geq$	$-3\mu_1$	$-4\mu_2$	$-5(1-\mu_3)$
Type [1 0 1]:	$-7.86 \geq$	$-3\mu_1$	$-4(1-\mu_2)$	$-5\mu_3$
Type [1 0 0]:	$-7.86 \geq$	$-3\mu_1$	$-4(1-\mu_2)$	$-5(1-\mu_3)$
Type [0 1 1]:	$-7.11 \geq$	$-3(1-\mu_1)$	$-4\mu_2$	$-5\mu_3$
Type [0 1 0]:	$-7.11 \geq$	$-3(1-\mu_1)$	$-4\mu_2$	$-5(1-\mu_3)$
Type [0 0 1]:	$-1.14 \geq$	$-3(1-\mu_1)$	$-43(1-\mu_2)$	$-5\mu_3$
Type [0 0 0]:	$-1.14 \geq$	$-3(1-\mu_1)$	$-4(1-\mu_2)$	$-5(1-\mu_3)$

No choice of beliefs can sustain full-participation equilibrium

Consider Equilibrium in which:

Types [1 0 1], [1 0 0], [0 1 1], [0 1 0], [0 0 1], [0 0 0] do not enter.

Types [1 1 1] and [1 1 0] enter and declare full type.

Therefore, trait realizations $\omega_1 = 0$ or $\omega_2 = 0$ are taboo.

Consider payoff to types from participating vs. not participating:

	Participation	Non-participation
Types [1 1 1], [1 1 0]:	-5.54	-5.57
Types [1 0 1], [1 0 0]:	-7.47	-7.29
Types [0 1 1], [0 1 0]:	-4.72	-4.71
Types [0 0 1], [0 0 0]:	-7.49	-6.43

Clearly the taboo traits are self-enforcing.

In the equilibrium of example 11, we must check the optimality conditions for both agents who Attend in equilibrium and those who choose Not Attend. Assuring that Attend is

optimal for types $[1\ 1\ 1]$ and $[1\ 1\ 0]$ given what is inferred following the decision to Not Attend is straightforward since beliefs following both of these choices are defined by Bayes rule.

The second condition we must check is that it is optimal for the agents who choose Not Attend to take that action. This condition is complicated by the larger set of possible deviations and the need to form off-path beliefs for the receiver in some events. We focus our discussion on the optimality condition for a sender with a taboo trait since senders have a larger set of deviations to consider than receivers. First we note that a potential advantage of choosing Attend for a sender with a taboo trait realization is that the deviator will have “tricked” the receiver into believing $\omega_1 = \omega_2 = 1$.

In the event that a sender with a taboo trait deviates to Attend and the receiver with which she is matched chooses Attend, the sender must choose a sequence of messages to issue to the receiver. In example 11, if the deviating sender reveals ω_3 and the receiver has a matching trait realization, the receiver will choose Match and the sender will get a payoff of $-L$ from the unproductive match. We assume that L is sufficiently large that it is never optimal to reveal ω_3 .¹⁷

If a deviating sender instead issues an off-path message, the receiver knows that a productive match is impossible, chooses Leave (verifiably confirming or disconfirming whether he shares the sender’s trait(s)), and the receiver holds beliefs that are not specified by Bayes’ rule. In the example above we choose these off-path beliefs to make it optimal for the sender to choose Not Attend instead of such a deviation. Receivers will not deviate given a large L for similar logic: deviating and participating in the conversation requires verifiable signals of traits, which either leads to an off-the-path action or a positive probability of an unprofitable match.

The comparative statics we derive for the No Inference case no longer hold in the Inference case or hold only under special conditions. We first study how changes in the variance of the trait realization affects our ability to find a full-participation equilibrium. In order to prove our result, we need to assume that M is sufficiently large that we can order the types in terms of the tightness of the respective incentive constraints of the program defined by equation 6.1. The following lemma shows that for sufficiently large M the participation payoffs are increasing in ω . Note that when ρ_j increases, there are three effects on the payoffs of an agent with $\omega_j = 1$. First, the probability of successfully matching increases, which enhances payoffs. Second, the probability of matching on trait j increases, which also increases the

¹⁷If we augmented our game to allow senders to refuse a receiver’s Match, a deviating sender with a taboo trait realization could avoid unproductive matches. However, this would constitute a deviation from the equilibrium path of play and again allow us to choose PBE off-path beliefs to make this deviation suboptimal.

probability of revealing high value traits in the event a match fails to occur later in the conversation. Third, in the event that the conversation terminates earlier, the sender suffers higher information penalties since the receiver is more confident that the sender possesses his true trait realization. For sufficiently large M , the first positive effect outweighs the remaining two negative effects. One can show that the required M need not be so high that full participation is insured¹⁸ (and hence proposition 8 is not vacuous).

Lemma 5. *Let $\Pi(\omega, \rho, v)$ denote the payoff from participation. For M sufficiently large, we have $\omega > \omega'$ implies $\Pi(\omega, \rho, v) \geq \Pi(\omega', \rho, v)$ and $\rho > \rho'$ implies $\Pi(\omega, \rho, v) \geq \Pi(\omega, \rho', v)$ and $\Pi(\omega', \rho, v) \leq \Pi(\omega', \rho', v)$*

Once we have ordered the participation payoffs, it is clear that only a subset of the lower contour of Ω will have tight participation constraints. By focusing on this subset we can show the results of Proposition 5 extend to the Inference Case. The basic intuition is that Lemma 5 tells us that decreasing ρ_j reduces the payoffs of types with $\omega_j = 1$, but these are agents whose participation constraint is known to be slack. Those agents with $\omega_j = 0$, whose participation constraints bind more tightly, are shown to have their participation constraints slackened. The net effect is to extend the scope for enforcement of full participation.

Proposition 10 (Inference Case). *Suppose M is sufficiently large that lemma 5 applies for ρ and ρ' and that a full-participation equilibrium can be sustained given $\rho = (\rho_1, \dots, \rho_N)$. Then full participation can be sustained for $\rho' \leq \rho$.*

One might hope that a similar extension of proposition 6 to the Inference Case held for comparative statics with respect to information values. Comparative statics with respect to v are difficult to generate due to two countervailing effects. First, increasing v raises the expected information penalty from participation, which makes participation harder to support. Second, raising v decreases the payoff from choosing Not Attend since the beliefs held by the other player upon observing an (off-path) choice of Not Attend impose a larger information penalty. The second “punishment” effect can make it easier to find out-of-equilibrium beliefs that support full-participation. One can demonstrate examples that satisfy the assumption of lemma 5 and do not have full-participation equilibria, but for which marginal increases in v can render a full-participation equilibrium feasible. It is an open question as to whether useful conditions can be provided to separate cases where information penalties encourage or discourage taboos in equilibrium.

¹⁸Full participation is insured if the participation payoff is positive for all types. In other words, if the expected match payoff is larger than the expected information penalty earned from participation.

7 Extensions

7.1 Cheap Talk Messages

We define a *truthful equilibrium* to be one where the sender offers messages that contain information about her type that are truthful, while the receiver issues only messages that truthfully confirm that his type matches the information conveyed by the sender. Throughout the paper, we have assumed that the messages are verifiable. Without this assumption, there will not be an equilibrium with two or more stages of conversation where the agents communicate truthfully unless the agents can offer transfers that act as a signal of truthful behavior.

To see this, suppose that there were a truthful equilibrium with two or more stages of cheap-talk messages in which information is conveyed and no transfers are employed. In any such equilibrium, there is a positive probability of a history being realized wherein the sender issues a message such that the receiver knows surely that the sender and receiver types do not match. In a truthful equilibrium, the receiver must then issue a message revealing that the types do not match, and both the sender and receiver leave the conversation and suffer information penalties.

For an example of when this action may be suboptimal for the receiver when there is no verifiability, suppose $N = 2$, $v_1 = 1, v_2 = 2$ and $\rho_1 = \rho_2 = \frac{1}{2}$. Assume that a truthful equilibrium exists. Consider receiver-type $[0 \ 1]$ and an initial sender message $\{\omega_1 = 1\}$. Given this message, the receiver knows that his type does not match the sender's. If the receiver confirms that he does not match the sender's first trait, he earns a payoff of $-1 * v_1 - \frac{1}{2} * v_2 = -2$. Alternatively, if the receiver deviates (and lies) by confirming the sender's message and then leaving after the next message (an implicit claim to not match the receiver's second trait), the receiver expects at the time of deviation to earn a payoff of $0 * v_1 - \frac{1}{2}(0 * v_2) - \frac{1}{2}(1 * v_2) = -1$.¹⁹ Since the (non-truthful) deviation is profitable for the receiver, truthfulness cannot be an equilibrium. We now state a proposition that this issue holds more generally in the context of the sender-optimal grammar

Proposition 11. *For $N \geq 2$, there is no truthful full-participation equilibrium.*

To prove this proposition, we describe a general deviation from truthfulness that is welfare improving for any type of receiver. Consider any history prior to the final stage of the conversation where the sender reveals a trait that does not match the receiver's type. Instead of leaving the conversation, the receiver (non-truthfully) confirms the sender's message. The

¹⁹The expectation includes the probability $\frac{1}{2}$ event that the sender lies in the second period by claiming $\omega_2^R = 0$ and the probability $\frac{1}{2}$ that the sender truthfully reveals $\omega_2^R = 1$ to end the game.

receiver then confirms every message from the sender that does not match his own type until a message revealing a trait that does match the receiver's type is issued. The receiver then causes the sender to believe the two agents have different realizations of the final trait revealed by taking action Leave and ending the conversation. In the event that no such matching traits are revealed by the sender before the final stage of the conversation, the receiver chooses Leave at the last stage, in effect truthfully disconfirming the final trait revealed by the sender. Given a proposed truth-telling full-participation equilibrium, the expected utility of every receiver-type is strictly improved by following such a deviation since each lie told reduces the receiver's information penalty. In the event that the receiver exits without ever truthfully revealing information about his type, then the information penalty of the receiver is strictly lower than if he had failed to deviate. In the event that the receiver must truthfully reveal his N^{th} trait in the final stage of the conversation, the receiver does receive an information penalty for revealing his N^{th} trait. However, the expected information penalty of the receiver is reduced when such a deviation is followed.

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

Proposition 12. *Suppose in stage $t \in \{1, \dots, N - 1\}$ the receiver incurs a cost, denoted c_t , where $c_t \geq v_t + (2\rho_{t+1} - 1)v_{t+1}$. If M and L are sufficiently large, a truthful equilibrium exists.*

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

Interestingly, the receiver of the most common type has the strongest incentive to deviate and lie if the sender reveals a non-matching trait. All receivers gain v_t at stage t if they

lie and convince the sender the receiver's trait matches the sender's message when it in fact does not. Ideally the receiver would cause the conversation to end by lying a second time and claiming a mismatch with the traits revealed by the sender in a later stage. Receivers with common trait realizations are more likely to reap this second benefit in a future stage, which makes lying in the current stage more tempting.

In lieu of a cost, the receiver could provide a transfer to the sender. Note that this transfer cannot be paid up front and must either be provided dynamically as the conversation progresses or as a (previously contracted) final payment at the close of a conversation where no match occurs. This provides an additional rationalization for increasing fees as the due-diligence process of mergers occurs or break-up fees in the event of failed merger negotiations.

7.2 Mediated Conversation

The focus of our paper has been direct communication between the sender and receiver. In the real-world much communication is mediated. Common examples include investment banking firms who attempt to match a client with possible acquisition targets and match-makers who attempt to pair couples with a high potential for marriage. In this section, we model these examples by assuming the existence of a third player, a neutral mediator.

We model a mediator mechanism as follows. The messages sent to the mediator mechanism are the sender's type ω^S and receiver's type ω^R . The output is a pair of messages issued to the sender and receiver, where the messages are members of a finite set of arbitrary cheap-talk messages $\mathcal{M}' = \{m'_1, m'_2, m'_3, \dots\}$ with at least two elements. Given the types of the sender and receiver, the output of the mediator mechanism is two probability distributions over the space of messages, $f_S(\mathcal{M}'|\omega^S, \omega^R)$ and $f_R(\mathcal{M}'|\omega^S, \omega^R)$. After observing the message, the receiver makes a decision to Match or Leave. We focus on equilibria with *optimal matching* in which the receiver matches if and only if he shares the sender's type. A mediator mechanism that maximizes ex-ante social welfare is surprisingly simple:

Proposition 13. *A socially optimal mediator mechanism with optimal matching uses two messages. If the types of sender and receiver match, both receive the first message. Otherwise, both receive the second message. The expected information penalty of a player of type ω^i is:*

$$-\Pr(\omega^i) \sum_{j=1}^N v_j - \sum_{\omega' \in \Omega} \Pr(\omega') \{ \sum_{j=1}^N v_j \mu_R(\omega_j = \omega_j^* | \omega^* \neq \omega') \} \quad (7.1)$$

In this optimal mechanism, the mediator simply tells the players if they match or not. The formal proof is fairly straightforward. Given any two distinct messages issued with positive probability when there is no 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 optimal matching players must learn if they have matching types. An optimal mechanism acts as an “information sink” revealing only 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. It is interesting to note that, while the mediator mechanism often Pareto dominates the dynamic equilibrium discussed above, it is possible for an individual type to be worse off with this mechanism.

In Example 12, we visually demonstrate the mediated mechanism, using the same parameters the previous visual representations of the static and dynamic equilibria:

Example 12: (Mediated Static Communication)					
Assume that $N=3$ and that $\rho_1=.8$, $\rho_2=.6$, and $\rho_3=.7$.					
Focus on sender-type [110]. Consider static payoff with mediator:					
Stage 1					
Sender	Message	Receiver (Potential Types)	Receiver Inference (About Sender)	Receiver Response	Sender Payoff
110	"No Match"	000	Not 000	Leave	$-.82v_1 - .61v_2 - .28v_3$
	"No Match"	001	Not 001	Leave	$-.85v_1 - .64v_2 - .32v_3$
	"No Match"	010	Not 010	Leave	$-.83v_1 - .59v_2 - .27v_3$
	"No Match"	011	Not 011	Leave	$-.87v_1 - .56v_2 - .33v_3$
	"No Match"	100	Not 100	Leave	$-.79v_1 - .66v_2 - .23v_3$
	"No Match"	101	Not 101	Leave	$-.74v_1 - .77v_2 - .39v_3$
	"I am a 110"	110	110	Match	$M - v_1 - v_2 - v_3$
	"No Match"	111	Not 111	Leave	$-.70v_1 - .40v_2 - .45v_3$
Expected payoff to sender-type [110]:					$.144M - .79v_1 - .63v_2 - .46v_3$
Expected payoff, dynamic communication:					$.144M - v_1 - .92v_2 - .64v_3$
Expected payoff, static communication:					$.144M - v_1 - v_2 - v_3$

7.3 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 .

As before, we define $\rho_j^*(\omega)$ as the ex-ante probability that type ω '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. When π is linear, all types prefer that traits be ordered in conversation from low information value to high information value (Proposition 2). This result is independent of ρ and the 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, players with different types may have different preferences over grammars.

The notion of preferences over grammars is not a well-defined concept in that the agent's preferences over any grammar depends on the endogenous inferences made by the other player. To make the notion of preferences concrete, we assume that when the sender's preferences are formed as if the receiver's beliefs satisfy straightforward inference (section 5.4). 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 agents use the same grammar, in which case the equilibrium beliefs of the receiver must satisfy straightforward inference. If senders all have the same most preferred grammar, we could then apply an argument along the lines of section 5.4 and proposition 2 to justify selecting this equilibrium.

Proposition 14. *Given straightforward inferences, type ω prefers to reveal trait j before trait k if and only if*

$$v_j \frac{1 - \pi(\rho_j^*(\omega))}{1 - \rho_j^*(\omega)} \leq v_k \frac{1 - \pi(\rho_k^*(\omega))}{1 - \rho_k^*(\omega)} \quad (7.2)$$

Equation 7.2 implies that an agent 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 provide sufficient conditions for 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 $\rho_j^*(\omega)$.*

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 agents are assumed to use the same grammar in equilibrium.

Example 13 (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)

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

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

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

Payoffs are equal if π is linear.

Payoff from Grammar 2 is better if π is concave.

Payoff from Grammar 3 is better if π is convex.

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 when traits have sufficiently different information values, v_j and v_k , the nonlinearity of $\pi(\rho)$ is not an issue. It is only when the information values are of a comparable level that the relative rarity of a trait realization drives disagreement between the agents.

Corollary 3. *All agents prefer to reveal trait j before trait k if both of the following hold*

$$\begin{aligned} v_j \frac{1 - \pi(\rho_j)}{1 - \rho_j} &\leq \min \left\{ v_k \frac{1 - \pi(\rho_k)}{1 - \rho_k}, v_k \frac{1 - \pi(1 - \rho_k)}{\rho_k} \right\} \\ v_j \frac{1 - \pi(1 - \rho_j)}{\rho_j} &\leq \min \left\{ v_k \frac{1 - \pi(\rho_k)}{1 - \rho_k}, v_j \frac{1 - \pi(1 - \rho_k)}{\rho_k} \right\} \end{aligned} \quad (7.3)$$

Proof. Equation 7.3 follows by requiring equation 7.2 to hold for all possible realizations of traits j and k . \square

Once types have different preferences over orderings, it becomes more difficult to select between the multiplicity of equilibria. However, it is possible to choose a unique equilibrium

ordering if players are sufficiently myopic. For example, consider the case where $v_j = v_k = 1$ and π is strictly concave. Reorder the traits so that $\rho_i \geq \rho_j$ if $i < j$. Then all agents with $\omega_1 = 1$ strictly prefer an ordering where trait 1 is revealed first (although they may disagree about the ideal order of revelation for subsequent traits). If it is commonly known that the senders are myopic, then senders with $\omega_1 = 1$ can credibly argue to the receiver as follows

I am going to reveal trait 1 that is set to 1. You should not assume anything about the traits I do not reveal since any type who can reveal that trait 1 is equal to 1 would prefer to reveal that trait prior to any other trait. Furthermore, you should assume that any agent type who could make this argument would have done so.

Given myopic agents, this message is credible. If such a message is issued at each stage, then the traits are revealed in order of decreasing ρ_i . However, if the agents have foresight regarding stages beyond the present, this message loses credibility. Specifically, while the most preferred ordering of all agents with $\omega_1 = 1$ has trait 1 revealed first, a forward-looking agent would realize that repeatedly issuing the speech suggested above in successive stages will lead to an ordering of trait 1, trait 2, trait 3, etc., which might be less preferred than some alternative equilibrium where trait 1 is not revealed first. If this occurs, then the speech above would no longer be credible and the refinement loses its appeal. The powerlessness of our refinement is a direct result of the disagreement about trait revelation order by different types of agents in this setting, and we leave this difficult problem for future work.

7.4 Correlated Traits

Our preference structure has the simplifying, but potentially unrealistic, feature that information the other party holds about correlations between traits does not constitute a loss of privacy. One might imagine that correlations are important either because they constitute a loss of privacy (and agents have a direct preference over this loss) or because the correlations would enable the other party to easily cause a further loss of privacy through an (unmodeled) action following the conversation. To see an extreme example, consider the following:

Example 14 (Correlations)

Assume that there are N traits, $\rho_i = .5$ and $v_i = 1$ for all $i \in \{1, \dots, N\}$

An agent of type $[0 \ 0 \dots 0]$ is indifferent between revealing no information and revealing that he is either of type $[0 \ 0 \dots 0]$ or of type $[1 \ 1 \dots 1]$

If one had instead based preferences over the posterior likelihood of an agent's true type (as opposed to the individual traits), the sort of correlation described in Example 14 would

yield a significant welfare loss. However, many of the signature predictions of our model, such as our analysis of endogenous taboos and the description of the amount and order of information released in equilibrium, 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 general analogs of our results could even be stated (much less proven).

There are a number of other fashions in which one might modify our model that yield correlations that render the model intractable. For example, if the trait realizations are correlated, the beliefs of the agents will naturally reflect this correlation as the conversation evolves except in non-generic cases. If agents can issue messages that are more complex, for example if they can identify themselves as a member of an arbitrary set of types, then beliefs of the players could be correlated along the equilibrium path. It has proven extremely difficult to generate sharp, useful characterizations of the equilibrium set under any of these extensions. In situations where correlations of this form are plausible, our assumptions about preferences are likely inappropriate. We leave discovering and analyzing tractable models of these environments for future work.

7.5 Two-Sided Conversations

We have focused on the case where the sender issues signals and the receiver confirms that his type matches the traits revealed in the sender’s message. Many forms of information exchange in the real-world involve the sequential revelation of information - in other words, the roles of sender and receiver can be exchanged over time. Our goal in this section is to argue that our analysis extends to the sender’s choice of what message to issue in this alternative model.

Our analysis, which is founded on the incentives of the sender, extends naturally to more elaborate frameworks where the identity of the sender can change over time. Proposition 2 applies equally to any agent in stages where that agent is required to send a message. Therefore, all types of that agent prefer to issue a message that reveals the trait with the lowest value of v_j amongst those traits that have yet to be revealed. Given the common preference amongst senders over the order of trait revelation, it is straightforward to generalize the refinement argument of section 5.4 to imply that on the equilibrium path the sender, whomever plays this role, prefers an equilibrium where all types reveal trait j at stage j .

Our argument suggests that the basic patterns of conversations we have identified will occur regardless of how the identity of the sender and receiver evolve over time. Crucially conversations will proceed with minimal amounts of information revealed in each stage, and

the information that is revealed will have the lowest information value possible. Given our equilibrium refinement, agents are indifferent between playing the role of the sender and the receiver - the information revealed about their types is effectively the same. We leave elaborations of the model that draw further differentiation between these roles or endogenize the role of sender to future work.

8 Conclusion

The focus of our paper is analyzing a model wherein agents must exchange information to discover whether they can have a productive relationship, but the agents each have a preference for privacy - in the event that a match is impossible, neither agent would like to reveal information about his or her type. 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 to study the structure of the exchange and identify both the quantity and the kind of information exchanged as the conversation progresses.

We focus on a stylized model of information exchange and 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 revealed 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 revealing traits and not on the rarity of the traits. This surprising result is driven by the dynamic nature of the conversation - early revelation of rare trait realizations is more likely to incur an information penalty, but removes the possibility of histories where the rare trait causes the halt of conversations after even more traits have been revealed.

Our model provides a natural explanation for taboos - traits that are not discussed in conversations about types. Since agents prefer not to reveal information, an agent with a rare trait realization might prefer to avoid conversation entirely (and miss out on matching) rather than revealing this trait to the other player. We show that it is harder to sustain full participation equilibria (and therefore easier for taboos to emerge endogenously) when there is lower variance in population trait realizations and traits have higher information values. Furthermore, it is easier to sustain full participation equilibria when agents can make inferences based on the choice of an agent to not converse.

In the case where agents make inferences from the choice of others to not participate, we argue that taboos cannot exist in isolation - multiple traits must be taboo to provide uncertainty regarding why the agent chose to not participate (i.e., which of the taboo traits

the agent possesses). It is straightforward to provide examples where there is no full-participation equilibrium, but equilibria with taboo traits exist. We also show that some, but not all, of our intuitive comparative statics results carry over to settings where agents make inference from non-participation.

We provide a number of extensions of our basic model. First we analyze the incentive issues that arise when the messages are not verifiable, and we show that if the receiver pays a cost for receiving a message that truthfulness can be preserved. We also discuss the possibility of altering the sender-receiver structure of our model and argue that our results would be qualitatively similar under a number of such alterations. We also 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.

While our principal goal is studying how preferences for privacy influence the structure and timing of information exchange, we hope to incorporate our model into more general models that have an information exchange component that would allow us to endogenize the match value, M . Potential settings include bargaining over merger decisions in models of market competition, bargaining over policy in political economy settings, and principal-agent problems that require information exchange between the agents.

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] Geanakoplos, J.; D. Pearce and E. Stacchetti (1989) “Psychological Games and Sequential Rationality,” *Games and Economic Behavior*, 1, pp. 60 - 79.
- [8] Ghosh, P. and D. Ray (1996) “Cooperation in Community Interaction without Information Flows,” *Review of Economic Studies*, 63, pp. 491-519.
- [9] Glazer, J. and A. Rubinstein (2003) “Optimal Rules for Persuasion,” *Econometrica*, 72 (6), pp. 1715 - 1736.
- [10] Gradwohl, R. (2012) “Privacy in Implementation,” *mimeo*.
- [11] Honryo, T. (2011) “Dynamic Persuasion,” *mimeo*.
- [12] Hörner, J. and A. Skrzypacz (2011) “Selling Information,” *mimeo*.
- [13] Kamenica, E. and M. Gentzkow (2011) “Bayesian Persuasion,” *American Economic Review*, 101, pp. 2590–2615.
- [14] 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.
- [15] Mandler, M.; P. Manzini; and M. Mariotti (2008) “A Million Answers to 20 Questions: Choosing by Checklist,” *mimeo*.
- [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.

A Appendix

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 in cases where all agents verifiably reveal a subset of traits from a set V of previously unrevealed traits. Given this restriction, the agents can signal up to roughly 50% more traits than are verifiably revealed.

Proposition 15. *Consider history h consistent with an equilibrium that satisfies block inference. Suppose there exists a set $V \subseteq U(h)$ where $|V| = k > 0$ such that all senders verifiably reveal traits from V using messages of length less than or equal to k . Then at a successor history h' we have $|K(h')| < |K(h)| + k * \log_2 3$.*

Proof. Consider an arbitrary history h of an equilibrium that satisfies block inference. At any successor history of h , which we denote h' , it must be the case that $K(h) \subset K(h')$. Let $n = |K(h')| - |K(h)|$ denote the number of traits that are revealed by the messages sent at history h . For n traits to be revealed, 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 the set of n traits revealed at history k is of size

$$\sum_{m=1}^k \binom{k}{m} 2^m$$

where the combinatorial term accounts for the different sets of m traits that can be verifiably revealed, and 2^m refers to the possible realizations of these traits. Note that 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

$$\begin{aligned} n &\leq \log_2(3^k - 1) \\ &< k \log_2 3 \end{aligned}$$

□

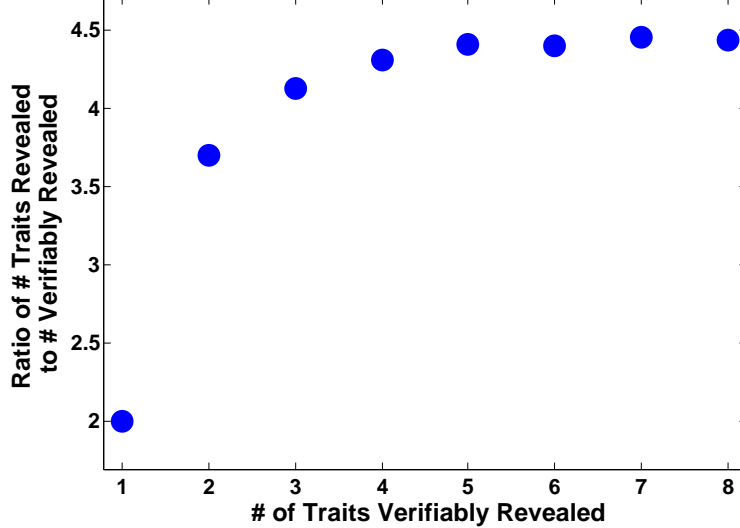


Figure 1: Numerical Results

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

$$\sum_{m=1}^k \binom{n}{m} 2^m$$

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

$$\sum_{m=1}^k \binom{n}{m} 2^m \geq 2^n$$

Although there do not exist closed forms for partial sums of this form, the figure above 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 a ratio of (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 messages.

B Proofs

Proposition 1. *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. We construct an equilibrium in which, if the sender and receiver types match, the receiver continues the conversation until the sender completely reveals his or her type (which occurs along any path of play due to the assumption that a complete grammar is used along any such path). Suppose that following any deviation to an out-of-equilibrium message the receiver believes his type does not match the sender's.²⁰ When this occurs, the receiver will not allow a match to occur.

Given that the sender and receiver match along any path on which their types are known to match by the receiver, the sender is willing to follow the grammar and completely reveal his or her type (instead of deviating and insuring a match cannot occur) if M is sufficiently high. The only exception to this argument is if the sender deviates by issuing a message that completely and verifiably reveals all his traits (at which point the receiver is forced to believe there is a match of sender and receiver types if there is a match). However this message is weakly worse than any possible grammar by the argument of Lemma 1.

Given that the sender employs a grammar that completely reveals his type along the path of play, the receiver does not wish to leave the conversation before learning that the sender and receiver types do not match. A fully optimal best response to the sender strategy (that must satisfy this property) can be constructed through backward induction. We need not consider beliefs following or sender responses following off-path actions by the receiver, since any such deviation must be a Leave or Match action that ends the game. \square

Lemma 2. *(σ', μ'_R) leads to payoffs equal to those under (σ, μ_R) 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 have not changed. Once we append the message verifiably

²⁰To fully define the off-path beliefs we need to define the beliefs of the receiver about the beliefs of the sender about the receiver's type. This is important only to the extent that it influences the optimal receiver action in the event that the receiver learns that his type does not match the sender's type.

revealing the realizations of the signaled traits, the receiver's straightforward inferences are the same as under (σ, μ_R) . Therefore the payoffs in the event of a failure to match are the same under (σ', μ'_R) and (σ, μ_R) . \square

Lemma 3. *Consider a situation in which a sender-type uses an arbitrary grammar g along which 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 prefers to use the grammar $g \oplus (m_i) \oplus (m_j) \oplus g'$ to the grammar $g \oplus (m_{ij}) \oplus g'$ where g' is an arbitrary grammar that completes $g \oplus (m_i) \oplus (m_j)$.*

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 . Let

$$\rho_k^* = \begin{cases} \rho_k & \text{if } \omega_k^S = 1 \\ 1 - \rho_k & \text{if } \omega_k^S = 0 \end{cases}$$

Assume that grammar g reveals traits $\mathcal{T} \subset \{1, \dots, N\}$. 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

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

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

$$-(1 - \rho_i^* \rho_j^*) \left(\sum_{t \in \mathcal{T}} v_t + v_i + v_j + \sum_{t \notin \mathcal{T} \cup \{i, j\}} \rho_t^* * v_t \right) \quad (\text{B.2})$$

Subtracting equation B.2 from B.1 yields

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

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

Lemma 4. *Let \mathcal{G}^* denote the set of grammars that reveal one trait per message. All agents prefer the grammar $g \in \mathcal{G}^*$ that reveals traits in order of increasing v_i .*

Proof. Suppose some type of sender ω^S has the most preferred grammar $g \in \mathcal{G}^*$ of the form $g = \{m_1, m_2, \dots, m_N\}$ where message m_i reveals trait $\beta(i)$. Suppose for some $i \in \{1, \dots, N-1\}$ 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 $g' = (m_1, \dots, m_{i-1}, m_{i+1}, m_i, m_{i+1}, \dots, m_N)$ 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

$$\begin{aligned} & - (1 - \rho_{\beta(i)}^*) * \left(\sum_{k=1}^i v_{\beta(k)} + \sum_{k=i+1}^N \rho_{\beta(k)}^* v_{\beta(k)} \right) \\ & - (1 - \rho_{\beta(i+1)}^*) \rho_{\beta(i)}^* * \left(\sum_{k=1}^{i+1} v_{\beta(k)} + \sum_{k=i+2}^N \rho_{\beta(k)}^* v_{\beta(k)} \right) \end{aligned} \quad (\text{B.3})$$

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

$$\begin{aligned} & - (1 - \rho_{\beta(i)}^*) \rho_{\beta(i+1)}^* * \left(\sum_{k=1}^{i+1} v_{\beta(k)} + \sum_{k=i+2}^N \rho_{\beta(k)}^* v_{\beta(k)} \right) \\ & - (1 - \rho_{\beta(i+1)}^*) * \left(\sum_{k=1}^{i-1} v_{\beta(k)} + v_{\beta(i+1)} + \rho_{\beta(i)}^* v_{\beta(i)} + \sum_{k=i+2}^N \rho_{\beta(k)}^* v_{\beta(k)} \right) \end{aligned} \quad (\text{B.4})$$

Subtracting equation B.3 from equation B.4 yields

$$(1 - \rho_{\beta(i)}^*) (1 - \rho_{\beta(i+1)}^*) (v_{\beta(i)} - v_{\beta(i+1)}) > 0$$

where the inequality follows from our assumption that $v_{\beta(i)} > v_{\beta(i+1)}$. But this implies that the sender strictly prefers g' to g , violating our assumption that the sender found g optimal. \square

Proposition 2. $\sigma^*(\omega, h)$, $\chi^*(\omega, h)$, $\mu_S^*(h)$, $\mu_R^*(h)$ is a PBE for a sufficiently large M . Furthermore, if this equilibrium exists, it provides a higher payoff to all sender-types and a strictly higher payoff for some type than any equilibrium with full participation by all types of senders in which $\sigma(\omega, h) \neq \sigma^*(\omega, h)$ on the equilibrium path.

Proof. Consider an arbitrary alternative equilibrium (σ, μ) and let the terminal node where a match occurs for type ω under (σ, μ) be denoted $h_T(\omega)$. If (σ_{OPT}, μ^*) is a perfect Bayesian equilibrium, lemmas 2, 3, and 4 imply (as described in the text) that (σ_{OPT}, μ^*) yields a higher utility for all sender types than any alternative grammar. To show that (σ_{OPT}, μ^*) is a perfect Bayesian equilibrium, we appeal to proposition 1. \square

Proposition 3 (No Inference Case). *Given vectors ρ and v , there is some M^* such that a full participation equilibrium exists if and only if $M \geq M^*$.*

Proof. Let M^* denote the smallest value of M such that the participation constraints are satisfied in the sender-preferred equilibrium. Since this equilibrium relaxes the symmetric participation constraints for buyers and sellers as much as possible, for $M < M^*$ there exists no full participation equilibrium. For the other direction of the proposition, consider match values $M > M^*$. Note that the participation constraints must be satisfied under M as well. \square

Proposition 4 (No Inference Case). *Assuming that all agents participate, we have $\omega \geq \omega'$ implies $\bar{M}(\omega) \leq \bar{M}(\omega')$ where $\omega \geq \omega'$ only if $\omega_j \geq \omega'_j$ for all $j \in \{1, \dots, N\}$.*

Proof. If an agent of type ω' is willing to participate, it must be the case that the following inequality holds

$$M \prod_{i=1}^N \rho_i^* - \sum_{i=1}^N \left(\prod_{j=1}^{i-1} \rho_j^* \right) (1 - \rho_i^*) \left[\sum_{j=1}^i v_j + \sum_{j=i+1}^N \rho_j^* v_j \right] \geq \sum_{j=1}^N \rho_j^* v_j \quad (\text{B.5})$$

where we use the notation $\rho_i^* = \rho_i$ if $\omega'_i = 1$ and $\rho_i^* = 1 - \rho_i$ otherwise. Rewriting equation B.5 yields

$$M \prod_{i=1}^N \rho_i^* - \sum_{i=1}^N \left(\prod_{j=1}^{i-1} \rho_j^* \right) (1 - \rho_i^*) \left(\sum_{j=1}^i (1 - \rho_j^*) v_j \right) \geq 0 \quad (\text{B.6})$$

Increasing ρ_k^* has three effects. First, it increases the expected payoff from matching. Second, it decreases the information penalty from trait k , $(1 - \rho_k^*)v_k$, in those instances where the conversation ends following the revelation of trait k . Third, it decreases the probability of the conversation ending at trait k . These three effects all increase the payoff to the agent. The fourth effect is that it increases the probability of failure following the revelation of trait k . We can write the marginal value of the fourth effect as

$$- \sum_{i=k+1}^N \left(\prod_{j=1, j \neq k}^{i-1} \rho_j^* \right) (1 - \rho_i^*) \left(\sum_{j=1}^i (1 - \rho_j^*) v_j \right)$$

Note however that we have that the marginal positive effect of increasing the match probabilities is

$$M \prod_{i=1, i \neq k}^N \rho_i^*$$

We know that

$$M \prod_{i=1, i \neq k}^N \rho_i^* \geq \sum_{i=k+1}^N \left(\prod_{j=1, j \neq k}^{i-1} \rho_j^* \right) (1 - \rho_i^*) \left(\sum_{j=1}^i (1 - \rho_j^*) v_j \right)$$

since for an agent of type ω' to be willing to participate equation B.5 must be satisfied. Therefore

$$\begin{aligned} M \prod_{i=1}^N \rho_i^* &\geq \sum_{i=1}^N \left(\prod_{j=1}^{i-1} \rho_j^* \right) (1 - \rho_i^*) \left(\sum_{j=1}^i (1 - \rho_j^*) v_j \right) \\ &\geq \sum_{i=k+1}^N \left(\prod_{j=1}^{i-1} \rho_j^* \right) (1 - \rho_i^*) \left(\sum_{j=1}^i (1 - \rho_j^*) v_j \right) \end{aligned}$$

This implies that increasing ρ_k^* makes equation B.5 slack for type ω , which implies that if ω and ω' differ only on trait k and $\omega \geq \omega'$ we have that $\bar{M}(\omega) \leq \bar{M}(\omega')$.

We can extend this argument to types ω and ω' that differ on more than one trait by applying our proof for the one-trait difference to any chain of types $(\omega_1, \dots, \omega_N)$ where $\omega_1 = \omega$, $\omega_N = \omega'$, $\omega_i \geq \omega_{i+1}$ and ω_i, ω_{i+1} differ on one trait. \square

Proposition 5 (No Inference Case). *Given $v \in \mathbb{R}^N$, for $\rho, \rho' \in \mathbb{R}^N$ where $\rho \geq \rho'$ we have $M^*(\rho, v) \geq M^*(\rho', v)$.*

Proof. From proposition 4 we know that $M^*(\rho, v)$ is defined by the cutoff threshold for sender of type $[0 \ 0 \ \dots \ 0]$. The argument of proposition 4 also proves that, conditional on an agent of type $[0 \ 0 \ \dots \ 0]$ being willing to participate, we have that the agent's participation constraint slackens as ρ_i^* , which is defined as $1 - \rho_i$, increases. Therefore we have that $\rho \geq \rho'$ implies that $M^*(\rho, v) \geq M^*(\rho', v)$. \square

Proposition 6 (No Inference Case). *Given $\rho \in \mathbb{R}^N$ and $v, v' \in \mathbb{R}^N$ where $v \geq v'$, we have $M^*(\rho, v) \geq M^*(\rho, v')$*

Proof. Consider type $\omega = [0 \ 0 \ \dots \ 0]$ that defines the lower bound on M required for full participation. Suppose for some $j \in \{1, \dots, N\}$ we have that $v_j > v'_j$ and $v_k = v'_k$ for all $k \neq j$. In this event the payoff from not participating in the match is changed by $-(1 - \rho_j)(v_j - v'_j)$ when moving from information value v' to information value v , which makes participation easier to sustain. The expected information penalty for participation is changed by no less than $-(1 - \rho_j)(v_j - v'_j)$ since every terminal node following participation suffers a change in information penalty of at least $-(1 - \rho_j)(v_j - v'_j)$ with those terminal nodes where trait j has been revealed having a strictly larger change in information penalties. Together these facts

imply the increased penalty in the event of nonparticipation is dominated by the increased information penalty in the event of participation. \square

Proposition 7 (No Inference Case). *There is always an equilibrium in which some subset of types choose Attend.*

Proof. Consider an equilibrium where only senders and receivers of type ω participate. Assume that the participants follow the sender-optimal equilibrium strategy. If any off-path messages are issued by the sender, the receiver uses his prior belief regarding any traits that have not been revealed verifiably. Agents of type ω are clearly willing to Attend since conversation partners are always good matches in equilibrium. Other types strictly prefer to Not Attend since there are no good match partners available amongst those agents willing to Attend and choosing Attend causes a weakly higher information penalty to be realized than choosing Not Attend. \square

Proposition 8 (Inference Case). *For a given ρ, v , the cutoff value $M^*(\rho, v)$ in the Inference Case is weakly lower than $M^*(\rho, v)$ in the No Inference Case*

Proof. In a full participation equilibrium of the Inference Case, we have the freedom to choose off-path beliefs. Note that one such belief is simply the prior probabilities as in the No Inference Case. Since we can only do better than to employ this off-path belief, it must be that our cutoff $M^*(\rho, v)$ is weakly lower in the Inference Case. \square

Proposition 9 (Inference Case). *Suppose for parameters (v, ρ) there exists an equilibrium with a single taboo trait realization. Then there also exists an equilibrium with full participation.*

Proof. Consider an arbitrary equilibrium with a single taboo trait. Assume without loss of generality that all agents with $\omega_j = 0$ choose Not Attend while all agents of type $\omega_j = 1$ choose Attend. Now consider two types of agents, denoted ω and ω' , where $\omega_j > \omega'_j$ and $\omega_i = \omega'_i$ for all $i \neq j$. Consider an equilibrium where all agents participate, reveal trait j in the first message, and then proceed to send messages as per the sender optimal equilibrium. Note that types ω and ω' earn the same payoff as an agent of type ω in the original equilibrium, which was sufficient to insure type ω chose Attend in that equilibrium.²¹ Therefore types ω and ω' are willing to Attend in the new equilibrium. As per the arguments of section 5.4, the senders would improve their payoff under the sender-optimal grammar and hence still be willing to Attend. Since the receiver's earn the same utility as the senders under

²¹The choice of Not Attend is weakly worse for ω in the new equilibrium since the off-path beliefs regarding trait j can only yield worse information penalties for type ω . This makes the choice of Attend easier to support in equilibrium.

the sender-optimal equilibrium, receivers are also willing Attend. Hence our constructed equilibrium with full participation must exist. \square

Lemma 5. *Let $\Pi(\omega, \rho, v)$ denote the payoff from participation. For M sufficiently large, we have $\omega > \omega'$ implies $\Pi(\omega, \rho, v) \geq \Pi(\omega', \rho, v)$ and that $\rho > \rho'$ implies $\Pi(\omega, \rho, v) \geq \Pi(\omega, \rho', v)$ and $\Pi(\omega', \rho, v) \leq \Pi(\omega', \rho', v)$*

Proof. Consider ω, ω' where there is $j \in \{1, \dots, N\}$ such that $\omega_j > \omega'_j$ and for all $i \neq j$ we have $\omega_i = \omega'_i$. Note then that

$$\Pi(\omega, \rho, v) = (2\rho_j - 1) * \frac{\partial \Pi(\omega, \rho, v)}{\partial \rho_j} + \Pi(\omega', \rho, v)$$

To prove our claim, it suffices to show that $\frac{\partial \Pi(\omega, \rho, v)}{\partial \rho_j} \geq 0$ for all such pairs ω, ω' . To see this note

$$\begin{aligned} \frac{\partial \Pi(\omega, \rho, v)}{\partial \rho_j} = & M * \prod_{i \neq j} \rho_i^* - v_j \sum_{i=1}^{j-1} \left(\prod_{k=1}^{i-1} \rho_k^* \right) (1 - \rho_i^*) + \left(\prod_{i=1}^{j-1} \rho_i^* \right) \left(\sum_{k=1}^j v_k + \sum_{k=j+1}^N \rho_k^* v_k \right) - \\ & \sum_{i=j+1}^N \left(\prod_{k < i, k \neq j} \rho_k^* \right) \left(\sum_{k=1}^i v_k + \sum_{k=i+1}^N \rho_k^* v_k \right) \end{aligned}$$

For M sufficiently large this term will be positive, and since the number of ω, ω' pairs is finite we can chose an M as required by our lemma. To prove our final point it suffices to note that $\frac{\partial \Pi(\omega, \rho, v)}{\partial \rho_j} \geq 0$ implies $\Pi(\omega, \rho, v) \geq \Pi(\omega, \rho', v)$ and that $\frac{\partial \Pi(\omega', \rho, v)}{\partial \rho_j} = -\frac{\partial \Pi(\omega, \rho, v)}{\partial \rho_j}$ so $\Pi(\omega', \rho, v) \leq \Pi(\omega', \rho', v)$. \square

Proposition 10 (Inference Case). *Suppose M is sufficiently large that lemma 5 applies for ρ and ρ' and that a full-participation equilibrium can be sustained given $\rho = (\rho_1, \dots, \rho_N)$. Then full participation can be sustained for $\rho' \leq \rho$.*

Proof. From lemma 5 we know that $\omega > \omega'$ implies $\Pi(\omega, \rho, v) \geq \Pi(\omega', \rho, v)$. Let $p_i = \mu(\omega_i^S = 1 | h_{NA})$ denote the off-path beliefs of the receiver in the event a sender does not participate for the full participation equilibrium under ρ . Then we have for all ω

$$\Pi(\omega, \rho, v) \geq - \sum_{i=1}^N (p_i \omega_i + (1 - p_i)(1 - \omega_i)) v_i$$

Consider ω, ω' where there is $j \in \{1, \dots, N\}$ such that $\omega_j > \omega'_j$ and for all $i \neq j$ we have $\omega_i = \omega'_i$. Suppose $p_i > \frac{1}{2}$ for some p that supports an equilibrium. Since $\omega > \omega'$ we then

have that

$$\Pi(\omega', \rho, v) \geq -(1 - p_j)v_j - \sum_{i \neq j} (p_i \omega_i + (1 - p_i)(1 - \omega_i)) v_i$$

implies

$$\Pi(\omega, \rho, v) \geq \Pi(\omega', \rho, v) \geq -\frac{1}{2}v_j - \sum_{i \neq j} (p_i \omega_i + (1 - p_i)(1 - \omega_i)) v_i$$

Therefore it is without loss of generality to assume $p_j \leq \frac{1}{2}$. If $p_j \leq \frac{1}{2}$ we have that

$$\begin{aligned} \Pi(\omega, \rho, v) &\geq \Pi(\omega', \rho, v) \geq -(1 - p_j)v_j - \sum_{i \neq j} (p_i \omega_i + (1 - p_i)(1 - \omega_i)) v_i \\ &\geq -p_j v_j - \sum_{i \neq j} (p_i \omega_i + (1 - p_i)(1 - \omega_i)) v_i \end{aligned}$$

so the participation constraint for type ω is slack (and strictly slack if $p_j < \frac{1}{2}$). Since lemma 5 yields $\Pi(\omega', \rho, v) \leq \Pi(\omega', \rho', v) \leq \Pi(\omega, \rho', v)$, we have that under ρ' the participation constraints are slack relative to ρ and a weakly larger set of p satisfy the participation constraints under ρ' . \square

Proposition 11. *For $N \geq 2$, there is no truthful full-participation equilibrium using the sender-optimal grammar*

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

$$-\sum_{i=1}^j v_j - \sum_{i=j+1}^N \rho_j^* v_j \tag{B.7}$$

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

$$-\sum_{i=1}^{j-1} v_j - \sum_{i=k+1}^N \rho_j^* v_j$$

which is an improvement over equation B.7.

The remaining event is where the receiver deviated from truthfulness and his deviation

requires him to either choose Match and suffer loss $-L$ and reveal all of his traits or reveal (truthfully) that trait N of the sender and receiver do not match by choosing Leave. Obviously choosing Leave is suboptimal. When the receiver reveals that he does not have the same realization of trait N as the sender by choosing Leave, the receiver gets a payoff of

$$-\sum_{i=1}^{j-1} v_j - v_N$$

The expected payoff from this event at the time of deviation (i.e., the stage at which the first lie occurs) is

$$-\sum_{i=1}^j v_j - \rho_N^* v_N$$

which is lower than equation B.7. Therefore our deviation is utility improving relative to truthfulness. \square

Proposition 12. *Suppose in stage $t \in \{1, \dots, N - 1\}$ the receiver incurs a cost, denoted c_t , where $c_t \geq v_t + (2\rho_{t+1} - 1)v_{t+1}$. If M is sufficiently large, a truthful equilibrium exists.*

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

Once the sender conveys a message that reveals to the receiver that Match is not optimal, the receiver has an incentive to lie to distort the sender's beliefs about the receiver's type (and hence reduce the receiver's information penalty). Notice that at stage N that truthfulness is incentive compatible - if the sender and receiver types are revealed to not match at this stage, it is optimal for the receiver to reveal the failure to match rather than face a penalty of $-L > 0$ for matching with a sender of another type as well as suffering the information penalties.

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

$N - 1$) and a lie (claiming a match) have respective payoffs

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

To make truth-telling optimal, we must have

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

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

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

is required.

Turning to stage $N - 2$, assume the receiver lies in stage $N - 2$. We now discuss the dynamic concerns of the receiver in stage $N - 1$, at which point the receiver may have to choose between telling the truth at stage $N - 1$ and ending the conversation or lying a second time and allowing the conversation to proceed to stage N . Our inequality on c_N implies that the latter form of lie is suboptimal, so (if the receiver lies) he will choose to end the conversation in stage $N - 1$ even if this necessitates truthfully revealing trait $N - 1$.²² Comparing the payoff from truthfully ending the conversation at stage $N - 2$ to lying and ending the conversation at stage $N - 1$ we find

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

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

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

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

Proposition 13. *A socially optimal mediator mechanism with optimal matching uses two messages. If the types of sender and receiver match, both receive the first message. Otherwise, both receive the second message. The expected information penalty of a player of type*

²²Note that this also implies that if the receiver can end the conversation with a lie at stage $N - 1$, he will clearly choose to do so.

ω^i is:

$$- \Pr(\omega^i) \sum_{j=1}^N v_j - \sum_{\omega' \in \Omega} \Pr(\omega') \left\{ \sum_{j=1}^N v_j \mu_R(\omega_j = \omega_j^* | \omega^* \neq \omega') \right\}$$

Proof. For exposition we consider the information penalty generated by the beliefs of the receiver about the sender's type, but symmetric arguments will apply for the information penalty of the receiver.

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.²³

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

$$\begin{aligned} & \sum_{\omega^S \in \Omega(\omega_j=1)} \Pr(\omega^S) f_R(m_1 | \omega^R, \omega^S) \left(v_j \frac{\sum_{\omega' \in \Omega(\omega_j=1)} \Pr(\omega') f_R(m_1 | \omega^R, \omega^S)}{\sum_{\omega' \in \Omega} \Pr(\omega') f_R(m_1 | \omega^R, \omega^S)} \right) \\ & + \sum_{\omega^S \in \Omega(\omega_j=0)} \Pr(\omega^S) f_R(m_1 | \omega^R, \omega^S) \left(v_j \frac{\sum_{\omega' \in \Omega(\omega_j=0)} \Pr(\omega') f_R(m_1 | \omega^R, \omega^S)}{\sum_{\omega' \in \Omega} \Pr(\omega') f_R(m_1 | \omega^R, \omega^S)} \right) \end{aligned} \quad (\text{B.8})$$

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

$$v_j \frac{\left[\sum_{\omega^S \in \Omega(\omega_j=1)} \Pr(\omega^S) f_R(m_1 | \omega^R, \omega^S) \right]^2 + \left[\sum_{\omega^S \in \Omega(\omega_j=0)} \Pr(\omega^S) f_R(m_1 | \omega^R, \omega^S) \right]^2}{\sum_{\omega^S \in \Omega} \Pr(\omega^S) f_R(m_1 | \omega^R, \omega^S)} \quad (\text{B.9})$$

²³Any elements of \mathcal{M} sent with 0 probability can be discarded from the mechanism without affecting the player's information penalties.

Correspondingly for message m_2 we have

$$v_j \frac{\left[\sum_{\omega^S \in \Omega(\omega_j=1)} \Pr(\omega^S) f_R(m_2|\omega^R, \omega^S) \right]^2 + \left[\sum_{\omega^S \in \Omega(\omega_j=0)} \Pr(\omega^S) f_R(m_2|\omega^R, \omega^S) \right]^2}{\sum_{\omega^S \in \Omega} \Pr(\omega^S) f_R(m_2|\omega^R, \omega^S)} \quad (\text{B.10})$$

Note that the total information penalty induced by sending m_1 and m_2 is the sum of equations B.9 and B.10.²⁴

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

$$v_j \frac{\left[\sum_{\omega^S \in \Omega(\omega_j=1)} \Pr(\omega^S) f_R(m_{12}|\omega^R, \omega^S) \right]^2 + \left[\sum_{\omega^S \in \Omega(\omega_j=0)} \Pr(\omega^S) f_R(m_{12}|\omega^R, \omega^S) \right]^2}{\sum_{\omega^S} \Pr(\omega^S) f_R(m_{12}|\omega^R, \omega^S)} \quad (\text{B.11})$$

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

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

for any $\Omega' \subseteq \Omega$. Therefore we can write equation B.11 as follows

$$v_j \frac{\left[\sum_{\omega^S \in \Omega(\omega_j=1)} \Pr(\omega^S) f_R(m_1|\omega^R, \omega^S) + \sum_{\omega^S \in \Omega(\omega_j=1)} \Pr(\omega^S) f_R(m_2|\omega^R, \omega^S) \right]^2 + \left[\sum_{\omega^S \in \Omega(\omega_j=0)} \Pr(\omega^S) f_R(m_1|\omega^R, \omega^S) + \sum_{\omega^S \in \Omega(\omega_j=0)} \Pr(\omega^S) f_R(m_2|\omega^R, \omega^S) \right]^2}{\sum_{\omega^S \in \Omega} \Pr(\omega^S) f_R(m_1|\omega^R, \omega^S) + \sum_{\omega^S \in \Omega} f_R(\omega^S) \Pr(m_2|\omega^R, \omega^S)} \quad (\text{B.12})$$

To find the difference in information penalties caused by combining messages m_1 and m_2 , we subtract equation B.12 from the sum of equations B.9 and B.10. This difference is equal to

²⁴We 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 B.8.

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

which is weakly positive. \square

Proposition 14. *Given straightforward inferences, type ω prefers to reveal trait j before trait k if and only if*

$$v_j \frac{1 - \pi(\rho_j^*(\omega))}{1 - \rho_j^*(\omega)} \leq v_k \frac{1 - \pi(\rho_k^*(\omega))}{1 - \rho_k^*(\omega)}$$

Proof. Suppose some type of sender ω^S has the most preferred grammar $g \in \mathcal{G}^*$ of the form $g = \{m_1, m_2, \dots, m_N\}$ where message m_i reveals trait $\beta(i)$. Suppose for some $i \in \{1, \dots, N-1\}$ 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 $g' = (m_1, \dots, m_{i-1}, m_{i+1}, m_i, m_{i+1}, \dots, m_N)$ 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

$$\begin{aligned} & - (1 - \rho_{\beta(i)}^*) * \left(\sum_{k=1}^i v_{\beta(k)} + \sum_{k=i+1}^N \pi(\rho_{\beta(k)}^*) v_{\beta(k)} \right) \\ & - (1 - \rho_{\beta(i+1)}^*) \rho_{\beta(i)}^* * \left(\sum_{k=1}^{i+1} v_{\beta(k)} + \sum_{k=i+2}^N \pi(\rho_{\beta(k)}^*) v_{\beta(k)} \right) \end{aligned} \quad (\text{B.13})$$

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

$$\begin{aligned}
& - (1 - \rho_{\beta(i)}^*) \rho_{\beta(i+1)}^* * \left(\sum_{k=1}^{i+1} v_{\beta(k)} + \sum_{k=i+2}^N \pi(\rho_{\beta(k)}^*) v_{\beta(k)} \right) - \\
& \quad (1 - \rho_{\beta(i+1)}^*) * \left(\sum_{k=1}^{i-1} v_{\beta(k)} + v_{\beta(i+1)} + \pi(\rho_{\beta(i)}^*) v_{\beta(i)} + \sum_{k=i+2}^N \pi(\rho_{\beta(k)}^*) v_{\beta(k)} \right)
\end{aligned} \tag{B.14}$$

Subtracting equation B.13 from equation B.14 yields a positive quantity if and only if

$$\frac{v_{\beta(i+1)}}{v_{\beta(i)}} \leq \frac{(1 - \rho_{\beta(i+1)}^*) (1 - \pi(\rho_{\beta(i)}^*))}{(1 - \rho_{\beta(i)}^*) (1 - \pi(\rho_{\beta(i+1)}^*))}$$

which is equivalent to equation 7.2 for traits $\beta(i)$ and $\beta(i+1)$. \square

Corollary 2. *Assume $v_i = v_j$ for all i and j and that π is differentiable. If π is strictly concave, players prefer the equilibrium in which traits are discussed from highest ρ_j^* to lowest ρ_j^* . If π is strictly convex, players prefer the equilibrium in which traits are discussed from lowest ρ_j^* to highest ρ_j^* .*

Proof. For a single agent, equation 7.2 can be written

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

Since π is differentiable we can write

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

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