UNACKNOWLEDGED HETEROGENEITY IN COMMUNICATION

KYLE P. CHAUVIN

NOVEMBER 12, 2020

JOB MARKET PAPER

[Latest Draft]

ABSTRACT. Successful communication with natural language requires a sender and receiver to share the same understanding of the meaning of words, or linguistic convention. What happens when people think they share the same convention, but in fact they do not? This paper shows how unacknowledged heterogeneity leads to systematic bias. I analyze a sender-receiver model in which both parties’ conventions are drawn independently from a population satisfying a basic regularity condition. Relative to linguistic anchor states that all conventions express identically – such as minimum, maximum, or midpoint states – the receiver on average exaggerates the sender’s intended message. The model therefore predicts over-reaction to information across a range of communication settings, such as economic forecasting, risk calibration, and doctor-patient relations. This occurs even when the parties have completely aligned interests. A sender and receiver with aligned interests can mitigate miscommunication by rephrasing the form of the message sent, garbling the sender’s information, introducing a mediator, or including a redundant second sender. When the two parties have conflicting interests, the receiver always suffers from unacknowledged heterogeneity, but a sender seeking to warp the receiver’s action can benefit from the exaggeration effect. When multiple, similarly biased, agents aggregate information by communicating in sequence, all agents will mistakenly perceive the sequence converging to a point belief, when in fact the limiting distribution is noisily centered on an exaggeration of the true state.

JEL Classification: C72, D83, D91, Z13

Keywords: Cheap-talk, Projection Bias, Social Learning, Natural Language

Department of Economics, Harvard University. Email: kchauvin@hbs.edu. My thanks to Roland Bénabou, Alison Wood Brooks, Christine Exley, Tristan Gagnon-Bartsch, Ben Golub, Faruk Gul, Anna Hopper, Wolfgang Pesendorfer, Muriel Niederle, Matthew Rabin, Andrei Shleifer, Joshua Schwartzstein, and Mike Yeomans for helpful conversations about this project.
1. Introduction

A sender and a receiver need two things to communicate successfully. The receiver must trust what the sender says, and the receiver must correctly interpret what the sender says. While the factors that promote or preclude trust have long been studied, the prospect of incorrect interpretation has received comparatively little attention. The reason is simple: standard equilibrium concepts impose Bayesian rationality, which means that however the sender expresses her message, and however it might be garbled in transmission, the receiver always understands what he observes.

There are nonetheless multiple reasons to worry about misunderstanding. The sender and receiver have to coordinate on a linkage between meaning and expression – what Lewis (1969) termed a convention – and coordination is difficult. Communication games feature a multiplicity of equilibria with varied selection and refinement criteria; the real world features tremendous diversity across and within languages. Coordination is also stymied when people do not adapt their conventions. Natural language is often taught prescriptively, creating the illusion of a single, ‘correct,’ way to use words. Additionally, the well-documented preference for honesty is predicated on words having objective meaning. Lastly, and most importantly, people are prone to projection bias, by which they over-extrapolate from their own circumstances and fail to acknowledge differences in others.

1Dating from Crawford and Sobel (1982), the literature on cheap-talk, in which a sender with private but unverifiable information sends a message to a receiver, has yielded many insights about communication. As Sobel (2012) notes, trust is paramount: “the quality of information communicated in equilibrium and the potential benefits of communication increase when the conflict between the Sender and Receiver decreases.” However, Rabin (1990) and Zapater (1997), among others, observe the following. If we consider any equilibrium of a cheap-talk game, successful communication requires not just (partial) separation of sender types, but also successful coordination by both parties on the meaning of messages.

2Different approaches to selection and equilibrium refinement can be found in, among others, Rabin (1990); Matsui (1991); Wärneryd (1991); Farrell (1993); Zapater (1997); Baliga and Morris (2002); Demichelis and Weibull (2008); Chen, Kartik and Sobel (2008); Lo (2020). See Farrell and Rabin (1996) for an engaging discussion.

3Regarding the chasm between linguists’ descriptive approach to language usage and teachers’ prescriptive approach, see McWhorter (2011). For a recent and economics-focused discussion of linguistic diversity, see Ginsburgh and Weber (2020).

4As further discussed in Section 6, both honesty preference (e.g. Abeler, Nosenzo and Raymond (2019)) and interpersonal projection bias (e.g. Loewenstein, O’Donoghue and Rabin (2003); Van Boven and Loewenstein (2003)) lead communicating parties to neglect how their counterpart might interpret words and instead concentrate on what they regard to be the ‘objectively true’ convention. Casella et al. (2018) find that while lab subjects can adjust to situation-specific feedback (e.g. learning that ‘I’ll give you $6’ actually means ‘I’ll give you $5’), their initial behavior reflects an assumption of objective meaning. Diversity of interpretation has been documented with respect to technical reports (Budescu, Por and Broomell, 2012) and doctor-patient communication (Ong et al., 1995). Weber and Camerer (2003) show a similar effect in a lab experiment simulating corporate culture clash.
This paper analyzes miscommunication arising from projection bias. It asks: what happens when a sender and receiver each believe the other shares the same convention, but in fact they do not? I show that failure to appreciate linguistic differences does not simply add noise to a classical equilibrium outcome. Instead, it creates a kind of systematic overreaction to information that I label the exaggerated effect. This has welfare consequences for all communication settings, notably ones without strategic frictions – such as an analyst talking to an investor within the same firm, a patient describing pain to her doctor, or a consultant presenting research to a client – where standard game theory predicts seamless coordination.

Consider an example. Susan the analyst tells Roberto the investor to expect ‘solid growth’ in the upcoming year. They both want his level of investment to accord with her true assessment of the economy, which happens to be 2% growth. She communicates with natural language instead of numbers as a matter of habit and congeniality, particularly since both parties believe they share a common linguistic convention. Suppose Susan’s and Roberto’s conventions are actually drawn independently from two equally likely conventions, one exuberant and one reserved. The word that the exuberant convention uses to label $X\%$ growth is used by the reserved convention to label $2X\%$ growth. Hence, if Susan and Roberto are either both exuberant or both reserved, Roberto will correctly interpret her statement as 2% growth. However, if Susan is exuberant and Roberto is reserved, he will interpret 4% growth, and if the conventions are reversed, he will interpret 1% growth. On average, his interpretation is 2.25%. Relative to the truth, that is a 12.5% exaggeration.

The content of this paper branches out from the example in two distinct ways. In Section 2, it employs a minimalistic framework to characterize how universally the exaggeration effect occurs. Sections 3 through 5 place additional assumptions on the structure of communication to analyze a fully-specified game, evaluate corrective policy measures, and consider extensions such as persuasion and social learning.

The general framework in Section 2 includes the following components. There is an interval of messages, an unstructured set of words, and a set of conventions, each of which defines a bijection between messages and words. In the example above, messages can be any growth percentage, and words are descriptors such as ‘slow,’ ‘encouraging,’ or ‘phenomenal’ growth. While the sender’s intended message is held fixed, the sender’s and receiver’s conventions are drawn independently from a single population distribution. To guarantee a base level of mutual coherence, the conventions
are assumed to rank words in the same order according to implied meanings. Apart from this requirement, conventions can have any functional form. Although each party wrongly assumes the other shares his/her own convention, sender and receiver nonetheless communicate successfully for particular messages, termed linguistic anchors, which all conventions express identically. In the example above, the anchor was 0% growth. Messages around an anchor can either be exaggerated away from the anchor or understated towards it.

The main result establishes that, under a simple regularity condition, average miscommunication exaggerates messages around anchors. The result is driven by differences in exuberance between conventions. One convention is more exuberant than another if it always uses more extreme words to express the same messages. Accordingly, the regularity condition is exuberance-consistency: for every pair of distinct conventions, one of them is more exuberant than the other.\(^5\) This matters because a message is exaggerated conditional on the sender being more exuberant than the receiver, while the message is understated when the sender is less exuberant. Because the sender’s and receiver’s conventions are drawn i.i.d., both possibilities are equally likely. However, a message is also distorted by a greater magnitude when the sender is more exuberant than when the sender is less exuberant. On average, exaggeration therefore supersedes understatement.

Several additional results add nuance to the exaggeration effect. When there is no anchor message that all conventions express with the same word, but instead each pair of conventions agrees at its own pairwise anchor, then messages outside the range of pairwise anchors are exaggerated with respect to all of the pairwise anchors. When more exuberant conventions also use higher words to express the same messages than do less exuberant ones, messages are systematically shifted upwards in expectation.\(^6\) Intuitively, the directional shift stems from the lower bound on messages acting like an anchor, with exaggeration manifesting as a directional shift. When conventions satisfy exuberance-consistency only on a neighborhood of an anchor, as can happen if the anchor is an endpoint of the message interval, then exaggeration is guaranteed locally rather than globally. Lastly, when there are only two conventions, then – even absent exuberance consistency – each anchor features messages arbitrarily close to it that are exaggerated.

\(^5\)For example, a more exuberant convention might describe ‘phenomenal’ growth, whereas a more reserved convention would call it merely ‘encouraging.’ The regularity condition asks, essentially, that one convention’s relative proclivity to extreme words vis-a-vis another varies gradually across the set of messages. See Section 2.1 for a formal discussion.

\(^6\)Conversely, when more exuberant conventions use lower words to express the same messages than do less exuberant ones, messages are systematically shifted downwards.
Section 3 studies the consequences of heterogenous conventions in a fully-specified communication game. The sender no longer has a fixed message, but rather, she privately observes a signal correlated with an unobserved state, and this informs her choice of message. The state and signal are jointly Gaussian, the sender’s message is an affine function of the signal, and both parties want the receiver’s action to minimize mean-squared error with the state. In the reference case where sender and receiver do share the same convention, there is a continuum of outcome-equivalent equilibria. Each equilibrium involves the sender’s message distorting the signal in some way and the receiver correctly reversing the distortion so that his action is always the conditional mean of the state.

Convention heterogeneity impacts the game as follows. With an additional regularity assumption about relative exuberance between conventions, the receiver’s interpretation is an random affine function of the sender’s intended message. Then, for any given equilibrium, both the mean-squared error and the mean bias of the receiver’s action are made strictly larger by the exaggeration effect. Moreover, both error and bias increase in the informativeness of the signal, the extent of convention heterogeneity, and the average misalignment between the sender’s message and the linguistic anchor. Sender and receiver do best in equilibria where the sender’s average message equals the anchor. They also suffer less from miscommunication when the sender sends a highly inflated version of her signal; that is, ironically, when they coordinate on strategies that involve exaggeration.

Section 4 discusses how the sender and receiver, or a benevolent third party, could mitigate miscommunication. First, if the sender and receiver were aware of convention heterogeneity but unable to identify the other’s convention, they would instead coordinate on down-weighting the signal in estimating the state. Alternatively, garbling the sender’s information by augmenting her signal with uninformative noise can help both parties when the mean of the noise works against the direction of miscommunication bias and the variance of the noise is not too high. Placing a mediator between the sender and receiver can likewise re-align what the receiver hears and eliminate bias. However, the mediator’s own convention introduces an additional source of error, bounding the mediator’s potential effectiveness. Finally, a receiver hearing from two senders is better off when the information they send is highly correlated, illustrating a benefit of redundancy in communication. One caveat with mitigation is that it introduces its own coordination problems. For example, when sender and receiver are aware of convention heterogeneity, attempt to down-weight their collective
estimate, but fail to coordinate properly, they can create even greater mismatch and end up worse for trying.

Section 5 analyzes two different ways to extend the model. First, as in the cheap-talk literature, we suppose that the sender and receiver have conflicting interests. While the receiver still aims to match the state, the sender balances two concerns: inducing the receiver’s action to match the state plus a privately-observed persuasion motive and honestly communicating her belief about the state.\(^7\) With convention heterogeneity, the receiver loses unambiguously from miscommunication. Across a wide range of parameters, the sender loses strictly less than the receiver does; in extreme cases, she even benefits from miscommunication. Intuitively, a sender hoping to convert the receiver’s action from matching the state to matching her personal preferred point can gain substantially from exaggeration pushing the receiver’s action to be more extreme. The second extension places convention heterogeneity in a social learning context. A sequence of agents, each with an independently realized convention, iteratively form opinions about the state by incorporating their private signals into (what they understand to be) their predecessors’ beliefs. As the sequence grows, the words used to express the agents’ opinions converge to a limit point, but the agents’ actual opinions converge to a non-degenerate asymptotic distribution centered on an exaggerated version of the true state.

Section 6 relates this paper to prior literature and discusses ways to extend the paper’s line of inquiry. Appendix A includes omitted proofs, and Appendix B interrogates the model’s key assumptions by reviewing potential variations. In particular, the exaggeration effect is qualitatively unchanged if, instead of drawing conventions \(i.i.d\), the joint distribution over the sender’s and receiver’s conventions is merely symmetric. Replacing the definition of a convention as a one-to-one function with a noisy correspondence between messages and words likewise preserves the effect. Systematically breaking the exaggeration effect requires more radical departures from model, including correlation between the state and the sender’s exuberance, correlation between exuberance and the pairwise anchor, and separate distributions of conventions for senders and receivers.

\(^7\)The only innovation in the framework used here is the Gaussian distribution of fundamental noise terms. A preference for honest communication has been introduced in other cheap-talk models, e.g. Kartik (2009); Khalmetski (2019). A principal consequence of partial honesty is language ‘inflation,’ whereby the sender’s message is shifted in the direction of her preferred state; I document this effect as well in the setting without convention heterogeneity.
2. The Exaggeration Effect

A privately-informed sender $S$ (‘she’) wishes to communicate a numeric message $m$ to an action-taking receiver $R$ (‘he’). Their interests are aligned, so they both want the receiver to learn the true $m$. However, the sender communicates indirectly. She first expresses her intended message $m$ as a ‘word’ $w$, which the receiver interprets as $\hat{m}$. This section treats $m$ as fixed, and, postponing consideration of other features of the environment, concentrates solely on the relationship between $m$ and $\hat{m}$. Subsequent sections formally situate the two parties in a fully-specified communication game using a restricted class of conventions.

Messages lie in an interval $\mathcal{M}$ of the real line, and words come from an unstructured set $\mathcal{W}$. Each linguistic convention $C \in \mathcal{C}$ defines a link between messages and words as codified by an expression function $w_C : \mathcal{M} \rightarrow \mathcal{W}$ and its inverse, an interpretation function $m_C = w^{-1}_C : \mathcal{W} \rightarrow \mathcal{M}$. The sole restriction on the set of conventions $\mathcal{C}$ is that different conventions rank the words in the same order according to the messages they represent. Formally, $m_C(w) \geq m_C'(w')$ if and only if $m_{C'}(w) \geq m_{C'}(w')$ for all convention pairs $(C,C')$ and all word pairs $(w,w')$. The monotonicity assumption ensures a baseline level of comparability between conventions; without it, the ordering of the message set loses all significance.

Sender $S$ and receiver $R$ use conventions drawn i.i.d. from $\mathcal{C}$ and independently of message $m$. Abusing notation slightly, the sender’s and receiver’s conventions are also denoted by $S$ and $R$. Hence, when $S$ intends to communicate message $m$, she expresses it with word $w_S(m)$. Similarly, $R$ interprets $w$ according to his own convention, interpreting message $m_R(w_S(m))$. The composition $h \equiv m_R \circ w_S : \mathcal{M} \rightarrow \mathcal{M}$, which itself depends on how the conventions are realized, is termed the communication function.

This framework models various environments. For example, $S$ could be a financial analyst communicating her growth forecast $m$ to $R$, an investor. Her word is a descriptor like ‘solid,’ ‘firm,’ ‘promising,’ or ‘robust,’ and the set of messages $\mathcal{M}$ is the entire real line. Alternatively, $S$ could be a consultant reporting the probability a firm’s bankruptcy, a plane crash, successful vaccine development, or other event, to client $R$. In this case, the set of messages is the unit
interval \([0,1]\). In another alternative, \(S\) is a doctor diagnosing patient \(R\) with a ‘mild,’ ‘severe,’ or ‘worrying’ ailment. Or \(S\) could be a patient describing her ‘irritating’ or ‘excruciating’ pain to doctor \(R\). The point is that people communicate with natural language across a wide spectrum of social environments. Figure 1 below illustrates the complete process for the analyst / investor example.

![Figure 1. Example Conventions. Analyst \(S\) communicates growth forecast \(m\) to investor \(R\) using words in the center of the figure. The anchor between the two is \(m = 0\), which both \(S\) and \(R\) express as ‘no growth.’ Non-anchor messages are exaggerated away from the anchor when \(S\) communicates to \(R\). Conversely, if the two conventions were flipped, communication would compress non-anchor messages towards the anchor.]

2.1. Main Result

Miscommunication between the sender and receiver can take various forms depending on which conventions are realized. The receiver’s interpretation may be higher or lower than the sender’s intention, and nothing more specific can be said from the ex-post perspective. However, the primary observation of this paper is that the ex-ante distribution over the receiver’s interpretation has systematic features which depend very little on the set of conventions. To characterize these formally, we concentrate on the average communication function \(E[h(\cdot)]\).

We start by observing that many environments feature particular messages, like endpoints or midpoints (e.g. 0% growth in Figure 1), which all conventions express identically. This is termed a
linguistic anchor: \( a \in \mathcal{M} \) such that \( w_S(a) = w_R(a) \) for all \( S, R \in \mathcal{C} \). The sender and receiver do not realize that anchors have particular significance, yet, by definition, no miscommunication occurs at an anchor. However, anchors can be used to describe miscommunication at non-anchor messages. Say that average communication is an \textit{exaggeration} with respect to anchor \( a \) if \( |\mathbb{E}[h(m)] - a| > |m - a| \). Intuitively, exaggeration ‘stretches’ intended messages away from the anchor. The opposite phenomenon, by which messages are compressed towards the anchor, is termed \textit{understatement}. Figure 2 illustrates these possibilities.

What critically determines the shape of average communication around an anchor is the proclivity of conventions to use extreme words, a quality I label ‘exuberance.’

\textbf{Definition.} Convention \( C \) is \textbf{more exuberant} than \( C' \) if

\[ |m_C(w) - m_C(w')| < |m_{C'}(w) - m_{C'}(w')| \]

for all words \( w \neq w' \).

Intuitively, a more exuberant convention utilizes a broader range of words. In Figure 1, for example, the sender’s convention is more exuberant than the receiver’s. As growth increases from 0% to 4%, the sender transitions from describing ‘no growth’ to ‘terrific’ growth, while the receiver merely
reaches ‘encouraging’ growth. To quantify exuberance, we rephrase it in terms of the ranges of messages. A more exuberant convention understands a given pair of words to signify a shorter interval of meanings. For example, whereas the sender in Figure 1 understands the difference between ‘no growth’ and ‘terrific’ growth to be 4%, the less exuberant receiver thinks those same words span an interval of 12%.

Relative exuberance between a pair of conventions may vary with the pair of comparison words. This paper considers different possibilities, concentrating on the simplest arrangement.

**Definition.** Conventions are **exuberance-consistent** if, for any pair of distinct conventions, one is more exuberant than the other.

Consistency requires that, whenever $S$ is found to be more exuberant than $R$ according to one pair of words, then $S$ is guaranteed to be everywhere more exuberant than $R$. It models an environment in which some speakers are equally profligate with their words when describing high states or low states, good news or bad news, etc., while other speakers are consistently reserved. Under this regularity condition, average communication exhibits a stark pattern: exaggeration.

**Proposition 1.** If conventions are exuberance-consistent and share linguistic anchor $a$, then average communication is an exaggeration relative to $a$. Proof in the appendix.

The exaggeration effect established by Proposition 1 is generated by the confluence of several factors. Intuitively, because the sender’s and receiver’s conventions are drawn i.i.d., the communication function is generated in a two-step process. First, two conventions are drawn from the population, and then they are independently assigned to the sender and receiver with equal probability. Hence, the average communication function $\mathbb{E}[h(\cdot)]$ is equal to the expectation of $(1/2)(h_{SR}(\cdot) + h_{RS}(\cdot))$. Demonstrating exaggeration for each pair of distinct conventions is therefore sufficient to prove the proposition.

To that end, consider any fixed pair of distinct conventions $S$ and $R$, where without loss of generality $S$ is everywhere more exuberant than $R$. Imagine $h_{SR}$ as describing ‘forward’ communication from $S$ to $R$, while $h_{RS}$ captures ‘reverse’ communication from $R$ to $S$. Then, forwards
communication results in a greater magnitude of miscommunication than does reverse communication: $|h_{SR}(m) - m| > |h_{RS}(m) - m|$.\(^9\) Moreover, forwards communication exaggerates the message relative to the anchor, $|h_{SR}(m) - a| > |m - a|$, while reverse communication understates it: $|h_{RS}(m) - a| < |m - a|$.\(^10\) These two properties show that the magnitude of exaggeration in the forwards direction is greater than the magnitude of understatement in the reverse direction. Thus, their unweighted average is guaranteed to be an exaggeration.

### 2.2. Complementary Results

The two suppositions in Proposition 1 – that conventions are exuberance-consistent and share a common anchor – can each be loosened without qualitatively changing the exaggeration effect. This subsection demonstrates four different ways to do so.

First, we address the possibility that the message domain lacks an endpoint, midpoint, status-quo point, or other natural candidate for a linguistic anchor. For example, colors, aromas, and other subjective sensory perceptions are often conceptualized on as a scale, but without a fixed reference point that would serve as an anchor. In place of a common anchor, instead suppose that each convention pair $(S, R)$ has its own, potentially distinct, pairwise anchor $a_{SR}$ such that $w_{S}(a_{SR}) = w_{R}(a_{SR})$.

**Proposition 2(a).** If conventions are exuberance-consistent, and each pair $S \neq R$ shares a pairwise anchor $a_{SR}$ contained in the range $[a, \pi]$, then average communication is an exaggeration of all values $m > \pi$ and $m < a$ relative to all pairwise anchors. *Proof in the appendix.*

The result is that extreme messages are exaggerated. Intuitively, when we condition on a particular convention pair, messages are exaggerated away from the pairwise anchor per Proposition 1. Therefore, any message that is greater than all pairwise anchors is unambiguously miscommunicated upwards, and any message lower than all pairwise anchors is pushed downwards. Messages inside the support of the pairwise anchors are buffeted by opposing forces – exaggerated upwards.

\(^9\)In the forwards direction, the magnitude of miscommunication $|h_{SR}(m) - m|$ equals the receiver’s perception of the difference in meaning between the sender’s word for the message, $w_{S}(m)$, and the receiver’s word for the message, $w_{R}(m)$. In reverse, it equals the sender’s perception of the gap between those two words, which, because the sender is assumed to be more exuberant, is strictly smaller.

\(^10\)In the forwards direction, the difference between the (miscommunicated) message and the anchor is equal to the receiver’s perception of the gap between the sender’s word for the message and the sender’s word for the anchor. Because the receiver is less exuberant, that perception is strictly less than the sender’s perception of the gap between the two words, which is equal to the difference between the actual message and the anchor. This means the miscommunication is an exaggeration. Understatement in the reverse direction follows from a parallel argument.
conditional on low pairwise anchors but exaggerated downwards by high pairwise anchors – making the aggregate effect ambiguous. When the support narrows to a single point, the statement of the result matches Proposition 1.

We next consider the possibility that conventions share no anchors, common or pairwise. Because of the foundational assumption that conventions rank words in the same order, two conventions can only fail to share a pairwise anchor if one of them always uses a higher word to express the same message, or, equivalently, if one of them always interprets the same word as signifying a lower message. Formally, say that convention $S$ is higher than $R$ if $m_S(w) < m_R(w)$ for all words $w$.

**Proposition 2(b).** Suppose conventions are exuberance-consistent. If $S$ is higher than $R$ whenever $S$ is more exuberant, then $E[h(m)] > m$ for all $m \in M$. If $S$ is lower than $R$ whenever $S$ is more exuberant, then $E[h(m)] < m$ for all $m \in M$. Proof in the appendix.

The result is that miscommunication is directional: messages as miscommunicated upwards when more exuberant conventions are also higher, and messages are miscommunicated downwards when more exuberant conventions are also lower. At first glance, this may seem to describe something other than exaggeration, but it is closely related to Proposition 1. Suppose for the sake of illustration that more exuberant conventions are higher and that the message space is all of $\mathbb{R}$. For any pair of conventions $(S, R)$, where $S$ is both more exuberant and higher than $R$, the value of $h_{SR}(m) - m$ converges to a constant $K \geq 0$ as $m \to -\infty$. Similarly, $h_{RS}(m) - m$ approaches $-K$ as $m \to -\infty$. Hence, $-\infty$ resembles an anchor, although it is not even included in the message set, so the statement $E[h(m)] > m$ is essentially claiming that average communication is an exaggeration relative to $-\infty$. In the opposite case, when more exuberant conventions are lower, it is $+\infty$ that acts like an anchor, causing messages to be miscommunicated downwards instead of upwards.

Next, we return to dealing with a common anchor but relax the assumption that exuberance-consistency holds globally. In many cases, it is either impossible or improbable that one convention $S$ will be more exuberant than convention $R$ across all words and messages. For example, if the message set is closed and bounded – as when messages are probability values – then all conventions
agree that the highest word\textsuperscript{11} corresponds to the highest message, the lowest word corresponds to the lowest message, and the gap between the highest and lowest words is the length of $\mathcal{M}$. Hence no convention can be everywhere more exuberant than another. However, we can still consider exuberance consistency and exaggeration on a local scale. To bypass the technical quibble that different conventions interpret a fixed neighborhood of messages $M \subset \mathcal{M}$ to correspond to different neighborhoods of words, we call a word $w$ ‘relevant’ to $M$ if there exists any convention $C$ such that $m_C(w) \in M$. Then, say that conventions are \textit{locally} exuberance-consistent on $M \subset \mathcal{M}$ if, for any pair of distinct conventions, one is more exuberant than the other between all pairs of distinct words relevant to $M$.

\textbf{Proposition 2(c).} \textit{If conventions are locally exuberance-consistent on interval $M$ containing anchor $a$, then average communication is an exaggeration relative to $a$ on $M$. Proof in the appendix.}

The result confirms that the same dynamics which yield Proposition 1 also produce local exaggeration under local exuberance-consistency. Aside from endpoint cases, such as low probabilities exaggerated upwards from 0 and high probabilities exaggerated downwards from 1, the local result also applies to scenarios in which exuberance ordering switches from one side of an anchor to the other. For example, take the case of growth forecasting and suppose that some conventions are ‘optimists’ – exuberantly expressing positive growth while reservedly expressing negative growth – and that some conventions are similarly pessimists. If the conventions are locally exuberance-consistent on the range of positive growth, and separately exuberance-consistent on the range of negative growth, then average communication is an exaggeration \textit{everywhere}, despite the lack of global exuberance-consistency.

Last, we relax exuberance-consistency entirely as a way of further clarifying its role in the main result. However, immediate consequence of dropping the condition is a technical complication: without exuberance-consistency, one anchor may be the limit point of a sequence of other anchors, so describing the effect of miscommunication ‘around’ the anchor loses meaning. The dynamics of individual convention pairs also fail to aggregate cleanly. However, if we side-step these hiccups by restricting attention to isolated anchors with only a single pair of distinct conventions, a clear pattern of exaggeration remains.

\textsuperscript{11}There is a single highest word when there is a single highest message. (See Appendix B.3 for a discussion of continuity between conventions).
Proposition 2(d). If there are two conventions, and \( a \) is an isolated anchor, then for all \( \varepsilon > 0 \) there exists a message \( m \in B_\varepsilon(a) \) such that average communication is an exaggeration of \( m \) relative to \( a \). Proof in the appendix.

The result states that, although we cannot guarantee exaggeration uniformly around an anchor, exaggeration nonetheless occurs infinitely often around anchors. In other words, with two conventions, exaggeration is the only kind of miscommunication that can be found on an interval around an anchor. Consistent understatement, local or global, is impossible. Actually produce examples of understatement requires scrapping the fundamental monotonicity assumption as well.\(^{12}\)

2.3. Special Case: Affine Communication Functions

The exaggeration effect can be seen even more transparently if we strengthen exuberance consistency as follows. Fix a pair of distinct conventions \( S \) and \( R \). Then, instead of assuming the ratio
\[
\frac{m_R(w_2) - m_R(w_1)}{m_S(w_2) - m_S(w_1)}
\]
remains either strictly greater than 1 (if \( S \) is more exuberant than \( R \)) or strictly less than 1 (if \( S \) is less exuberant than \( R \)) as we vary the pair of comparison words, suppose instead that it equals some constant \( b_{SR} \neq 1 \). Conventions with constant exuberance ratios are guaranteed\(^{13}\) to have affine communication functions. That is, \( h_{SR}(m) = m + (b_{SR} - 1)(m - a) \), where \( a \) is the anchor and \( b_{SR} \) is the exuberance ratio between \( S \) and \( R \). (Appendix B.2 discusses the implications of a random pairwise anchor \( a_{SR} \).) Affine communication functions constitute a tractable functional subclass that facilitates incorporating convention heterogeneity into richer economic models. Moreover, they allow us to prove exaggeration in a simple but illuminating way.

In fact, there are two distinct arguments. Both methods establish that \( \mathbb{E}[b_{SR}] > 1 \), which is equivalent to exaggeration around anchor \( a \). The first proof mirrors the logic of Proposition 1: as

\(^{12}\)As a simple example, suppose that \( h_{SR}(m) = h_{RS}(m) = -m \), and that one convention has probability 2/3 and the other 1/3. It follows that average communication is \( \mathbb{E}[h(m)] = (5/9)m + (4/9)(-m) = m/9 \), an understatement.

\(^{13}\)Proof: take \( S \neq R \). Monotonicity implies \( b_{SR} > 0 \), hence \( m_R(w_2) - m_R(w_1) = b_{SR}(m_S(w_2) - m_S(w_1)) \). Let \( m_0 \) be an arbitrary, temporary, reference point and let \( m \) be any message. Setting \( w_2 = w_S(m) \) and \( w_1 = w_S(m_0) \) gives \( h_{SR}(m) = h_{SR}(m_0) + b_{SR}(m - m_0) \). By the monotonicity assumption, \( b_{SR} \neq 1 \), so there is a unique value \( a_{SR} \) at which \( h_{SR}(a_{SR}) = a_{SR} \). Setting \( m_0 = a_{SR} \) produces the above formula.
communication $h_{SR}$ is equally likely as its inverse $h_{RS}$,

$$E[b_{SR}] = \frac{1}{2} E[b_{SR} + b_{RS}] = \frac{1}{2} E \left[ b_{SR} + \frac{1}{b_{SR}} \right] = E \left[ \frac{b_{SR}^2 + 1}{2b_{SR}} \right] > E[1] = 1.$$  

The second method fixes an arbitrary convention $C^* \in \mathcal{C}$ as a numeraire. Then, noting any exuberance ratio $b_{SR}$ can be expressed as $b_{SC^*}/b_{RC^*}$, where the factors $b_{SC^*}$ and $b_{RC^*}$ are independent of each other,

$$E[b_{SR}] = E \left[ \frac{b_{SC^*}}{b_{RC^*}} \right] = E[b_{SC^*}] E \left[ \frac{1}{b_{RC^*}} \right] = E[b_{SC^*}] E \left[ \frac{1}{b_{SC^*}} \right] > E[b_{SC^*}] \frac{1}{E[b_{SC^*}]} = 1.$$  

Most critical here is the convexity of the reciprocal function, so that the inequality follows immediately from Jensen’s inequality. Perhaps surprisingly, this argument appears not to generalize beyond the affine special case. Exaggeration in the general case is a statement about $E[h(m)]$, where – in contrast to Jensen’s inequality – the argument is fixed and the function is random. While the role of convexity in the special case proof hints that some more powerful cousin of Jensen’s inequality is working behind the scenes more broadly, none has been found yet. Certainly, the communication functions themselves need not be convex. An example, if there are two conventions with communication function

$$h_{SR}(m) = \int_0^m 2 + \sin \frac{1}{\tilde{m}} \ d\tilde{m}$$

and its inverse, the conventions are exuberance-consistent on all of $\mathbb{R}$, but function $h_{SR}$ is neither convex (nor differentiable) on any neighborhood of $a = 0$.

### 3. Communication Game

Moving beyond the general results of the last section, we now study a more restricted but fully specified environment. Suppose that the sender’s message regards an unobserved fundamental state $\omega$. The sender $S$ privately observes signal $\theta$, which is jointly normally distributed with $\omega$. Then, she sends message $m$ to the receiver $R$, and $R$ chooses action $x \in \mathbb{R}$. The parties share a common utility function $U(x|\omega) = -E[(x-\omega)^2]$, which means they both want $R$’s action to minimize mean-squared error with the state.

---

14 Means, variances, and covariances are denoted as usual with $\mu$ and $\sigma$; the correlation coefficient is denoted $\rho_{\omega\theta} = \sigma_{\omega\theta}/(\sigma_{\omega}\sigma_{\theta})$. Conditional means and variances are denoted e.g. $\sigma_{\omega\theta}^2$. Regression coefficients and values that act similarly are denoted with $\beta$. For example, $E[\omega|\theta] = \mu_\omega + \beta_{\omega|\theta}(\theta - \mu_\theta)$, where $\beta_{\omega|\theta} = \sigma_{\omega\theta}/\sigma_\theta^2$.  

14
To study the effect of convention heterogeneity, we first consider $S$ and $R$ playing a classical equilibrium of the communication game. The equilibrium concept is perfect Bayesian equilibrium. We restrict attention to equilibria in affine strategies, where $S$’s message and $R$’s action are of the form

$$m(\theta) = \mu_m + \beta_S(\theta - \mu_\theta) \quad \text{and} \quad x(m) = \mu_x + \beta_R(m - \mu_m).$$

Except for where explicitly indicated otherwise, the players’ strategies are not restricted to affine functions. We then add convention heterogeneity. As in Section 2.3, it is assumed different conventions have constant exuberance ratios between conventions, which guarantees that communication functions are affine. Hence, when $S$ intends message $m$, $R$ in fact receives $\hat{m} = m + (b-1)(m-a)$, where $b$ is the (random) exuberance ratio between $S$ and $R$ and $a$ is the linguistic anchor. Information distortion and loss are measured with two quantities. The statistical bias $E[x-\omega]$ measures the expected difference between the receiver’s action and the true state. Similarly, the mean-squared error $E[(x-\omega)^2]$ quantifies of the amount of information transmitted in equilibrium.

3.1. Equilibrium

We first consider equilibria of the game without convention heterogeneity. Note that for a fixed pair of sender and receiver strategies, the receiver’s action is equal to $x = \mu_x + \beta_S\beta_R(\theta - \mu_\theta)$. In equilibrium, the values of $\mu_x$ and $\beta_S\beta_R$ are such $x$ is a regression of the state on the signal. The value of $\beta_S\beta_R$ is the effective regression coefficient.

**Fact 1.** Strategies $m(\cdot)$ and $x(\cdot)$ are an equilibrium if only if (1) the receiver’s action is centered on the mean state, $\mu_x = \mu_\omega$, and (2) the effective regression coefficient is

$$\beta_S\beta_R = \frac{\sigma_{\omega\theta}}{\sigma_\theta^2}.$$

Equilibrium bias and error are minimal: $E[x(m) - \omega] = 0$, and $E[(x(m) - \omega)^2] = \sigma_\omega^2/\sigma_\theta^2$.

Equilibrium features the sender scaling the signal residual $(\theta - \mu_\theta)$ by an arbitrary value of $\beta_S$, which the receiver matches with $\beta_R$ so that the product $\beta_S\beta_R$ is the correct regression coefficient. Hence, all equilibria are informationally equivalent. If the sender sends her opinion of the state given her private information, $m = E[\omega|\theta]$, then the receiver’s corresponding equilibrium strategy is to follow the message: $x = m$. If the sender directly passes along her private information, $m = \theta$, then the receiver first conditions the state on that information, $x = \mu_\omega + (\sigma_{\omega\theta}/\sigma_\theta^2)(m - \mu_\theta)$. Because
the sender’s opinion of the state is always perfectly recoverable, the receiver’s action in equilibrium is always \( x = \mathbb{E}[\omega|\theta] \). It is therefore unbiased, and the error reflects residual uncertainty about the state that is not informed by the signal. Mathematically, error increases with the ex-ante variance of the state and decreases in the magnitude of the correlation coefficient between the state and the signal.

With convention heterogeneity, \( R \) interprets \( \hat{m} = m + (b - 1)(m - a) \) and therefore takes action

\[
x = \mu_\omega + b \cdot \beta_R \beta_S (\theta - \mu_\theta) + (b - 1) \cdot \beta_R (\mu_m - a).
\]

There are two modifications relative to the frictionless baseline. First, the effective regression coefficient on the signal, \( \beta_S \beta_R \), is augmented as \( \beta_S \beta_R \cdot b \), and second, there is an additional offset term proportional to the misalignment \( \mu_m - a \). From this expression it is immediate that \( \mathbb{E}[x|m] = \mathbb{E}[\omega|\theta] + \beta_R \mathbb{E}[b - 1](m - a) \), i.e. the conditional exaggeration of the receiver’s action given the affine functional form. The following Fact computes the unconditional bias as well as the error that arises with convention heterogeneity.

**Fact 2.** With convention heterogeneity, bias and error given equilibrium strategies \( m(\cdot) \) and \( x(\cdot) \) are

\[
\mathbb{E}[x(\hat{m}) - \omega] = \beta_R \cdot \mathbb{E}[b - 1] \cdot (\mu_m - a)
\]

and

\[
\mathbb{E}[(x(\hat{m}) - \omega)^2] = \mathbb{E}[(x(m) - \omega)^2] + (\sigma_\omega^2 - \sigma_{\omega|\theta}^2) \cdot \mathbb{E}[(b - 1)^2] \cdot \left[ 1 + \left( \frac{\mu_m - a}{\beta_S \cdot \sigma_\theta} \right)^2 \right].
\]

On average, the extra term in \( R \)'s action introduces bias in the direction of \((\mu_m - a)\), the gap between the average state and the linguistic anchor. The magnitude of bias increases both in the size of that gap and the level of heterogeneity in exuberance as captured by \( \mathbb{E}[b - 1] \). (See below for a discussion about how measures such as \( \mathbb{E}[b - 1] \) and \( \mathbb{E}[(b - 1)^2] \) are determined by fundamentals of the communication environment.) Similarly, error relative to the frictionless baseline increases with the expectation of \((b - 1)^2\). As with bias, error also increases with the absolute value of the gap \((\mu_m - a)\). Error also increases both with the ex-ante variance of the state and the correlation coefficient between state and signal. Intuitively, the more that the agents have to learn from \( S \)'s signal, the higher the stakes of the game, and therefore the more that they have to lose from miscommunication.
What is most intriguing about the bias and error expressions in Fact 2 is that, while all equilibria in the baseline game are outcome equivalent, there is a meaningful interaction between the particular equilibrium played by S and R and the welfare consequences of convention heterogeneity.

**Corollary.** Bias and error are jointly minimized when either (1) the sender’s average message aligns with the anchor: \( \mu_m \rightarrow a \), or (2) the sender inflates her message, which the receiver deflates: \( \beta_S \rightarrow \infty \) and \( \beta_R \rightarrow 0 \).

As noted by the corollary, S and R are best off when they rephrase the form of the sender’s message to align with the anchor. Perfectly aligning the two takes away all bias and diminishes the mean squared error. Less intuitively, larger values of \( \beta_S \) also diminish error, and, as \( \beta_S^2 \rightarrow \infty \), the error is reduced to its minimum possible level. By coordinating on exaggeration, S and R actually end up fighting against the exaggeration bias from miscommunication.

**Measuring Heterogeneity and Welfare Loss.** Multiple statistics of random exuberance ratio \( b \) are incorporated into bias, error, and welfare calculations. In particular, quantities \( \mathbb{E}[b - 1] \) and \( \mathbb{E}[(b - 1)^2] \) appear often. Although the precise relationships between the different statistics vary by parameterization, the statistics are generally qualitatively equivalent.

Note that, because the probability of \( b \) must equal the probability of \( 1/b \), any valid distribution over \( b \) is log-symmetric about zero: \( F(b) + F(b^{-1}) = 1 \) for all \( b \). Distribution \( F_2 \) is said to be more heterogeneous than \( F_1 \) if \( F_2(b) < F_1(b) \) for all \( b > 1 \), or, equivalently, if \( F_2(b) > F_1(b) \) for all \( b < 1 \). Intuitively, \( F_2 \) is more heterogenous if it ‘spreads out’ the mass of \( F_1 \). Using this definition, we can clarify what it means for a statistic of the \( b \) distribution to capture the degree of heterogeneity in \( b \). Then, given a family of distributions \( \mathcal{F} \) over \( b \), statistic \( g : \mathcal{F} \rightarrow \mathbb{R} \) is a **heterogeneity indicator** if \( g(F_0) = 0 \) for the degenerate distribution \( F_0 \), and \( g(F_2) > g(F_1) \) whenever \( F_2 \) is more heterogeneous than \( F_1 \).

The most common statistics of \( b \) in this paper’s calculations are heterogeneity indicators, as cataloged in the following fact. Some of the statistics are not guaranteed to be heterogeneity indicators over all distributions, but are so when restricting to natural subclasses. One important parametrized family is that of two equally likely conventions: a 1/2 probability of the exuberance ratio being 1, a 1/4 probability of it being \( B \) for some value \( B \neq 1 \), and a corresponding probability 1/4 of it being \( 1/B \). Greater heterogeneity corresponds to larger values of \( B \). Second is a log-normal distribution over \( b \): \( \log b \sim \mathcal{N}(0, \sigma_b^2) \), where greater heterogeneity corresponds to greater...
values of $V[\log b] \equiv \sigma_b^2$. As illustrated in the Figure 3, all six of the statistics covered in Fact 3 have qualitatively the same shape.

**Fact 3.** Functions $E[b-1]$, $E[(b-1)^2]$, and $E[b(b-1)]$ are heterogeneity indicators over the set of all $b$ distributions. Additionally, $V[b]$, $E[b^2]/E[b]-1$, and $C[b_1,b_2]$ are heterogeneity indicators for the two-convention and log-normal families. *Proof in the appendix.*

An immediate corollary of Fact 3 is that $E[b^2] > E[b]$ and $E[(b-1)^2] > E[b-1]$ for any non-degenerate distribution over $b$.

![Figure 3](image.jpg)

**Figure 3.** Measures of Convention Heterogeneity. Left: two conventions; exuberance ratio $b = 1$ with probability $1/2$, and $b = B$ and $b = 1/B$ with probability $1/4$ each. Right: lognormal conventions; exuberance ratio $b$ with $\log b \sim N(0,\sigma_b^2)$. Graphs show $E[b-1]$ (blue), $E[(b-1)^2]$ (orange), $E[b^2-b]$ (purple), $V[b]$ (red), $E[b^2]/E[b]-1$ (green), and $C[b_1,b_2]$ (brown) for $B \in [1,3]$ (left) and $\sigma_b^2 \in [0,2/3]$ (right).

### 4. Mitigation Methods

This section examines four ways to mitigate welfare loss from miscommunication. First is awareness, in which the sender and receiver realize they may have different conventions, but each is unable to identify the other’s convention. The next two methods feature a benevolent third party who intervenes, either by modifying the signal observed by the sender or by introducing an additional, mediating, agent between $S$ and $R$. Last is redundancy, in which the sender is replaced by two separate senders reporting independent or correlated information to a single receiver.

#### 4.1. Awareness

Suppose $S$ and $R$ are cognizant of convention heterogeneity, yet remain unable to discern the other’s actual convention. This hypothetical can be understood both as one form of mitigating
loss as well as a counterfactual by which to evaluate other mitigation measures. From their newly enlightened perspective, sender and receiver regard convention mismatch like technological garbling of information. The sender sends her message anticipating that
\[ R \] will receive \( \hat{m} = m + (b - 1)(m - a) \), knowing the anchor \( a \) but not the random exuberance ratio \( b \). We now consider equilibria of this modified game.\(^{15}\)

**Fact 4.** When sender and receiver are aware of convention heterogeneity, strategies \( m = \mu_m + \beta_S(\theta - \mu_\theta) \) and \( x = \mu_x + \beta_R(m - \mu_m) \) are an equilibrium if
\[
\mu_x = \mu_\omega, \quad \mu_m = a, \quad \text{and} \quad \beta_S \beta_R = \frac{E[b]}{\sigma_\theta^2} \cdot \frac{\sigma_\omega^2}{\sigma_\theta^2}.
\]

In any equilibrium, bias is zero and error is
\[
E[(x - \omega)^2] = \sigma_\omega^2 (1 - \rho^2) + \sigma_\omega^2 \rho^2 (1 - E[b]^2/E[b^2]).
\]

*Proof in the appendix.*

Unlike in the baseline case, where \( S \) and \( R \) are unaware of convention heterogeneity, the sender’s message is pinned down in equilibrium. As any non-anchor message mean \( \mu_m \neq a \) would prompt the aware sender to reduce exaggeration by deviating her strategy in the direction of \( a \), the unique equilibrium value is \( \mu_m = a \). The effective regression coefficient on the sender’s signal, \( \beta_S \beta_R \), also reflects the parties’ increased consciousness. Fact 4 shows that the appropriate regression coefficient strictly discounts the baseline value of \( \sigma_\omega^2/\sigma_\theta^2 \) by the factor \( E[b]^2/E[b^2] \), which decreases in the level of heterogeneity. However, the task of imbuing the receiver’s action with additional conservatism can be split in any fashion between the two parties. Lastly, the awareness case provides an upper bound on the parties’ welfare in the presence of convention heterogeneity. The gap between the error in the unaware benchmark minus its counterpart in the awareness counterfactual simplifies to
\[
\sigma_\omega^2 \cdot \rho^2 \cdot \frac{V[b]}{E[b^2]},
\]
which increases with greater heterogeneity.

Somewhat paradoxically, the solution to mis-coordination itself involves a practically identical coordination problem. If the sender and receiver do not properly coordinate on the remedy, they may be worse off for trying. To illustrate this point, suppose that each agent assumes that \( \beta_S =

\(^{15}\)The restriction to affine strategies does bind here: because \( \hat{m} \) is not jointly normally distributed with the state, Bayes’ rule does not have an affine functional form.
\( \alpha(\mathbb{E}[b]/\mathbb{E}[b^2])(\sigma_{\omega\theta}/\sigma_{b}^2) \), where factor \( \alpha \) is also drawn \( i.i.d. \) and independently from the other random variables. Then, the realized regression coefficient is

\[
\beta_S \beta_R = (\alpha_S/\alpha_R)(\mathbb{E}[b]/\mathbb{E}[b^2])(\sigma_{\omega\theta}/\sigma_{b}^2),
\]

yielding mean squared error

\[
\mathbb{E}[(x - \omega)^2] = \sigma_{\omega}^2(1 - \rho_{\omega\theta}^2) + \sigma_{\omega}^2 \rho_{\omega\theta}^2 \mathbb{E}[(\alpha_S \alpha_R^{-1} b - 1)^2],
\]

and this, per the same logic as in Proposition 1, can be arbitrarily large as heterogeneity in \( \alpha \) increases.

4.2. Third-Party Garbling

As shown by the awareness counterfactual, mitigating welfare loss from convention mismatch involves both correcting the signal-anchor misalignment and introducing greater conservatism in the parties’ collective regression of the state on the signal. One plausible method to achieve these goals directly targets \( S \)'s signal. A third-party, such as a higher level manager overseeing both \( S \) and \( R \), could change the distribution of information \( S \) can access. If nothing else, the manager could employ a ‘useful idiot’ to weaken the quality of \( S \)'s signal by offering uninformed opinions.

Mathematically, we consider \( S \)'s signal as modified from \( \theta \) to \( \theta + \varepsilon \), where \( \varepsilon \) is independent of both \( \theta \) and \( \omega \). The noise term \( \varepsilon \) accordingly ‘garbles’ the signal.

The most critical determinant of optimal garbling is whether or not the sender and receiver are aware of it. If they remain unaware, the third party garbler can tailor the specification of noise \( \varepsilon \) to the particular message form employed by the parties. Otherwise, the third party must anticipate \( S \) and \( R \) re-adjusting in the presence of garbling.

**Fact 5.** **Optimal third-party garbling is determined as follows.** If \( S \) and \( R \) are aware of it, and

- if heterogeneity is sufficiently small, then no garbling is optimal: \( \varepsilon^* \equiv 0 \).
- if heterogeneity is sufficiently large, optimal garbling acts to stop communication altogether: \( \sigma_{\varepsilon}^2 \nearrow \infty \).

If \( S \) and \( R \) are unaware, then the optimal garbling offsets the message’s misalignment with the anchor: \( (\mu^*_\varepsilon - a)(\mu_m - a) < 0 \); garbling can only be welfare-improving when the misalignment and degree of heterogeneity are sufficiently small. **Proof in the appendix.**
As Fact 5 establishes, the scope of garbling in the presence of a savvy sender and receiver is highly limited. The mean of noise term \( \varepsilon \) can do nothing itself, as \( S \) and \( R \) de-mean the variable when conditioning on it. Any increase in the variance of \( \varepsilon \) weakens the quality of the sender’s signal. Accordingly, there are only two possibilities. If the miscommunication from heterogeneity is mild, and \( S \) and \( R \) benefit from communication in the absence of garbling, then garbling can only make them worse off. On the other hand, if the communication process is so corrupted that \( S \) and \( R \) lose ex-ante, then the third party can use the garbling to weaken the signal strength to an arbitrary degree, effectively foreclosing the communication channel altogether.

When \( S \) and \( R \) are unaware of the garbling, the third party can take advantage of their earnestness to shift the message into better alignment with the anchor. The optimal mean of \( \varepsilon \) is

\[
\mu^*_\varepsilon = -\left(1 - \frac{\mathbb{E}[b]}{\mathbb{E}[b^2]} \right) \beta_S (\mu_m - a),
\]

which becomes more extreme in the degree of heterogeneity, the level of the sender’s coefficient \( \beta_S \), and the misalignment \( \mu_m - a \). Because \( S \) and \( R \) do not respond to it, variance in \( \varepsilon \) is entirely counterproductive. The garbler would thus like the variance \( \sigma^2_\varepsilon \) to be as low as possible. However, even if some positive level of variance is unavoidable, welfare can still be improved through realignment.

4.3. Third-Party Mediation

Suppose that a mediating third party is placed between the sender and receiver. Instead of sending directly to the receiver, the sender instead passes \( m_S = \mu_{m_S} + \beta_S (\theta - \mu_\theta) \) to the mediator, who in turn passes \( m_M = \mu_{m_M} + \beta_M (m_S - \mu_{m_S}) \) to the receiver. All three parties’ conventions are drawn i.i.d. The receiver takes action \( x = \mu_x + \beta_R (m_M - \mu_{m_M}) \). Absent heterogeneity, \( R \)’s action is thus

\[
x = \mu_x + (\beta_R \beta_M \beta_S)(\theta - \mu_\theta).
\]

We assume the mediator first announces his strategy, after which the sender and receiver adjust accordingly. Hence, equilibrium dictates \( \mu_x = \mu_\omega \) and \( \beta_R \beta_M \beta_S = \sigma_{\omega \theta} / \sigma^2_\theta \). The relevant question is, can the mediator help the sender and receiver, despite introducing another source of convention heterogeneity?
Fact 6. The mediator can eliminate bias by setting message $\mu_{mM} = a$. However, if the sender’s message is already aligned with the anchor, $\mu_{mS} = a$, then the presence of the mediator strictly decreases welfare. Proof in the appendix.

As the above fact states, the mediator can help the sender and receiver, but only insomuch as he is able to realign the message with the anchor. Part of the problem is that the sender and receiver adjust to whatever strategy the mediator chooses. To further illustrate this point, suppose that the sender’s message is already aligned with the anchor, but that $S$ and $R$ wrongly believe that the mediator is a neutral messenger who simply sets $m_M = m_S$. The mediator can exploit this unawareness by setting $\mu_{mM} = \mu_{mS} = a$ and setting $\beta_M$ to minimize the resulting expression for error:

$$E[(x - \omega)^2] = \sigma^2_\omega(1 - \rho^2_{\omega\theta}) + \sigma^2_\omega \rho^2_{\omega\theta} E[(\beta_M b - 1)^2].$$

The optimal choice is $\beta_M = E[b]/E[b^2]$, precisely the conservatism factor that appeared in the awareness equilibrium. In this way, a mediator can further assist the sender and receiver when they do not best-respond to his actual strategy.

4.4. Redundancy

The final approach to improving communication supposes that, instead of a single sender with a single message, information is passed to $R$ via two separate senders whose information can be independent or arbitrarily closely correlated. To facilitate a direct comparison between the two cases, assume the sender(s) report their signal directly and that the receiver does all of the appropriate discounting. Furthermore, we concentrate on a particular family of information structures that produce identical outcomes in the absence of convention heterogeneity. For the one-sender case, suppose the sender’s signal takes the form $\theta = \omega + \varphi$, where $\varphi$ is mean-zero Gaussian noise with variance $\sigma^2_\varphi$. It follows

$$x = \mu_\omega + \frac{\sigma^2_\omega}{\sigma^2_\omega + \sigma^2_\varphi} (\theta - \mu_\omega),$$

leading to zero bias and error

$$E[(x - \omega)^2] = \sigma^2_\omega \left(1 - \frac{\sigma^2_\omega}{\sigma^2_\omega + \sigma^2_\varphi}\right).$$
For the two-sender case, suppose each of senders \( i = 1, 2 \) observes and transmits \( \theta_i = \omega + \psi + \varepsilon_i \) to \( R \), where \( \psi \) is a common mean-zero Gaussian noise term with variance \( \sigma^2_{\psi} \), and the \( \varepsilon_i \) are i.i.d. mean-zero Gaussian with variance \( \sigma^2_{\varepsilon} \). Now, \( R \)'s optimal action is

\[
x = \mu_{\omega} + \sum_{i=1,2} \frac{\sigma^2_{\omega}}{2(\sigma^2_{\omega} + \sigma^2_{\psi} + \sigma^2_{\varepsilon})}(\theta_i - \mu_{\omega}),
\]

leading to zero bias and error

\[
E[(x - \omega)^2] = \sigma^2_{\omega} \left(1 - \frac{\sigma^2_{\omega}}{\sigma^2_{\omega} + \sigma^2_{\psi} + \sigma^2_{\varepsilon}/2}\right).
\]

These expressions show that the receiver and sender are indifferent between the one- and two-sender cases so long as that \( \sigma^2_{\psi} + \sigma^2_{\varepsilon}/2 = \sigma^2_{\varphi} \).

Fact 7. With communication heterogeneity, bias is the same in the one- and two-sender cases. Redundancy decreases error in the two-sender case, and, for sufficiently high redundancy, the two-sender case has strictly less error than the one-sender case. Proof in the appendix.

Bias is equal in the one- and two-sender cases, intuitively, as the receiver’s regression coefficient on each of the two senders is half of what he puts on the single sender. Error in the two-sender case deviates from the baseline in two ways. First, the \( E[(b - 1)^2] \) term in the error expression (as in Fact 2) is replaced by

\[
E[(b - 1)^2] - \frac{1}{2}(V[b] - C[b_1, b_2]) < E[(b - 1)^2],
\]

reflecting the fact that the exuberance ratios of the two different senders, though not independent, are effectively averaged in the error computation, and their average has lower variance than one individually. On the other hand, the second discrepancy is that there is an extra error term associated with the individual information component \( \varepsilon \). Hence, redundancy lowers error. In the extreme, when the information is completely redundant, this term disappears, making the two-sender error unambiguously lower than its one-sender counterpart.
5. Persuasion and Social Learning

This section analyzes two extensions. First, it revises the model in Section 3 to incorporate conflicting interests between the sender and receiver. Second, it considers the consequences of unacknowledged heterogeneity for social learning.

5.1. Persuasion

As before, sender $S$ sends message $m$ to receiver $R$, and $R$ takes action $x \in \mathbb{R}$ given utility function $U_R(x|\omega) = -(x - \omega)^2$. We now assume that the there is a prescribed message format – the sender is supposed to send her conditional mean belief about the state – but that the sender’s goals are divided between honestly reporting that message and persuading the receiver away from the true state. In addition to signal $\theta$, the sender also privately observes a persuasion motive $\xi$, which is also Gaussian, and, for simplicity, assumed to be independent of both the state and the signal. Honesty concerns aside, the sender wants $R$’s action to match the offset value $\omega + \xi$ instead of $\omega$. The relative weight placed on honesty relative to shifting $R$’s action is captured by parameter $\eta > 0$.

Aggregating these concerns, the sender’s ex-post payoff is

$$U_S(x|\omega, \xi, m) = -(x - (\omega + \xi))^2 - \eta(m - \mu_\omega(\theta))^2.$$

The specific functional forms specified in the game above are used for two reasons: they integrate cleanly with the affine special case of communication functions, and they facilitate the calculation of comparative statics. The only innovation in the framework used here relative to standard cheap-talk models is the Gaussian distribution of fundamental noise terms. A preference for honest communication has been introduced in other cheap-talk models, e.g. Kartik (2009); Khalmetski (2019), and, as shown below, the present model without convention heterogeneity replicates the qualitative features of those models.\(^{16}\)

As before, the equilibrium concept is perfect Bayesian equilibrium, and we restrict attention to equilibrium in affine strategies. Unlike in the pure coordination case, equilibrium with the added features of persuasion is unique. To see how this comes about, first consider the sender’s best

\(^{16}\)A key consequence of partial honesty is language ‘inflation’, whereby the sender’s message is shifted in the direction of her preferred state; I document this effect as well in the setting without convention heterogeneity.
response to any receiver strategy. Supposing $R$ adopts strategy $x = \mu + \beta_R(m - \mu_m)$, the sender’s best response is a convex combination of two competing goal messages:

$$m = \left(\frac{\eta}{\eta + \beta_R^2}\right) \cdot \mu(\theta) + \left(1 - \frac{\eta}{\eta + \beta_R^2}\right) \cdot \left(\frac{\mu_m + \mu(\theta) - \mu + \xi}{\beta_R} \right) \cdot x^{-1}(\mu(\theta) + \xi).$$  \hfill (1)

Honesty pushes $S$’s message closer to the prescribed message $\mu(\theta)$. Her persuasion motive pushes $m$ closer to $x^{-1}(\mu(\theta) + \xi)$, the message that, given $R$’s strategy, would induce him to take her preferred action $\mu(\theta) + \xi$. Equilibrium obtains when the receiver’s strategy coincides with Bayes’ rule: $\mu_m = \mathbb{E}[m]$ and $\beta_R = \mathbb{C}[\omega, m]/\mathbb{V}[m]$.

**Proposition 3.** There is a unique equilibrium, consisting of strategies $m$ as in Equation (1) and $x = \mu + \beta_R(m - \mu_m)$, where $\mu_m = \mu + (\beta_R/\eta) \cdot \mu_\xi$ and $\beta_R$ is the unique positive solution to

$$\beta_R = \frac{(\eta + \beta_R^2)(\eta + \beta_R)}{(\eta + \beta_R)^2 + \beta_R^2[\sigma_\xi^2/\sigma_\theta^2]}.$$

The equilibrium satisfies the following properties:

1. The receiver strictly discounts $S$’s message: $\beta_R < 1$.
2. As honesty preference $\eta$ increases from 0 to $\infty$, $\beta_R$ increases strictly from 0 to 1.
3. Coefficient $\beta_R$ is independent of $\mu_\xi$; it decreases strictly in $\sigma_\xi^2$, with $\beta_R \rightarrow 1$ as $\sigma_\xi^2 \rightarrow 0$.
4. The receiver’s action has zero bias; error is strictly larger than in the coordination case, and it decreases strictly in $\eta$.

*Proof in the appendix.*

The properties enumerated in Proposition 3 illustrate several features of the persuasion game in the absence of convention heterogeneity. First is language inflation, as has been documented in the literature on cheap-talk with (partially) honest senders. The average message in equilibrium is shifted in the direction of the sender’s average persuasion motive, although the receiver understands this, and properly demeans it so that his action has no bias. The extent of that shift, as well as all other outcomes of the game, depend on the equilibrium value of $R$’s regression coefficient $\beta_R$. As with prior cheap-talk models, the greater the gap between the sender’s and receiver’s preferences, here modeled as greater $\sigma_\xi^2$ and/or lower $\eta$, the lower is $\beta_R$. That is, $R$ further discounts $S$’s message; hence less information is transmitted in equilibrium. On the other hand, as $R$ further
discounts S’s message, S loses influence over R’s action, diminishing her incentive to lie and keeping the equilibrium parameters interior.

In the presence of convention heterogeneity, R takes action \( x = \mu \omega + \beta_R b(m - \mu_m) + \beta_R(b - 1)(\mu_m - a) \), which accordingly influences error and bias.

**Fact 8.** In the equilibrium of the persuasion game with convention heterogeneity, bias is

\[
\mathbb{E}[x(\hat{m}) - \omega] = \beta_R \cdot \mathbb{E}[b - 1] \cdot (\mu_\omega - a) + (\beta_R^2/\eta) \cdot \mathbb{E}[b - 1] \cdot \mu_\xi,
\]

and error is

\[
\mathbb{E}[(x(\hat{m}) - \omega)^2] = \sigma_\omega^2 \cdot (1 - \rho_{\omega m}^2) + \sigma_\omega^2 \cdot \rho_{\omega m}^2 \cdot \mathbb{E}[(b - 1)^2] \cdot \left[ 1 + \left( \frac{\mu_m - a}{\sigma_m} \right)^2 \right].
\]

In the expression for bias, the first term is analogous to the coordination case. Because equilibrium dampens the absolute value of R’s opinion shift due to signal \( \theta \), this component of bias is similarly dampened. On the other hand, the second term shows how the persuasion motive \( \xi \) contributes to additional bias in the sender’s preferred direction. The expression for error shows that, analogous to coordination case, there is a pure variance effect and a shift effect that depends on the magnitude of the gap \( (\mu_m - a) \). Furthermore, error increases unambiguously with miscommunication; thus convention heterogeneity never helps R.

While R’s expected utility is described by the error, the sender’s payoff incorporates additional factors. The sender’s equilibrium utility decreases in error, decreases in the variance of the persuasion motive, but increases in the correlation between the persuasion motive and the error.

**Fact 9.** The sender loses less from miscommunication than the receiver does if either \( |\mu_\xi| \) is sufficiently large or if \( \mu_\xi \cdot (\mu_\omega - \mathbb{E}[a]) \geq 0 \). The sender gains from miscommunication when \( \beta_R \) is sufficiently small and \( |\mu_\xi| \) is sufficiently large. **Proof in the appendix.**

Intuitively, the miscommunication effects S and R asymmetrically for the following reason. Because R understands the sender’s incentives, he accounts for the distribution of her persuasion motive; absent miscommunication, his action is unbiased, and miscommunication only thwarts his efforts to guard against S’s persuasion tactics. Although she has an honesty preference, the sender would be happy ex-post if nature intervened to push R’s belief in the direction of \( \mu_\omega(\theta) + \xi \). The exaggeration effect from miscommunication can achieve this. In the case that \( \mu_\xi \cdot (\mu_\omega - \mathbb{E}[a]) \geq 0 \), then, on average,
the offset term $\xi$ pushes $S$ away from the state $\omega$ in the same direction as the state diverges from the linguistic anchor $a$. Hence, for however much $R$ suffers from taking a biased action, that loss for the sender is ameliorated by the fact that $R$’s bias is towards $S$’s ideal action. In the second scenario, $|\mu_\xi|$ is very large. $S$ would like to move $R$’s belief substantially, but, without convention heterogeneity, $R$ knows to appropriately discount $S$’s $\mu_\xi$-biased messages. Hence, the greater is $|\mu_\xi|$, the greater the potential for exaggeration to help $S$. As Fact 9 verifies, nature’s assistance in shifting $R$’s action can be so great that $S$ benefits on net.

5.2. Sequential Communication

This section considers how does miscommunication affect the aggregation of different pieces of information by examining the sequential communication of agents with conditionally independent signals.

Suppose there is a sequence of agents at times $t = 1, 2, \ldots$. Each agent $t$ has private signal $\theta_t = \omega + \varepsilon_t$, where all of the $\varepsilon$ terms are i.i.d. Gaussian with mean zero and variance $\sigma_\varepsilon^2$. Furthermore, agent $t$ hears the opinion $\mu_{t-1}$ of her immediate predecessor. To simplify notation, this subsection uses precision $\tau_X \equiv 1/\sigma_X^2$ in place of variance. Denote $\tau_t = \tau_\omega + t \cdot \tau_\varepsilon$ to be the subjective precision of the agent $t$. Then, the agents’ beliefs over time evolve as

$$
\mu_t = \frac{\tau_\varepsilon}{\tau_t} \theta_t + \frac{\tau_{t-1}}{\tau_t} \mu_{t-1}.
$$

With convention heterogeneity, the sequence of beliefs evolves as

$$
\mu_t = a + \frac{\tau_\varepsilon}{\tau_t} \sum_{s=1}^t b_{st}(\theta_s - a) + \frac{\tau_\omega}{\tau_t} b_{1t}(\mu_\omega - a).
$$

From the perspective of an arbitrary fixed agent, $N$, the sequence of reported beliefs would appear to be

$$
\hat{\mu}_t^N \equiv a + b_{tN}(\mu_t - a) = a + \frac{\tau_\varepsilon}{\tau_t} \sum_{s=1}^t b_{sN}(\theta_s - a) + \frac{\tau_\omega}{\tau_t} b_{1N}(\mu_\omega - a).
$$

The second term in both expressions above captures the fact that the common prior is incorporated directly into the first agent’s belief and subsequently passed along from agent to agent. Furthermore, note that because of the way exuberance ratios telescope when composed (e.g. $b_{rs}b_{st} = b_{rt}$), the exact linear structure of the information aggregation does not play any meaningful role. We could
just as well consider agent $t$’s belief as aggregating from a tree structure, and the evolution of beliefs would still match the expression above.

**Fact 10.** From the perspective of observer agent $N$, the sequence of beliefs converges:

$$
\lim_{t \to \infty} E[\hat{\mu}^N_t | \omega] = a + E[b_tN](\omega - a), \quad \text{and} \quad \lim_{t \to \infty} V[\hat{\mu}^N_t | \omega] = 0.
$$

However, the actual belief sequence follows

$$
\lim_{t \to \infty} E[\hat{\mu}_t | \omega] = a + E[b_t](\omega - a), \quad \text{and} \quad \lim_{t \to \infty} V[\mu_t | \omega] = \omega^2 \cdot C[b_1, b_2] > 0.
$$

Limit error is $\lim_{t \to \infty} E[(\mu_t - \omega)^2] = (E[(b - 1)^2] + C[b_1, b_2]) (\sigma^2_\omega + (\mu_\omega - a)^2)$. *Proof in the appendix.*

As this fact shows, from the perspective of any fixed agent, the sequence of agents’ beliefs converges. What happens is that the sequence of *words* converges in the sense described in Section B.3. However, the implied limit message depends on the convention of the observer agent $N$. In the tail of the sequence, agents place nearly complete stock in the opinion of their predecessor, hence each agent’s belief is approximately her (subjective) interpretation of the limit word. While the sequence of actual beliefs converges *in mean* (to a strictly exaggerated value), the limit variance of beliefs remains strictly positive, reflecting the perpetual disagreement among agents about the meaning of the limit word.

From the ex-ante perspective, bias converges as $E[\mu_t - \omega] \to E[b - 1](\mu_\omega - a)$, in line with Fact 2. The expression for limit error in Fact 10 shows that the value of the information aggregation, as captured by the asymptotic error, improves on no information transmission only for sufficiently low heterogeneity. Furthermore, the aggregation of many agents’ conventions introduces a new source of error in the process. One corollary is that if heterogeneity is so great as to make two-player communication not worthwhile, that is if $E[(b - 1)^2] > 1$, then there is no way the sequence of agents can benefit from communication.

6. Discussion

This paper has identified a novel friction in communication games. That friction, the exaggeration effect, results from augmenting a sender-receiver framework with interpersonal projection, whereby an agent assumes overlooks key differences between others and him/herself. In contrast to
both classical and behavioral cheap-talk models, the *mean* of the receiver’s belief is shifted by biased reasoning. Beyond analyzing the exaggeration effect and studying how to mitigate its impact, the paper further expands the behavioral literature on projection biases. It introduces a taxonomy of miscommunication possibilities applicable beyond the particular friction studied. Finally, it employs a cheap-talk model with a Gaussian-Gaussian information structure and affine strategies, providing an additional and transparent way to understand existing results in the literature. In what follows, I further discuss connections with the literature and prospects for future work.

**Literature Connections.** This paper sits at the intersection of game theoretic models of cheap-talk and behavioral models of projection bias; it builds on and makes contributions to both. Most important is the psychological phenomenon of projection bias, the proclivity of humans to over-extrapolate from immediate personal circumstances. Loewenstein, O’Donoghue and Rabin (2003) and Van Boven and Loewenstein (2003) examine people substituting their current preferences for future preferences; Gagnon-Bartsch (2016) models inter-personal taste projection; Madarasz (2015) models information projection, by which one neglects how private information is not shared by others. Interpersonal projection is also found in the ‘false consensus effect’ (Ross, Greene and House, 1977), and debated in which subjects by which people *over*-extrapolate from their personal experiences when dealing with others. More broadly, advances in behavioral game theory (e.g. Nagel (1995); Stahl and Wilson (1995); Camerer, Ho and Chong (2004); Eyster and Rabin (2005, 2010)) have demonstrated that strategic naiveté can produce outcomes that are not merely noisy versions of their rational counterparts, but systematically distinct from existing predictions.

Much recent theoretical and experimental work has studied honesty preferences and guilt aversion (Battigalli and Dufwenberg, 2007). A number of lab experiments, such as Gneezy (2005); Charness and Dufwenberg (2011); Gibson, Tanner and Wagner (2013); Gneezy, Kajackaite and Sobel (2018), document subjects being unnecessarily truthful according to material payoffs. Likewise, studies such as Jin, Luca and Martin (2019) find that receivers are unnecessarily trusting of senders. Abeler, Becker and Falk (2014) document the same in a survey format, and Abeler, Nosenzo and Raymond (2019) provides a large meta-study analysis of various methodologies, all presenting a clear picture that people have a preference to stick to their understanding of truthful reporting. As

---

17The degree to which extrapolation from own experience reflects a bias versus a rational calculation is debated, most famously in (Dawes, 1989). Also see Engelmann and Strobel (2000). There is lab evidence for explicitly *undue* extrapolation (Krueger and Clement, 1994), but caution is warranted when assessing relevant data.
Sobel (2020) observes, one cannot have a preference for honesty or a desire to lie to one’s opponent without first understanding words to have some ‘objective’ meaning. Reasoning that one’s interlocutor might have a different conception of the meaning of words is thus fundamentally at odds with the concept honesty.

What distinguishes the approach in this paper from the most cheap-talk literature is that linguistic conventions are a primitive and not an outcome of the model. Several other studies have also done as much, but my approach is conceptually distinct. Blume and Board (2013) introduce a model of language ‘competence,’ in which one component of players’ private information in a Bayesian game defines their capacity for sending and distinguishing messages. (Also see Hagenbach and Koessler (2017) and Giovannoni and Xiong (2019) for applications of this approach.) Along similar lines, Sobel (2012) models a sender and receiver investing in a less explicit form of language ‘capacity,’ and shows it can generate the same kind of features as the traditional cheap-talk paradigm. In these papers, language differences function akin to a technological barriers between the parties, making communication more difficult. This is in contrast to heterogeneous conventions, which facilitate a kind of mistaken over-communication. In both Glazer and Rubinstein (2006) and Eliaz, Spiegler and Thysen (2020), the meaning of sender’s message is subject to a contextual interpretation that is not simply determined by equilibrium frequencies, but these papers treat receivers as willing to entertain multiple interpretations.

By failing to recognize convention heterogeneity, the agents in this paper’s model thus exhibit bounded rationality, but in a way distinct from prior literature on communication with level-K players (Crawford, 2003), credulous receivers (Chen, 2011; Kartik, Ottaviani and Squintani, 2007), ambiguity averse receivers Kellner and Le Quement (2018), receivers with coarse or analogy-based reasoning (Mullainathan, Schwartzstein and Shleifer, 2008; Ettinger and Jehiel, 2010; Jehiel, 2019; Hagenbach and Koessler, 2020), among other examples of behavioral communicators. A consistent theme from these papers is that bounded rationality breaks down the strategic forces that would otherwise prevent information transmission. A sophisticated sender facing a behavioral receiver benefits from an increased ability to persuade, but the receiver can also benefit as a result of the increased information. I also find biased agents being overly receptive to information, but it is because the rational response convention heterogeneity involves a team effort by sender and receiver to incorporate greater conservatism. Unacknowledged heterogeneity is akin to deception.
by ‘nature’ who wishes to exaggerate the receiver’s action. The receiver, wishing only to match
the state, is always in conflict with exaggeration and thus unambiguously loses; the sender only
benefits when her persuasion motive happens to align with the exaggeration effect. Moreover,
the quality of information passed, as measured by the mean-squared error in the receiver’s action,
always degrades as a result of the bias.

In contrast to the existing papers which evaluate the efficacy of mitigation protocols between
senders and receivers, such as Blume, Board and Kawamura (2007), Goltsman et al. (2009), and
Blume (2012) I find that added noise cannot improve welfare without managing to shift the mean of
agents’ beliefs. This stems partly from the focus in the present paper on coordination settings, but
it is also due to the fact that the sender and receiver are suffering from unwanted (and unacknowl-
edged) randomness between them. Another strategy for more effective communication is the notion
of language efficiency studied Crémer, Garicano and Prat (2007), which notes in the context of tech-
nical communication that it is worth having ‘precise words for frequent events and vague words for
unusual ones.’ Lim and Wu (2018) show that vague communication can be sender-optimal.

Future Directions. There are several avenues for extending the work in this paper. The
interaction between sender and receiver considered here is single-shot, but it would be useful to
understand how dynamic feedback could serve either to correct the mis-coordination or, were the
source of the conflict mis-identified, exacerbate it. Ultimately the exaggeration effect is a medium-
run phenomenon that would likely be corrected by a sufficient combination of time, feedback, and
stakes, but despite this, a dynamic extension could speak to the rate at which the initial gap in
conventions would be reduced. Along similar lines, this paper does not consider the original source
of convention heterogeneity. Implicit is that there are small differences in environment in a large
population, and that people adapt or are taught to use the local convention. Chen and Mitchell
(2018) consider a model of the evolution of language features along such lines, but does not model
the semantic mapping of a convention precisely. Perhaps most promising would be to consider
unacknowledged heterogeneity in game strategies more broadly. Just as failure to coordinate on
convention usage effectively pairs one half of one equilibrium with one half of another, so could the
same exercise be repeated in arbitrary two-player games.
References


Appendix A. Proofs

Proof of Proposition 1. The Proposition follows from four statements. For the first three, we concentrate on a single pair of distinct conventions $S \neq R$; we consider communication going in the ‘forwards direction’ from $S$ to $R$, taking $m$ to $h_{SR}(m)$, versus going in the ‘reverse’ direction from $R$ to $S$, taking $m$ to $h_{RS}(m)$.

(a) If $S$ is more exuberant than $R$ at message $m$, then the magnitude of miscommunication is greater in the forwards direction than in the reverse direction: $|h_{SR}(m) - m| > |h_{RS}(m) - m|$.

Consider the words $w_S(m)$ and $w_R(m)$ that the two parties use to describe $m$. With forwards communication, the magnitude of miscommunication is equal to what the receiver understands as the gap in meaning between the two words:

$$|h_{SR}(m) - m| = |m_R(w_S(m)) - m_R(w_R(m))|.$$  

Similarly, the magnitude of miscommunication in the reverse direction is equal to the sender’s perception of that gap, $|m_S(w_S(m)) - m_S(w_R(m))|$. By definition, the receiver’s gap is larger than the sender’s gap if and only if the sender is more exuberant than the receiver.

(b) If conventions are monotonic and $S$ is more exuberant than $R$ on $\mathcal{M}^*$ containing anchor $a$ and message $m$, then forwards communication stretches $m$ away from $a$ while reverse communication compresses $m$ towards $a$:

$$\begin{cases} a < m \implies a < h_{RS}(m) < m < h_{SR}(m) \\ m < a \implies h_{SR}(m) < m < h_{RS}(m) < a. \end{cases}$$

By definition, the difference between the actual message and the anchor is equal to what the sender understands as the gap between the words she herself would use to describe those values: $|m - a| = |m_S(w_S(m)) - m_S(w_S(a))|$. In the forwards direction, the difference between the miscommunicated message and the anchor is equal to what the receiver understands as the gap between those two words: $|h_{SR}(m) - a| = |m_R(w_S(m)) - m_R(w_S(a))|$. Hence, $S$ being more exuberant than $R$ on $\mathcal{M}^*$ means that the receiver’s perceived gap is larger, so $|h_{SR}(m) - a| > |m - a|$. Monotonicity ensures that the miscommunicated message is ‘on the same side of $a$,’ $h_{SR}(m) > a$ if and only if $m > a$, so it follows that the effect of forwards communication is to stretch $m$ away...
from $a$: either $h_{SR}(m) < m < a$ or $a < m < h_{SR}(m)$. An analogous argument establishes that reverse communication compresses $m$ towards $a$.

(c) If conventions are monotonic and exuberance-consistent on interval $\mathcal{M}^*$ containing anchor $a$, then $(1/2)(h_{SR} + h_{RS})$ is an exaggeration relative to $a$ on $\mathcal{M}^*$.

Suppose $m \in \mathcal{M}^*$ and, without loss of generality, that $m > a$. We can express

$$\frac{1}{2}(h_{SR}(m) + h_{RS}(m)) = m + \frac{1}{2}\left(\frac{(h_{SR} - m) - (m - h_{RS}(m))}{\Delta_{SR}}\right).$$

Statement (a) implies $|\Delta_{SR}| > |\Delta_{RS}|$, and statement (b) implies $\Delta_{SR}, \Delta_{RS} > 0$, so $\Delta_{SR} - \Delta_{RS} > 0$. It follows that $(1/2)(h_{SR}(m) + h_{RS}(m)) > m > a$, an exaggeration.

(d) If average communication conditional on each pair of distinct conventions is an exaggeration relative to $a$ on $\mathcal{M}^*$, then average communication is an exaggeration relative to $a$ on $\mathcal{M}^*$.

As the sender’s and receiver’s conventions are drawn i.i.d., the joint distribution over conventions is unchanged if we swap each sender-receiver pair, hence $E[h(m)] = E[h_{RS}(m)]$. Suppose $m \in \mathcal{M}^*$ and, without loss of generality, that $m > a$. Then, by (c),

$$E[h(m)] = \frac{1}{2}(E[h(m)] + E[h_{RS}(m)]) = E\left[\frac{1}{2}(h_{SR}(m) + h_{RS}(m))\right] > E[m] = m,$$

an exaggeration.

Proof of Proposition 2.

(a) The proof follows that of Proposition 1 with one additional observation: a value $m > \overline{a}$ is exaggerated relative to one pairwise anchor in the range $[a, \overline{a}]$ if and only if it is exaggerated relative to all of them.

(b) Fix a pair of distinct conventions $S$ and $R$, and suppose $S$ is both higher and more exuberant than $R$. Since $S$ is higher, $h_{SR}(m) > m$ for all $m$; hence, following the notation above, $m_+ > m > m_-$; As before, setting $w_2 = w_S(m_-)$ and $w_1 = w_S(m)$ in the exuberance condition $m_R(w_2) - m_R(w_1) > m_S(w_2) - m_S(w_1)$ yields the result $(1/2)(m_+ + m_-) > m$. The other half of the proof follows accordingly.

(c) Immediate from Proposition 1.

(d) Suppose for the sake of contradiction that no values $\tilde{v} \in [a, m]$ were exaggerated, where $m > a$ is arbitrary. Choose any $m_0$ in the isolation neighborhood of $a$ such that $a < v_0 < m$. Denote
\( m_k = (h_{SR})^{-k}(m_0) \), and suppose without loss of generality that \( h_{SR}(m_0) > v_0 \). This implies \( m_0 > v_1 > v_2 > \cdots > a \), although the sequence need not converge to \( a \). In order that there be no exaggeration at \( m_k \), the proof of Proposition 1 shows that

\[
v_k - v_{k+1} \geq v_{k-1} - v_k
\]

must hold. On the other hand, a bounded monotonic sequence cannot have increasing differences everywhere. Hence this chain of inequalities must break down at some point, identifying \( m_k \) in \((a,v)\) at which the inequality goes the other way, and so average communication is an exaggeration. An equivalent argument applies for \( m < a \). \( \diamond \)

**Proof of Fact 3.** The six statistics covered by the fact are, by inspection, equal to 0 when \( b \) is identically equal to 1. To show that the first three statistics are always increasing with heterogeneity, note that by the symmetry of the distribution over \( b \), it follows \( \mathbb{E}[f(b)] = (1/2)\mathbb{E}[f(b) + f(1/b)] \) for any measurable function \( f \). Leveraging this identity allows us to express

\[
\mathbb{E}[b - 1] = \frac{1}{2} \mathbb{E}[b - 1] + \frac{1}{2} \mathbb{E}\left[ \frac{1}{b} - 1 \right] = \mathbb{E}\left[ \frac{(b - 1)^2}{2b} \right],
\]

and as \((b - 1)^2/(2b)\) is symmetric around \( b = 1 \), it follows

\[
\mathbb{E}[b - 1] = \mathbb{P}[b = 1] + 2\mathbb{E}\left[ \frac{(b - 1)^2}{2b} \mid b > 1 \right].
\]

Furthermore, whenever \( F_2 \) is more heterogeneous than \( F_1 \), the distribution of \( F_2 \) restricted to \([1, \infty)\) first-order stochastic dominates the corresponding \( F_1 \) restriction. As \((b - 1)^2/(2b)\) is strictly increasing on \( b > 1 \), it follows \( \mathbb{E}[b - 1] \) is strictly greater under \( F_2 \) than under \( F_1 \). By similar rearranging we can also express

\[
\begin{cases}
\mathbb{E}[(b - 1)^2] = \mathbb{E}\left[ \frac{(b^2 + 1)(b - 1)^2}{2b^2} \right] \\
\mathbb{E}[b^2 - b] = \mathbb{E}\left[ \frac{(b^2 + b + 1)(b - 1)^2}{2b} \right],
\end{cases}
\]

showing they are increasing with heterogeneity as well. As for the other three statistics, note that for the two-convention case,

\[
\forall[b] = \frac{(B - 1)^2(3B^2 + 2B + 3)}{16B^2}, \quad \frac{\mathbb{E}[b^2]}{\mathbb{E}[b]} - 1 = \frac{(B - 1)^2(B^2 + B + 1)}{B(1 + B)}, \quad \text{and}
\]

38
\[ C[b_1, b_2] = \frac{(1 + B)^2(B^8 + 4B^6 - 4B^5 - 2B^4 - 4B^3 + 4B^2 + 1)}{64B^5}, \]
all of which are strictly increasing in \( B \). In the log-normal case,
\[ \forall [b] = (\exp{\sigma _b^2} - 1) \exp{\sigma _b^2}, \quad \frac{\mathbb{E}[b^2]}{\mathbb{E}[b]} - 1 = \exp{(3/2)\sigma _b^2} - 1, \quad \text{and} \quad C[b_1, b_2] = \exp{\sigma + \sigma /2} - \exp{\sigma ^2}, \]
all of which are strictly increasing in \( \sigma _b^2 \).

**Proof of Fact 4.** Denoting the receiver’s strategy \( x(\hat{m}) = \mu _\omega + \beta _R(\hat{m} - \mu _{\hat{m}}) \), the sender chooses \( m \) to maximize
\[ -\mathbb{E} \left[ (\mu _\omega + \beta _R(a + b(m - a) - \mu _{\hat{m}}) - \omega )^2 \right], \]
reflecting the completely classical approach in this case. Solving the first order condition yields
\[ m^* = \frac{\mathbb{E}[b](\mathbb{E}[\omega | \theta] - \mu _\omega) + \beta _R(\mathbb{E}[b^2 - b]a + \mathbb{E}[b]\mu _{\hat{m}})}{\beta _R\mathbb{E}[b^2]}, \]

In equilibrium,
\[ \mu _{\hat{m}} = \mathbb{E}[\hat{m}] = a + \mathbb{E}[b](\mathbb{E}[m] - a) \]
\[ = a + \mathbb{E}[b] \left( \frac{\mathbb{E}[b^2 - b]a + \mathbb{E}[b]\mu _{\hat{m}} - \mathbb{E}[a]}{\mathbb{E}[b^2]} \right) \]
\[ = \cdots \]
\[ \mu _{\hat{m}} = a, \]
from which it follows
\[ m^* = a + \frac{\mathbb{E}[b]}{\beta _R\mathbb{E}[b^2]} \frac{\sigma _\omega \theta}{\sigma _\theta ^2} (\theta - \mu _\theta). \]

From the receiver’s perspective,
\[ C[\hat{m}, \omega] = \mathbb{E}[b] \left( \frac{\beta _R\mathbb{E}[b]}{\beta _R^2\mathbb{E}[b^2]} \right) \cdot \frac{\sigma _\omega \theta}{\sigma _\theta ^2} \quad \text{and} \quad \forall [\hat{m}] = \mathbb{E}[b^2] \left( \frac{\beta _R\mathbb{E}[b]}{\beta _R^2\mathbb{E}[b^2]} \right)^2 \cdot \frac{\sigma _\omega \theta}{\sigma _\theta ^2}. \]

In equilibrium, \( \beta _R = C[\hat{m}, \omega]/\forall [\hat{m}] \), yielding the key identity \( \beta _S\beta _R = \frac{\mathbb{E}[b]}{\mathbb{E}[b^2]} \cdot \frac{\sigma _\omega \theta}{\sigma _\theta ^2} \). The expressions for bias and error follow mechanically from the setup. \( \diamond \)
Proof of Fact 5. When sender and receiver are aware of the garbling and respond to it, S’s message is of the form \( \mu_m + \beta S(\theta - \mu \theta + \varepsilon - \mu \varepsilon) \), and equilibrium requires

\[
\beta R \beta S = \frac{\sigma_{\omega \theta}}{\sigma_{\omega (\theta + \varepsilon) \varepsilon}} = \frac{\sigma_{\omega \theta}}{\sigma_{\theta \theta}^2 + \sigma_{\varepsilon \varepsilon}^2}.
\]

Then,

\[
x - \omega = -(\omega - \mu \omega) + \beta R \beta S b(\theta - \mu \theta + \varepsilon - \mu \varepsilon) + \beta R (b - 1)(\mu_m - a),
\]

so

\[
E[x - \omega] = \beta_R E[b - 1](\mu_m - a)
\]

and

\[
E[(x - \omega)^2] = \sigma_{\omega \theta}^2 (1 - \rho_{\omega(\theta + \varepsilon)}^2) + \sigma_{\omega \varepsilon}^2 \rho_{\omega (\theta + \varepsilon)}^2 E[(b - 1)^2] \left( 1 + \left( \frac{\mu_m - a}{\beta S \sigma_{\omega \varepsilon}} \right)^2 \right),
\]

demonstrating that the additional garbling does no more than weaken the information quality of the S’s signal. In the unaware case, \( \beta R \beta S = \sigma_{\omega \theta} \sigma_{\theta \theta}^2 \), without regard to the additional noise, hence

\[
x - \omega = -(\omega - \mu \omega) + \beta R \beta S b(\theta - \mu \theta + \varepsilon) + \beta R (b - 1)(\mu_m - a).
\]

Then

\[
E[x - \omega] = \beta_R \cdot E[b - 1] \cdot (\mu_m - a) + \frac{\sigma_{\omega \theta}}{\sigma_{\theta \theta}^2} E[b] \mu \varepsilon
\]

and

\[
E[(x - \omega)^2] = E + E[b^2] \beta R^2 \beta S^2 (\mu \varepsilon^2 + \mu \theta^2 + 2 \sigma \theta \varepsilon) + 2 E[b(b - 1)](\mu_m - a) \beta S \beta R \mu \varepsilon^2.
\]

Solving the first order condition shows the optimal mean is

\[
\mu \varepsilon = -\left( 1 - \frac{E[b]}{E[b^2]} \right) \beta S (\mu_m - a)
\]

and that this improves welfare if convention heterogeneity, message misalignment with anchor, and variance of noise term \( \varepsilon \) are all not prohibitively large:

\[
\left( 1 - \frac{E[b]}{E[b^2]} \right)^2 (\mu_m - a)^2 (2 - \beta S^2) - \sigma \varepsilon^2 > 0
\]

Proof of Fact 6. When we introduce heterogeneity,

\[
\hat{m}_S = m_S + b_{SM}(m_S - a) \quad \text{and} \quad \hat{m}_M = m_M + b_{MR}(m_M - a).
\]
It follows that
\[ x - \omega = -(\omega - \mu_\omega) + \beta R_\beta M_\beta S b_{MR} b_{SM}(\theta - \mu_\theta) + \beta R_\beta M_\beta b_{MR}(b_{SM} - 1)(\mu_{mS} - a) + \beta R(b_{MR} - 1)(\mu_{mM} - a). \]

Denoting \( \beta R_\beta M_\beta S = K, \)
\[ x - \omega = -(\omega - \mu_\omega) + K b_{SR}(\theta - \mu_\theta) + K(b_{SR} - b_{MR})(\frac{\mu_{mS} - a}{\beta S}) + K(b_{MR} - 1)(\frac{\mu_{mM} - a}{\beta S_\beta M}), \]
so
\[ E[x - \omega] = \beta R E[b - 1](\mu_{mM} - a). \]

This shows that the mediator can eliminate bias by choosing \( \mu_{mM} = a. \) Next, noting \( E[(b_{SR} - b_{MR})^2] = 2E[\bar{V}[b] - \bar{C}[b_1, b_2]] \) and \( E[(b_{SR} - b_{MR})] = -(\bar{V}[b] - \bar{C}[b_1, b_2]), \)
\[ E[(x - \omega)^2] = \sigma_\omega^2 + K^2 E[b^2] \sigma_\theta^2 + 2K^2 E[\bar{V}[b] - \bar{C}[b_1, b_2]] \left( \frac{\mu_{mS} - a}{\beta S} \right)^2 \]
\[ + K^2 E[(b - 1)^2] \left( -\frac{\mu_{mM} - a}{\beta S_\beta M} \right)^2 - 2K E[b] \sigma_\theta - 2K^2 E[\bar{V}[b] - \bar{C}[b_1, b_2]] \left( \frac{\mu_{mS} - a}{\beta S} \right) \left( \frac{\mu_{mM} - a}{\beta S_\beta M} \right), \]
which reduces to
\[ E[(x - \omega)^2] = \sigma_\omega^2 (1 - \rho_{\omega\theta}^2) + \sigma_\omega^2 \rho_{\omega\theta}^2 E[(b - 1)^2] \left( 1 + \left( \frac{\mu_{mM} - a}{\beta S_\beta M \sigma_\theta} \right)^2 \right) \]
\[ + 2E[\bar{V}[b] - \bar{C}[b_1, b_2]] \left( \frac{\mu_{mS} - a}{\beta S \sigma_\theta} \right) \left[ \left( \frac{\mu_{mS} - a}{\beta S \sigma_\theta} \right) - \left( \frac{\mu_{mM} - a}{\beta S_\beta M \sigma_\theta} \right) \right]. \]

As the mediator can set \( (\mu_{mM} - a)/(\beta S_\beta M \sigma_\theta) \) to any value, her welfare-maximizing choice is
\[ \left( \frac{\mu_{mM} - a}{\beta S_\beta M \sigma_\theta} \right) = \left( \frac{\bar{V}[b] - \bar{C}[b_1, b_2]}{E[(b - 1)^2]} \right) \left( \frac{\mu_{mS} - a}{\beta S \sigma_\theta} \right), \]
leading to error
\[ E[(x - \omega)^2] = \sigma_\omega^2 (1 - \rho_{\omega\theta}^2) + \sigma_\omega^2 \rho_{\omega\theta}^2 E[(b - 1)^2] \left( 1 + \left( \frac{\bar{V}[b] - \bar{C}[b_1, b_2]}{E[(b - 1)^2]} \right) \left( \frac{\mu_{mS} - a}{\beta S \sigma_\theta} \right)^2 \right). \]

Because the term at the right is strictly positive, error with the mediator is always strictly greater than without, provided \( \mu_{mS} = a. \)
Proof of Fact 7. The one-sender case with convention heterogeneity follows as a special case of Fact 2. Bias is

\[ E[x - \omega] = \frac{\sigma^2_\omega}{\sigma^2_\omega + \sigma^2_\varphi} E[b - 1](\mu_\omega - a), \]

and error is

\[ E[(x - \omega)^2] = \sigma^2_\omega \left( 1 - (1 - E[(b - 1)^2]) \left( \frac{\sigma^2_\omega}{\sigma^2_\omega + \sigma^2_\varphi} \right) \right) + \left( \frac{\sigma^2_\omega}{\sigma^2_\omega + \sigma^2_\varphi} \right)^2 E[(b - 1)^2](\mu_m - a)^2. \]

In the two-sender case with convention heterogeneity, the receiver’s action is

\[ x = \mu_\omega + \sum_{i=1,2} \beta \cdot b_i (m_i - \mu_m) + \sum_{i=1,2} \beta_R \cdot (b_{iR} - 1)(\mu_m - a), \]

where

\[ \beta_R = \frac{\sigma^2_\omega}{2(\sigma^2_\omega + \sigma^2_\varphi) + \sigma^2_\varepsilon}. \]

This has bias

\[ E[x - \omega] = \frac{\sigma^2_\omega}{\sigma^2_\omega + \sigma^2_\varepsilon + \sigma^2_\xi/2} E[b - 1](\mu_\omega - a) \]

and error

\[ E[(x - \omega)^2] = \sigma^2_\omega + 2\beta^2 \left\{ E[b^2]\sigma^2_m + E[b_i b_j] \sigma_{m,m_j} \right\} + 2\beta^2 \left\{ E[(b - 1)^2] + E[(b_i - 1)(b_j - 1)] \right\} (\mu_m - a)^2 - 4E[b]\sigma_{\omega m} \]

\[ = \sigma^2_\omega \left( 1 - \left[ E[(b - 1)^2] - \frac{1}{2}K \right] \left( \frac{\sigma^2_\omega}{\sigma^2_\omega + \sigma^2_\varphi + \sigma^2_\xi/2} \right) \right) \]

\[ + \left( \frac{\sigma^2_\omega}{\sigma^2_\omega + \sigma^2_\varphi + \sigma^2_\xi/2} \right)^2 \left( K \sigma^2_{\varepsilon} + \left( E[(b - 1)^2] - \frac{1}{2}K \right) (\mu_m - a)^2 \right), \]

where

\[ K = \mathbb{V}[b] - \mathbb{C}[b_{1R}, b_{2R}] > 0 \]

Proof of Proposition 3. Note

\[ \sigma_{\omega m} = \left( \frac{\eta + \beta_R}{\eta + \beta^2_{R}} \right) \left( \frac{\sigma^2_{\omega \theta}}{\sigma^2_\theta} \right) \quad \text{and} \quad \sigma^2_m = \left( \frac{\eta + \beta_R}{\eta + \beta^2_{R}} \right)^2 \left( \frac{\sigma^2_{\omega \theta}}{\sigma^2_\theta} \right) + \left( \frac{\beta_R}{\eta + \beta^2_{R}} \right)^2 \sigma^2_\varepsilon, \]
so equilibrium is characterized by

\[
\beta_R = \frac{\sigma_{\omega m}}{\sigma_m^2} = \frac{(\eta + \beta_R^2)(\eta + \beta_R)}{(\eta + \beta_R)^2 + \beta_R^2[\sigma_\omega^2/\sigma_\theta^2]}.
\]

Rearranging slightly, we see that \(\beta_R \geq 0\) constitutes an equilibrium if and only if it is a zero of the function

\[
f(\beta_R) \equiv K\beta_R^3 + \eta\beta_R^2 + \eta(\eta - 1)\beta_R - \eta^2,
\]

where \(K = \sigma_\xi^2/\sigma_\theta^2 > 0\). Note that \(f(0) = -\eta^2 < 0\) and that

\[
f(1) = K > 0,
\]

demonstrating that there is a zero of \(f\) between \(\beta_R = 0\) and \(\beta_R = 1\). Furthermore, as \(f(0) < 0\) and

\[
f''(\beta_R) = 2(\eta + 3K\beta_R) > 0,
\]

there can be only a single zero on \(\beta_R \geq 0\). To show dependence on \(\eta\) we use the implicit function theorem:

\[
\frac{\partial}{\partial \eta} \beta_R(\eta) = \frac{(\beta_R - 1)(\beta_R + 2\eta)}{2\eta\beta_R - (1 - \eta)\eta + 3K\beta_R^2} = \frac{\beta_R((1 - \beta_R)\eta^2 + K\beta_R^3)}{\eta(\eta\beta_R^2 + \eta^2 + 2K\beta_R^3)} > 0,
\]

where the second equality substitutes \(f(\beta_R(\eta)) = 0\). To see what happens as \(\eta \to \infty\), consider that \(f\) has the same zeros as

\[
\tilde{f}(\beta_R) = \frac{1}{\eta^2} \cdot \frac{\tau_{\theta}}{\tau_z} \left(1 + \frac{\tau_{\theta}}{\tau_s}\right) \beta_R^3 + \frac{1}{\eta} \beta_R(\beta_R - 1) + \beta_R - 1,
\]

and as \(\eta \to \infty\), the function \(\tilde{f}\) approaches, \(\beta_R - 1\), with a unique zero at \(\beta_R = 1\). Likewise, as \(\eta \to 0\), the original function \(f\) approaches

\[
f_{\eta \to 0}(\beta_R) = \frac{\tau_{\theta}}{\tau_z} \left(1 + \frac{\tau_{\theta}}{\tau_s}\right) \beta_R^3,
\]

with a unique zero at \(\beta_R = 0\). To show dependence on \(\tau_z\) we again use the implicit function theorem:

\[
\frac{\partial}{\partial \tau_z} \beta_R(\tau_z) = -\frac{-\frac{\tau_{\theta}}{\tau_z} \left(1 + \frac{\tau_{\theta}}{\tau_s}\right) \beta_R(\tau_z)^3}{f'(\beta_R(\tau_z))} > 0
\]

As \(\tau_z \to \infty\), the function \(f\) approaches \(f_{\tau_z \to \infty}(\beta_R) = \eta(\eta + \beta_R)(\beta_R - 1)\), with a single positive zero at \(\beta_R = 1\). Likewise, as \(\tau_z \to 0\), \(f(\beta_R) > 0\) for arbitrarily small values of \(\beta_R\), demonstrating
that the unique equilibrium value $\beta_R \rightarrow 0$. Lastly, regarding error in equilibrium, note:

$$E[(x - \omega)^2] = \sigma_\omega^2 + \beta_R^2 \mathbb{E}[m] - 2\beta_R \mathbb{E}[\omega, m] = \sigma_\omega^2 - \frac{\beta_R(\eta + \beta_R)}{\eta + \beta_R^2} \frac{\sigma_{\omega b}^2}{\sigma_b^2},$$

and, owing to the equilibrium characterization equation,

$$\frac{\beta_R(\eta + \beta_R)}{\eta + \beta_R^2} = \frac{(\eta + \beta_R)^2}{(\eta + \beta_R)^2 + K \beta_R^2} = \left(1 + K \left(\frac{\beta_R}{\eta + \beta_R}\right)^2\right)^{-1}.$$

Then,

$$\frac{d}{d\eta} \left(\frac{\beta_R(\eta)}{\eta + \beta_R(\eta)}\right) = -\frac{\beta_R^2(\eta^2 + \eta \beta_R + K \beta_R^2)}{(\eta + \beta_R)^2(\eta^2 + \eta \beta_R^2 + 2K \beta_R^2)} < 0,$$

so it follows $E[(x - \omega)^2]$ is strictly decreasing in $\eta$. Moreover, as $\eta \rightarrow \infty$,

$$E[(x - \omega)^2] \searrow \sigma_\omega^2 - \frac{\sigma_{\omega b}^2}{\sigma_b^2} = \sigma_\omega^2(1 - \rho_{\omega b}^2),$$

which is the level of error in the pure coordination case.

Proof of Fact 9. Observe

$$E[\mathcal{U}_S(x|\omega, \xi, m)] = -E[(x - [\omega + \xi])^2 + \eta(m - \mu_\omega(\theta))^2]$$

$$= -\left\{ E[(x - \omega)^2] + E[\xi^2 + \eta(m - \mu_\omega(\theta))^2] - 2E[\xi(x - \omega)] \right\},$$

and denote the extent to which $S$ and $R$ lose from miscommunication by $\Delta_S$ and $\Delta_R$. Then

$$\Delta_S - \Delta_R = 2E[\xi(x - \omega)]$$

$$= 2(\mu_\xi E[x - \omega] + \mathbb{C}[\xi, x - \omega])$$

$$= 2 \left( \frac{\beta_R^2}{\eta} \cdot E[b - 1] \cdot \mu_\xi^2 + \beta_R \cdot E[b - 1] \cdot \mu_\xi \cdot (\mu_\omega - E[a]) + \frac{\beta_R^2}{\beta_R^2 + \eta} \cdot E[b] \cdot \sigma_\xi^2 \right).$$

It follows that, $\Delta_S \geq \Delta_R$ if $\mu_\xi(\mu_\omega - E[a]) \geq 0$. Moreover, $\Delta_S$ increases arbitrarily high with $\mu_\xi^2$, so $\Delta_S > 0$ for sufficiently large $\mu_\xi^2$.

Proof of Fact 10. The sequences of expectations of beliefs are immediate from their definitions. The conditional variance of $\mu_t^N$ is

$$\mathbb{V}[\mu_t^N | \omega] = \mathbb{V}[b_{1N}] \left( \frac{\tau_\omega^2 + 2\tau_\epsilon \tau_\omega}{\tau_t^2} \right) \mu_\omega^2 + \left( \frac{\tau_\epsilon}{\tau_t} \right)^2 \sum_{i=1}^{t} \mathbb{V}[b_{sN} \theta_s | \omega] \rightarrow 0,$$
and the conditional variance of $\mu_t^N$ is

$$V[\mu_t | \omega] = \left( \frac{\tau_e}{\tau_t} \right)^2 \sum_{s,r=1}^{t} \omega^2 C[b_{st}, b_{rt}] + \left( \frac{\tau_e \omega}{\tau_t} \right)^2 (\mu_\omega - a)^2 V[b_{1t}] + \frac{\tau_e \omega}{\tau_t} \omega (\mu_\omega - a) \sum_{s=1}^{t} C[b_st, b_{1t}]$$

$$\rightarrow \omega^2 C[b_1, b_2] > 0.$$

To compute error, first express

$$\mu_t - \omega = \frac{\tau_s}{\tau_t} \sum_{s=1}^{t} ((b_{st} - 1)(\omega - \mu_\omega) + (b_{st} - 1)(\mu_\omega - a) + b_{st} \varepsilon_s) + \frac{\tau_\omega}{\tau_t} \omega (\mu_\omega - a) + (b_{1t} - 1)(\mu_\omega - a)).$$

Then:

$$E[(x - \omega)^2] = \left( \frac{\tau_e}{\tau_t} \right)^2 \sum_{s,r=1}^{t} (E[(b_{st} - 1)(b_{rt} - 1)](\sigma_\omega^2 + (\mu_\omega - a)^2) + 1_{(s=r)} E[b^2] \sigma_\varepsilon^2)$$

$$+ \left( \frac{\tau_\omega}{\tau_t} \right)^2 (\sigma_\omega^2 + E[(b - 1)^2](\mu_\omega - a)^2)$$

$$+ 2 \left( \frac{\tau_e \tau_\omega}{\tau_t^2} \right) \sum_{s=1}^{t} (-E[b - 1] \sigma_\omega^2 + E[(b_{1t} - 1)(b_{st} - 1)](\mu_\omega - a)^2)$$

$$\rightarrow (E[(b - 1)^2] + C[b_1, b_2]) (\sigma_\omega^2 + (\mu_\omega - a)^2).$$

---

APPENDIX B. ADDITIONAL EXTENSIONS

B.1. Noisy Encoding

Language conventions as defined in the main model are bijections between messages and words, in contrast to the standard approach in cheap-talk models, in which each word/message is associated with a subset of underlying states. This subsection demonstrates that the results of the main model are not being driven by this discrepancy. To do that, I introduce a variation on the definition of conventions that comes closer to the partitional approach in the literature.

Suppose that when $S$ intends to send message $m$, she in fact communicates the word she associates with $m + \varepsilon$, where $\varepsilon$ is mean-zero independent Gaussian noise. Unlike in Section 4.2, which considered garbling in $S$’s source of information, this noise is part of $S$’s language convention. Specifically, the variance $\sigma_{\varepsilon,S}^2$ is treated as an independent dimension of $S$’s convention, and when $R$ conditions his action on $S$’s message, he accounts for $\varepsilon$ using his own variance $\sigma_{\varepsilon,R}^2$. In the interest
of simplifying the exposition further, assume that the sender always sets \( \beta_S = 1 \). It follows that \( R \) sets

\[
\beta_R = \frac{\sigma_{\omega \theta}}{\sigma_{\theta}^2 + \sigma_{\varepsilon,R}^2},
\]

where the \( \sigma_{\varepsilon,R}^2 \) in the denominator reflects the receiver anticipating the sender’s message to incorporate his own level of noisiness, not the sender’s actual level \( \sigma_{\varepsilon,S}^2 \). When the noise term \( \varepsilon \) is added to the communication function as already defined, \( R \)'s action is

\[
x = \mu_m + \beta_R b(\theta - \mu_\theta) + \beta_R(b - 1)(\mu_m - a) + \beta_R \varepsilon.
\]

It follows from the above expression that bias is

\[
E[x - \omega] = E\left[\frac{\sigma_{\omega \theta}}{\sigma_{\theta}^2 + \sigma_{\varepsilon,R}^2} \cdot b - 1 \cdot (\mu_m - E[a])\right],
\]

and error is

\[
E[(x - \omega)^2] = \sigma_{\omega}^2 \cdot (1 - \rho_{\omega \theta}^2) + \sigma_{\omega}^2 \cdot \sigma_{\theta}^2 \cdot E\left[\frac{\sigma_{\theta}^2}{\sigma_{\theta}^2 + \sigma_{\varepsilon,R}^2} \cdot (b - 1)^2\right] + \sigma_{\varepsilon,R}^2 \cdot (b - 1)^2 \cdot E[(\mu_m - a)^2] + \sigma_{\varepsilon,R}^2 \cdot \sigma_{\varepsilon,S}^2 \cdot E[b^2].
\]

Bias differs from its counterpart in Fact 2 only in that the receiver’s regression coefficient \( \beta_R \) depends on his own convention, and is thus random. Fixing everything else, the magnitude of bias is lower in the noisy case, but this is only because the sender’s message is less informative, and thus the receiver discounts it further. The expression for error also bears close resemblance to Fact 2, but it contains an additional term that is solely due to the noisiness \( \varepsilon \). As an extreme case, supposing that variability in \( \sigma_{\varepsilon}^2 \) is the only source of convention heterogeneity, bias is zero, and error is given by the Bayesian level of error plus the additional noise term.

**B.2. Correlated Conventions**

This subsection considers additional variations to the model’s distributional assumptions. First, recall the general expression for the receiver’s action:

\[
x = \mu_\omega + \beta_S \beta_R \cdot b(\theta - \mu_\theta) + \beta_R(b - 1)(\mu_m - a).
\]
In deriving the expected bias, we assume several things: (1) that the exuberance ratio $b$ is independent of the state and the signal, (2) that there is a single anchor, and (3) that the sender’s and receiver’s conventions are *i.i.d.*

If the state or signal can be correlated with the exuberance ratio, then

$$\mathbb{E}[x - \omega] = \beta_R \mathbb{E}[b - 1](\mu_m - a) + \frac{\sigma_{\omega \theta}}{\sigma^2_b} \cdot \mathbb{C}[b, \theta].$$

This shows that the exaggeration effect is now augmented by the covariance between the signal and the exuberance ratio. In particular, if more exuberant senders also have lower signals, then the composite of the two effects might well be understatement and not exaggeration. Similarly, if there is no common anchor, but rather a random pairwise anchor that may be correlated with the exuberance ratio, then

$$\mathbb{E}[x - \omega] = \beta_R \left( \mathbb{E}[b - 1](\mu_m - \mathbb{E}[a]) - \mathbb{C}[a, b] \right),$$

which hinges critically on the covariance between the exuberance ratio and the pairwise anchor.

If the sender’s and receiver’s conventions are not *i.i.d.*, the consequences depend critically on how they are not *i.i.d.* In general,

$$\mathbb{E}[x - \omega] = \beta_R\mathbb{E}[b - 1](\mu_m - a)$$

as before, but there are now several possibilities. Let $C_*$ be an arbitrary numeraire convention, so $b = b_{SC_*} b_{RC_*}^{-1}$. If the distribution over conventions is symmetric, that is if $(b_{SC_*}, b_{RC_*})$ is as likely as $(b_{RC_*}, b_{SC_*})$, then the exaggeration effect is still guaranteed:

$$\mathbb{E}[b] = \frac{1}{2} \mathbb{E}[b_{SC_*}\cdot b_{RC_*}^{-1}] + b_{SC_*}\cdot b_{RC_*}] > 1.$$

If $b_{SC_*}$ and $b_{RC_*}$ are independent, but not from the same distribution, then

$$\mathbb{E}[b] = \mathbb{E}[b_{SC_*}]\cdot \mathbb{E}[b_{RC_*}^{-1}]$$

could be anything. If $b_{SC_*}$ and $b_{RC_*}$ are drawn from the same distribution but not independently, and without the symmetry condition above, then

$$\mathbb{E}[b] = \mathbb{E}[b_{SC_*}]\cdot \mathbb{E}[b_{RC_*}^{-1}] + \mathbb{C}[b_{SC_*}, b_{RC_*}],$$

47
which could be any value depending on the covariance term.

B.3. Monotonicity and Continuity

Finally, it is worth noting that the foundational assumption of monotonicity between conventions – that different conventions rank the words in the same order – can be rephrased as a statement about continuity. To even talk about the continuity of the encoding and decoding functions, however, requires the unstructured word set \( W \) to be augmented with a topology, and there is no reason \( W \) should have any particular topology. The following proposition instead gives five equivalent ways to understand two conventions as having the same topological relationship with \( W \), and establishes that these conditions are all but equivalent to monotonicity. At an intuitive level, the link between monotonicity and continuity springs from the fact that the standard topology on the real numbers is also the order topology.

**Proposition 4.** Conventions are monotonic if and only if each pair of conventions \( S \) and \( R \)

1. ranks at least one pair of words in the same order:

   \[ m_S(w_2) > m_S(w_1) \text{ and } m_R(w_2) > m_R(w_1) \]

   for some \( w_1, w_2 \in W \), and

2. satisfies one of the following equivalent continuity conditions:
   (a) For any word \( w \in W \) and any sequence of words \( \{w_k\}_{k=1}^\infty \), \( m_S(w_k) \to m_S(w) \) if and only if \( m_R(w_k) \to m_R(w) \).
   (b) Decoding functions \( m_S \) and \( m_R \) induce the same topology on \( W \).
   (c) Encoding function \( w_S \) is continuous if and only if \( w_R \) is continuous.
   (d) Decoding function \( m_S \) is continuous if and only if \( m_R \) is continuous.
   (e) Communication functions \( h_{SR} \) and \( h_{RS} \) are continuous.

The most readily interpretable of the continuity conditions is (a), that whenever one convention understands a sequence of words as converging to a limit, all others do too. Failure of that condition would mean, intuitively, that some language conventions see a pair of words as practically interchangeable while others understand them as radically different. Such discrepancy might be plausible in some real-world contexts, but this paper concentrates on situations where people are talking about degrees – pain, risk, profit, etc. – and a relevant counterexample is not readily
forthcoming.

**Proof of Proposition 4.** Let $S$ and $R$ be any two conventions. We first verify that the continuity conditions are equivalent. Recall that the topology induced by decoding function $m_C$ is the set \( \{ w_C(U) \mid U \subset R \text{ open} \} \) for $C = S, R$.

We show that condition (b), $\tau_S = \tau_R$, is equivalent to each of the following:

(c) $w_S$ is continuous if and only if $w_R$ is continuous. Note that if $W$ is endowed with topology $\tau$, $w_C$ is continuous if and only if $\tau \subset \tau_C$. Hence, if $\tau_S = \tau_R$, $w_S$ continuous implies $\tau \subset \tau_S = \tau_R$, which implies $w_R$ is continuous, and visa-versa. If $\tau_S \neq \tau_R$, then – without loss of generality – suppose $\tau_S \not\subset \tau_R$; then, under topology $\tau_R$, $w_R$ is continuous while $w_S$ is not.

(d) $m_S$ is continuous if and only if $m_R$ is continuous. Note that if $W$ is endowed with topology $\tau$, $g_C$ is continuous if and only if $\tau_C \subset \tau$. Hence, if $\tau_S = \tau_R$, $m_S$ continuous implies $\tau \supset \tau_S = \tau_R$, which implies $m_R$ is continuous, and visa-versa. If $\tau_S \neq \tau_R$, then – without loss of generality – suppose $\tau_S \not\supset \tau_R$; then, under topology $\tau_S$, $m_S$ is continuous while $m_R$ is not.

(e) $h_{SR} = m_R \circ w_S$ and $h_{RS} = m_S \circ w_R$ are continuous. If $\tau_S = \tau_R$, endow $W$ with $\tau = \tau_S = \tau_R$; by parts (a,b) it follows $w_S$, $w_R$, $m_S$, and $m_R$ are all continuous, and therefore the compositions $m_R \circ w_S$ and $m_S \circ w_R$ are also continuous. If $\tau_S \neq \tau_R$, then – without loss of generality – there exists $V \in \tau_S$ with $V \not\subset \tau_R$. Note that by definition $V = w_S(U)$ for some open set $U \subset R$. Then it follows $(m_S \circ w_R)^{-1}(U) = m_R \circ w_S(U) = m_R(m)$, but $m_R(m)$ cannot be open as $w_R(m_R(m)) = V \not\subset \tau_R$. This shows $m_S \circ w_R$ is not continuous.

Next, we show that condition (e) is equivalent to:

(a) For any word $w \in W$ and any sequence of words $(w_k)_{k=1}^{\infty}$, $m_S(w_k) \rightarrow m_S(w)$ if and only if $m_R(w_k) \rightarrow m_R(w)$. Suppose $h_{SR}$ and $h_{RS}$ are continuous, and, without loss of generality, let $(w_k)_{k=1}^{\infty}$ be a sequence such that $m_S(w_k) \rightarrow m_S(w)$. Denoting $m_k = m_S(w_k)$ and $m = m_S(w)$, it follows from continuity that $m_R(w_k) = h_{SR}(m_k) \rightarrow h_{SR}(m) = m_R(w)$. If, say, $h_{SR}$ is not continuous, then there exists message sequence $(m_k)_{k=1}^{\infty}$ such that $m_k \rightarrow m$ for some $m$ while $h_{SR}(m_k) \not\rightarrow h_{SR}(m)$. Denoting $w_k = w_S(m_k)$ and $w = w(m)$, it follows $m_S(w_k) = m_k \rightarrow m = m_S(w)$ while $m_R(w_k) = h_{SR}(m_k) \not\rightarrow h_{SR}(m) = m_R(w)$.

49
We now demonstrate that condition (e), coupled with condition (1) in the proposition statement, is equivalent to monotonicity. Recall from real analysis that: *if I is an interval of the real line, invertible function f : I → I is continuous if and only if it is strictly monotonic.* Thus, if condition (e) holds, then $h_{SR}$ and $h_{RS}$ must be either strictly increasing or strictly decreasing; if (1) also holds, they must be strictly increasing, and that implies monotonicity everywhere. Likewise, if monotonicity holds, then (1) follows trivially, and, as $h_{SR}$ and $h_{RS}$ are therefore strictly increasing, they must be continuous as well.

---

18 Proof: suppose $f$ is continuous. Were it not strictly monotonic, then there would exist points $x < y < z$ such that, without loss of generality, $f(x) \leq f(z) \leq f(y)$; by the intermediate value theorem, there would exist $y' \in [x, y]$ such that $f(y') = f(z)$, contradicting the invertibility of $f$. Now suppose $f$ is strictly monotonic. Let $(x, y)$ be an open interval in $I$, and without loss of generality say $f^{-1}(x) < f^{-1}(y)$. Then, $f^{-1}(x, y) = \{z > f^{-1}(x)\} \cap \{z < f^{-1}(y)\}$, an open set. As the set of open intervals is a basis for the (subspace) topology on $I$, it follows $f$ is continuous.