Strategic Information Transmission: A Survey of Experiments and Theoretical Foundations *

Andreas Blume[†] Ernest K. Lai[‡] Wooyoung Lim[§]

April 29, 2017

Abstract

This paper surveys the experimental literature on communication of private information via cheap-talk messages. Some players have private information that is relevant to the decisions made by others. The informed players have the option to send messages before decisions are made. Messages do not directly affect payoffs but can link decisions to information. This simple paradigm has been found useful in philosophy, linguistics, economics, political science and theoretical biology. The survey tracks the dialogue between the theory and experiments that have been inspired by this paradigm and is organized around the themes of available language, payoff structure and compositions of the sets of communicating players.

*We thank Vince Crawford and Joel Sobel for comments and encouragement.

[†]Department of Economics, The University of Arizona, ablume@email.arizona.edu

[‡]Department of Economics, Lehigh University, kwl409@lehigh.edu

[§]Department of Economics, The Hong Kong University of Science and Technology, wooyoung@ust.hk

1 Introduction

We survey the experimental literature on communication of private information via cheap-talk messages. The focus is on work that is firmly grounded in theory. We discuss the theoretical background (see also the recent survey by Sobel [94]), emphasize puzzles that theory poses for the lab, and try to identify possible novel perspectives.

There are large swaths of experimental work on communication that we do not cover, including communication of intentions (some of which is surveyed by Crawford [26]), costly signaling in the tradition of Spence [92] (for an example and additional references, see Kübler, Müller, and Normann [67]), disclosure as pioneered by Grossman [52] and Milgrom [74] (for some recent work and references, see Hagenbach and Perez-Richet [53]), and Bayesian persuasion *à la* Kamenica and Gentzkow [58] (see Nguyen [81] for an experiment on Bayesian persuasion). The principal omission is the burgeoning literature on lying and deception that was inaugurated by Gneezy [46]. Sobel [95] defines lying and deception in strategic situations, explores them as equilibrium phenomena, and discusses other theoretical treatments. Abeler, Nosenzo and Raymond [1] survey the experimental literature, conduct additional experiments, and propose preferences to organize the data.

We focus on well-defined game settings where both senders and receivers are strategic actors (as opposed to having the experimenter be the audience), the environment is public knowledge, and an attempt is made to control payoffs through monetary rewards (rather than trying to infer social, lying or deception preferences from choices). This is restrictive, and ignores the myriad psychological and moral motivations that impinge upon communication, as already noted by Adam Smith [91] in the *Theory of Moral Sentiments*, when he states: "But though the violation of truth is not always a breach of justice, it is always a breach of a very plain rule, and what naturally tends to cover with shame the person who has been guilty of it" (p. 206). We emphasize connections to philosophy of language and linguistics and the limitations of looking at communication of private information through the lens of sender-receiver games.

2 Theory background and questions for experiments

A great deal of communication serves to convey private information. This distinguishes it from other communicative acts such as signaling intentions, asking questions and making promises or threats. Those with private information may have an incentive to share some or all of it with others who are in a position to act on that information. In doing so, they act strategically. The messages they send are functions of their information. Likewise, the actions taken in return are functions of the messages received.

The simplest formal representations of such strategic information transmission are sender-receiver games. A single privately informed Sender sends a message to a Receiver who responds with an action that affects both parties' payoffs. Messages are not directly payoff relevant and therefore communication is purely informational. Sender-receiver games are the archetype of situations in which communication is essential; the only way to bring the sender's information to bear on the receiver's decision is via communication.

2.1 Origins of sender-receiver games – common interests

David Lewis [70] in his study of the conventionality of language introduces sender-receiver games by way of an example. The sexton of Old North Church signals to Paul Revere whether the redcoats "stay home," "set out by land," or "set out by sea." The available signals are to put either "no lantern," "one lantern," or "two lanterns" in the belfry. Paul Revere, after observing the signal, decides whether to "go home," "warn the countryside that the redcoats are coming by land," or "warn the countryside that the redcoats are coming by sea."

In the story there are three states of the world: the redcoats stay home, set out by land, or set out by sea. Denote these states by t_1, t_2 and t_3. The sexton is the sender, S. He has three messages, no lantern, one lantern, and two lanterns. We will represent these messages by m_1, m_2, and m_3. Paul Revere is the receiver, R. He has three actions, a_1, a_2, and a_3 corresponding to whether to go home, warn the countryside that the redcoats are coming by land, or warn that they are coming by sea. The preferences of S and R are perfectly aligned, as in the payoff table for Lewis's Game in Figure 1, where the first entry in each cell is the sender's payoff and the second entry the receiver's payoff for that state-action combination. Both receive a positive payoff if and only if the receiver's action matches the state of the world. This is the case, for example, for the state-action pair (t_2, a_2) where Paul Revere, correctly, warns the country side that the red coats set out by land.

	a_1	a_2	a_3
t_1	8,8	0,0	0,0
t_2	0,0	8,8	0,0
t_3	0,0	0,0	8,8

Figure 1: Lewis's Game

To turn this into a fully specified game, assume that players have a common prior belief π that assigns equal probability to all three states. A pure strategy $\sigma : T \to M$ for the sender is a map from the state space $T = \{t_1, t_2, t_3\}$ into the message space $M = \{m_1, m_2, m_3\}$. Likewise, a pure strategy $\rho : M \to A$ for the receiver is a map from the message space M into the action space $A = \{a_1, a_2, a_3\}$.

Their simplicity makes sender-receiver games ideally suited for experimentation. They can be used to address questions about the role of incentives in communication as well as about the significance of access to a common language. They are a natural starting point for the investigation of alternative communication protocols and ways of facilitating communication.

Even within the class of sender-receiver games, Lewis's game is special: state spaces, message spaces, and action spaces are small; players have common interests, i.e. their payoffs are the same in every state of the world; and, the cardinalities of state- message- and action spaces coincide. In common-interest games Lewis singles out strategy profiles (σ, ρ) with effective communication, such as $\sigma(t_i) = m_i$ and $\rho(m_i) = a_i$, $i = 1, 2, 3$, which he refers to as "signaling systems" and we would call "separating equilibria." In a separating equilibrium the sender's strategy reveals the state and the receiver chooses the optimal action for every state of the world.

Lewis is interested in the conventionality of language. Paraphrasing, Lewis defines a convention as a regularity of behavior of a population in a recurrent situation that is supported by a system of expectations and preferences that is common knowledge in the population. Agents achieve coordination by reference to a basis for common knowledge, which may be prior agreement, salience, or a precedent. The basis for common knowledge justifies their first- and higher order expectations, which in turn justify their behavior given their preferences. In the communication setting, a signaling convention implements a signaling system, or a separating equilibrium.

Lewis's discussion of the significance of agreement, salience, and precedent for achieving coordination is very much inspired and influenced by Thomas Schelling [87]. In both cases agreement, salience, and precedent are sources of matching first- and higher-order expectations in coordination games. Lewis pushes that reasoning to the limit by providing a definition of common knowledge and identifying agreement, salience, and precedent as bases for common knowledge.

It is worth noting that players in Lewis's game do not use natural language. The Sexton and Paul Revere agree on how to use a limited set of signals that do not belong to their natural language (the pre-play communication by which they reach agreement remains unmodeled). Their agreement applies to a single situation. They adhere to the agreement because they expect each other to adhere to the agreement. They explicitly agree on a separating equilibrium.

Separating equilibria are instances of Bayesian Nash equilibria (BNE). In a Bayesian Nash equilibrium each player maximizes his payoff given his information (his type) and given the strategies of other players. It is routine when studying sender-receiver games to focus on Perfect Bayesian Nash equilibria, a subset of Bayesian Nash equilibria with the property that strategies remain optimal also off the equilibrium path, given some beliefs. In terms of outcomes, i.e. joint distributions over types and actions, this makes no difference in sender-receiver games and we will henceforth simply talk about 'equilibria.'

The equilibrium perspective as such is silent about how players arrive at an equilibrium, at which equilibrium they arrive, or even if they arrive at an equilibrium. The focus is on characterizing the set of equilibria and studying their properties. In Lewis's game, besides the already mentioned separating equilibria, there are partial-pooling equilibria in which the sender sends the same message for two of the states and only distinguishes the third state, and there are pooling equilibria in which the sender sends the same message, regardless of the state. There are also variants of these equilibria in which the sender mixes over messages. The sender may for example "babble" and randomize over messages in a way that does not depend on the state. In that case messages are uninformative and it is optimal for the receiver simply to ignore messages.

The multiplicity of equilibria in Lewis's game is a natural target for conducting experiments. This is especially true given that common payoff-based refinements of (Bayesian) Nash equilibrium do not narrow down the set of outcomes (joint distributions over states and actions) when messages are cheap talk. Every outcome that is supported by a Nash equilibrium can be supported by a Perfect Bayesian Nash Equilibrium – simply replace responses and beliefs after unsent messages by responses and beliefs after sent messages. For the same reason, sequential equilibrium (Kreps and Wilson [64]) has no cutting power. Equilibrium refinements like the intuitive criterion (Cho and Kreps [24]) that rely on placing restrictions on beliefs following off-equilibrium messages can be very powerful in games with costly signaling à la Spence [92]. They are powerless in sender-receiver games since every equilibrium outcome can be supported by an equilibrium in which all messages are used with positive probability. Equilibrium refinements building on trembles and culminating in strategic stability (Kohlberg and Mertens, [61]) are equally ineffective. The game in Figure 1, for example, has an equilibrium in which both players assign positive probability to all of their pure strategies. That equilibrium is strategically stable, and no information is transmitted.

The experimental laboratory allows one to vary the conditions that Lewis names as possible bases for common knowledge, which are agreement, salience and precedent. It stands to reason that separating equilibria have a better chance of being attained if players can rely on the full resources of a shared natural language or, short of that, can at least rely on focal associations between messages and states or actions. If such focal associations are absent, it is implausible that players would achieve immediate coordination on a separating equilibrium. They may, however, be able to learn to build such associations and create the types of precedents that Lewis has in mind. Experiments are a natural tool for investigating what it takes for players to produce coordinated outcomes in sender-receiver games.

Looking at strategic information transmission through the lens of sender-receiver experiments forces questions that either matter little or are easily missed in the game-theoretic analysis: What exactly is the message space? How are messages, states, and actions described to the players? What is the population of players? What is the interaction pattern in the population? What is the information structure? Who gets what information about past interactions? What is the payoff structure? Is the payoff structure the same or different across utterance occasions? Is the description of messages, states, and actions the same or different across utterance occasions?

2.2 Sender-receiver games with (some) conflict of interest

Crawford and Sobel (CS) [25] shift the focus toward partial conflict of interest. They also impose structure on state and action spaces that lets one talk about proximity of states and actions. Specifically, they consider a unidimensional state space $T = [0, 1]$ and a unidimensional action space $A = \mathbb{R}$. The state is drawn from a common knowledge distribution with support T and density f. The payoff functions of the sender, $U^S(a, t)$, and the receiver, $U^R(a, t)$, are assumed to be strictly concave in a, with unique maximizers (ideal actions) $a^i(t)$, $i = S, R$, and satisfy a (single crossing) condition that ensures that each player i's ideal action $a^i(t)$ is monotonically increasing in the state, t. With these payoff functions both players prefer actions closer to their ideal actions.

The sender is assumed to be upward biased so that ideal actions satisfy $a^S(t) > a^R(t)$ for all t. An obvious consequence is that there are no separating equilibria; in any candidate separating equilibrium the sender would have an incentive to deviate toward mimicking a higher type. The key insight of CS is that in this environment all equilibria have a simple form: the state space can be partitioned into finitely many intervals, with all types belonging to the same interval inducing the same action and types from different intervals inducing distinct actions. With only moderate conflict of interest, there will be equilibria in which information is transmitted and the sender's message influences the receiver's actions. There is, however, a limit to how much information can be transmitted in equilibrium. For any given preferences and prior beliefs there is a finite upper bound on the number of equilibrium actions.

Communication is integral to most human activity. Much of that communication involves information sharing under conditions where preferences are not fully aligned. It is therefore not surprising that variations of the CS model have been widely used in applications. These applications include, to mention just a few, in political science the study of legislative debate (Austen-Smith [4]), in accounting explanations for limited voluntary disclosure (Gigler [44]), in finance rationalization of the coarseness of credit ratings (Goel and Thakor [47]), and in economics information sharing in duopoly (Goltsman and Pavlov [50]), the information content in stock reports (Morgan and Stocken [77]), and policy announcements by the federal reserve (Stein [96]).

Qualitatively, some of the forces present in the games considered by CS can be captured by the slight modification of Lewis's game in Figure 2, which we may call "Symbolic Crawford Sobel" (SCS). In SCS we change the payoffs from Lewis's game to create an incentive for type t_2 to mimic t_1 and for type t_3 to mimic t_2, and leave the prior and the message space $M = \{m_1, m_2, m_3\}$ unchanged. If we think of types with a lower index as higher types, we now have an incentive structure in which the sender has an incentive to exaggerate his type, but not by too much – type t_3 would like to pretend to be type t_2 but does not want to be misidentified as t_1. Clearly, there is no separating equilibrium; type t_2, for example, would want to deviate to sending the message sent by t_1 in any putative equilibrium in which the sender uses a separating strategy. There is, however, a partial pooling equilibrium in SCS in which types t_1 and t_2 both send message m_1 and type t_3 sends message m_3; note that message

m_2 is not sent in this equilibrium. Therefore, beliefs following (an unexpected) message m_2 are not pinned down by Bayes' rule. If we postulate that following m_2 the receiver assigns equal probability to t_1 and t_2, his unique optimal response to m_2 is action a_1 and therefore none of the three types has an incentive to start sending message m_2.

	a_1	a_2	a_3
t_1	8,8	0,0	0,0
t_2	9,5	8,8	0,0
t_3	0,0	9,5	8,8

Figure 2: Symbolic Crawford Sobel

From an experimental perspective, it is interesting that if preferences are not too far apart, there is considerable multiplicity of equilibria in CS games. There is an equilibrium with a maximal number of partition elements, and corresponding equilibrium actions. There is always an equilibrium with a single partition element, in which the receiver responds to all messages with the same action, the optimal action given prior beliefs. Furthermore, for any number of partition elements in-between these extremes there is a corresponding equilibrium. All these equilibria induce different equilibrium actions, and thus have different equilibrium outcomes. Experiments may provide guidance for how to think about equilibrium selection in CS games.

Call a CS game "uniform quadratic" if the payoff functions are quadratic, $U^R(a,t) = -(t-a)^2$ and $U^S(a,t) = -(t+b-a)^2$, are parameterized by a constant bias parameter $b > 0$, and the type distribution is uniform. With these payoff functions, given the sender's type t, the receiver's payoff is maximized by having the action equal the type; i.e., the receiver's ideal point satisfies $a^R(t) = t$ for all $t \in [0,1]$. The sender, in contrast, would prefer that action $t+b$ is taken; i.e., the sender's ideal point is $a^S(t) = t+b$. Having $a^S(t) - a^R(t) = b > 0$ means that the sender has a constant incentive to exaggerate his type.

In uniform quadratic CS games, the *ex ante* preferred equilibrium for both players is the one with the maximal number of intervals partitioning the type space; the maximal number of intervals is weakly increasing as preferences become more aligned; and, the payoff from the *ex ante* preferred equilibrium is increasing with the degree of incentive alignment. Thus under these (and somewhat more general) conditions there is a sensible comparative statics prediction: closer incentive alignment leads to better information transmission.

CS, partly by appealing to Pareto efficiency itself and partly by invoking Schelling to point out the salience of efficient equilibria, advocate for the *ex ante* efficient equilibrium. This prediction and the attendant comparative statics implications can be evaluated in the lab. Of particular interest is the qualitative prediction that closer incentive alignment improves information transmission.

CS leave the message space unspecified. It has no relevance for their formal analysis, as long as it is large enough to admit all equilibrium outcomes of the game. A sensible additional assumption to make is that the message space imposes no constraints on communication and, for example, contains meaningful expressions for describing states and actions. Only if the message space includes messages that either have preassigned meanings or at least some focal associations with the state or the action space is there a chance for any kind of information transmission. Even if the message space has a large set of messages with preassigned meanings, unlike in the common interest case, there is no straightforward intuition of how messages will be used. While for any equilibrium there is a "truthful" version in which types declare which interval they belong to or recommend the equilibrium action corresponding to "their" interval, this is not an *a priori* conspicuously natural use of messages. One suspects that there may be less of a "basis for common knowledge" for players to coordinate on an equilibrium than in Lewis's common interest game.

This offers both an opportunity and a challenge for experiments. It is a challenge, since one would like to employ the most natural message space, large message spaces are unwieldy, and limiting the size of the message space may require somewhat arbitrary choices. It is an opportunity, because it suggests the value of exploring the effects of using different message spaces.

2.3 Additional senders and/or receivers

Lewis and CS consider the simplest model of communication of private information. Communication is between a single sender and a single receiver; there is a single communication round; and, only the sender is privately informed. The literature in the footsteps of Lewis and CS does look at a variety of natural departures from this basic setup. Two that are of special interest involve varying the number of senders and receivers. We would like to understand, for example, the implications of communicating private information to an audience comprised of multiple receivers. We would also like to know whether and when there is a benefit to consulting multiple experts.

Farrell and Gibbons [37] tackle the first question. They consider a setup with a single sender and two receivers. The states of the world and the receivers' actions are binary, a receiver's payoff depends only on his own action (and the state), and the sender's payoff is the sum of two components, with the jth component depending on the state and the action of receiver j, $j = 1, 2$. Farrell and Gibbons ask how the incentives for truthful revelation to one receiver are affected by the presence of the other receiver and how welfare is affected by having communication conducted either in public or in private. Under public communication, the sender sends a single message that is observed by both receivers. Under private communication the sender sends separate messages to each receiver, and each receiver only observes his own message. Farrell and Gibbons show that, in equilibrium, truthful private communication with each receiver is possible only if truthful public communication is possible. The converse does not hold.

Depending on payoffs the possibilities for achieving separation with private and public communication are as follows. *No communication*: separation is impossible both with private and public communication. *Communication*: there is a separating equilibrium both with private and public communication. *One-sided discipline*: there is a separating equilibrium with only one of the receivers under private communication and there is a separating equilibrium with public communication. *Mutual discipline*: there is a separating equilibrium only with public communication. *Subversion*: there is a separating equilibrium in private communication with one of the receivers, but not with public communication. Regarding welfare, the receivers always (weakly) prefer a separating equilibrium whereas the sender's preferences over equilibria are ambiguous without additional conditions on payoffs.

Goltsman and Pavlov [49] revisit the questions raised by Farrell and Gibbons [37] in a setting with a richer payoff structure and access to a larger variety of communication protocols. Their environment extends the Crawford-Sobel setup to multiple receivers. The state space is the unit interval, the action space corresponds to the real numbers and payoffs depend on the distance between the state and the receiver actions. Each receiver's payoff depends only on his own action and the state, and the sender's payoff depends on the actions of both receivers and the state and is separable in the two actions. The degree of incentive alignment between sender and receivers depends on bias parameters that may differ across receivers.

An example of the class of environments considered by Goltsman and Pavlov is a variant of the uniform quadratic setup of Crawford and Sobel. A sender S faces two receivers, R_1 and R_2. The sender's type t is drawn from a uniform distribution on $[0, 1]$. The sender's payoff $-(a_1 - t)^2 - (a_2 - t)^2$ depends on his type t, action a_1 of receiver R_1, and action a_2 of receiver R_2. Receiver R_i's payoff if $-(a_i - t - b_i)^2$, with $b_i \in \mathbb{R}$, $i = 1, 2$. Depending on the values of b_i, $i = 1, 2$, receivers can be biased relative to the sender and to each other.

Like Farrell and Gibbons, Goltsman and Pavlov compare public with private communication. In addition they investigate "combined communication," which allows the sender simultaneously to send both private and public messages. The overall goal is to find an optimal communication protocol. To this end they further extend the scope of their investigation by analyzing both "mediated communication," in which the sender sends a message to a trusted mediator who then makes action recommendations to the receivers and "multi-stage communication," in which both sender and receivers repeatedly send messages to each other before actions are taken.

The sender's preference for private versus public communication depends on the average receiver bias, $\left|\frac{b_1+b_2}{2}\right|$. If that bias is high, the sender prefers private communication. If instead the average receiver bias is low and the receivers' biases are sufficiently different, the sender prefers public communication to private communication. It is frequently the case that combined communication improves upon both private and public communication. When both receivers are biased in the same direction the optimal mediated communication protocol has the sender employ independent private noisy channels. Alternatively, the optimum can be achieved with unmediated multi-stage communication.

The second question mentioned above, about when a decision maker might benefit from consulting multiple experts, was first addressed by Gilligan and Krehbiel [45]. Subsequently

Krishna and Morgan [65] (KM) provided sufficient conditions for the existence of fully revealing equilibria with a one-dimensional state space, when biases are not too large. The intuition for why full revelation is possible even with biased senders is simple. The senders can be played off against each other. Roughly, if their messages do not agree they get punished. It is worth noting the stark contrast between the lesson from the single-sender and the multiple-sender case. CS find, in the single-sender case, that communication becomes easier if conflict diminishes and that with any level of conflict full revelation is impossible. The work of KM, in contrast, suggests that with multiple senders at least for some range of biases full revelation is possible and conflict does not matter.

Battaglini [5] revisits the multiple-expert problem in the one dimensional case. He gives necessary and sufficient conditions for existence of fully revealing equilibria when there are opposing biases. Importantly, however, Battaglini demonstrates that full revelation in the one-dimensional case requires implausible out-of-equilibrium beliefs.

Battaglini goes on to investigate communication with two senders when the state space is multidimensional. He finds that in this environment for any sender biases full revelation can be supported by equilibria that do not require implausible off-equilibrium-path beliefs. The key to understanding this is to see that with a multidimensional state space, despite the conflict of interest, for each sender one can find a dimension along which that sender's preferences coincide with that of the receiver. Generically these dimensions differ across different senders and therefore span the state space. Given this result guaranteeing the existence of robust fully revealing equilibria with a multidimensional state space, the contrast with CS's intuition about the possibility of information transmission depending on the degree of incentive alignment comes into sharp relief. It is natural then to use experiments to sort out when conflict of interest governs how much information can be transmitted.

2.4 Language games and the puzzle of meaning

Simple sender–receiver games idealize a particular form of communication: it is unilateral; its role is information transmission; and, there are well defined message and action spaces. Message use in these games resembles the "builders' language" in Wittgenstein's [104] *Philosophical Investigations*:

"Let us imagine a language for which the description given by Augustine is right. The language is meant to serve for communication between a builder A and an assistant B. A is building with building-stones: there are blocks, pillars, slabs and beams. B has to pass the stones, and that in the order in which A needs them. For this purpose they use a language consisting of the words "block", "pillar", "slab", "beam". A calls them out; — B brings the stone which he has learnt to bring at such-and-such a call."

The builders' language both illustrates a naive conception of language, which Wittgenstein attributes to Augustine, and serves as an example of what Wittgenstein refers to as a "language game." It describes a scenario in which the meaning of a word can be understood as the object for which the word stands. Analogously, in sender-receiver games the meaning of a message can be thought of as the set of states the message refers to, thus identifying meaning with reference. The builders' language, however, is only an example of a multitude of language games tied together by similarity, or what Wittgenstein calls "family resemblance."

Among language games listed by Wittgenstein are: giving orders, and obeying them; describing the appearance of an object, or giving its measurements; constructing an object from a description (a drawing); speculating about an event; making up a story and reading it; making a joke, telling it; asking, thanking, cursing, greeting, playing. To us these can serve as useful reminders of the limitations of sender-receiver games and the naiveté of thinking of meaning as mere reference.

The multitude of language games in which a word can be deployed is used by Wittgenstein to make the point that words do not have a fixed meaning and that instead one is to look at their *use*. This meaning-use distinction has parallels in the distinction of literal meaning and strategic use of messages in sender-receiver games. In an equilibrium analysis of sender-receiver games the meaning of a message is entirely endogenous (an equilibrium phenomenon), determined by how the sender's strategy maps states into messages and the receiver's strategy maps messages into actions. Meaning then can be thought of as either "indicative meaning," the set of states at which the sender sends that message, or "imperative meaning," the receiver action induced by that message.

While equilibrium confers meaning on messages, this leaves the origin of equilibrium itself in need of explanation. Intuition and experimental evidence suggest that without some structure on the message space, which may stem from uses of the messages outside of the game or prior uses of the messages in repeated sender-receiver interactions, we cannot expect behavior consistent with equilibria in which useful information is transmitted. Unless players have access to messages with literal meaning (meaning outside the game itself) or to messages whose meaning has emerged endogenously in prior interactions there is no hope for meaningful communication.

While the availability of messages with prior meanings (literal, focal or established through prior interactions) is necessary for communication in sender-receiver games, outside of special cases there is generally no reason to expect that the observed use of messages coincides with their prior meaning. Consider a class of uniform-quadratic CS games with biases $b \in [\underline{b}, \overline{b}]$ such that for all biases in this range the maximally informative equilibria induce two equilibrium actions. In such an equilibrium one of the equilibrium actions will be induced by a low interval of types and the other by a high interval of types. If the message space is $\{$"High", "Low"$\}$ one may reasonably expect that the sender will send "Low" for states in the low interval and "High" for states in the high interval. The intervals, however, and thus the meaning of the messages "High" and "Low" will depend on the value of b. Thus, while prior meaning anchors the meaning in the game at hand, prior meaning does not fully determine meaning in a given strategic situation.

Wittgenstein is critical of the "philosophical concept of meaning" with its simplistic view of how language functions. Our discussion above illustrates how even in simple sender-receiver games, where one may expect the Augustinian conception of language to apply, meaning becomes fluid once one recognizes the interaction between prior meaning and incentives. Allowing for communication through noisy channels, mediation, uncertainty about the interlocutors' ability to produce and interpret messages further supports the conclusion that even in sender-receiver games meaning is at best inexact. Then again, as Wittgenstein (p.41) notes, "inexact" does not mean "unusable."

The interplay of prior meaning and incentives is central to lying and deception. Sobel [95] formally defines both; according to these definitions, the sender lies when he sends a messages he believes to be false, given a prior meaning of that message, and he deceives if his message induces an action that is inferior for the receiver to an action that could have been induced with a different message. For Wittgenstein (p.90) lying is one more language game. But how can there be lying and deception in sender-receiver games? In equilibrium the receiver knows the sender's strategy by fiat and thus, it seems, cannot be deceived. Furthermore, if we identify the meaning of a message with its equilibrium meaning, say the set of states for which that message is sent, it seems that there cannot be lying in equilibrium.

It is straightforward to understand how lying and deception can arise when agents are boundedly rational and messages have prior meanings (as pioneered in Crawford [27], and applied with considerable success to data from sender-receiver games by Cai and Wang [21], and Wang, Spezio and Camerer [100]). If receivers are credulous and take messages at face value, then clearly senders can manipulate receivers by using messages in a way that is inconsistent with their prior meaning. But how, one may ask, did the credulous receivers learn prior message meaning in the first place? One possibility is that the typical situation is one where it is in receivers' interest to interpret messages at their face value. But then their credulity is not much different from the behavior of a rational receiver in a partial pooling equilibrium of a sender-receiver game.

The last observation also suggests that it may be useful to revisit the question of lying in equilibrium when message do have prior meaning. Again, consider a uniform-quadratic CS environment with a bias low enough to allow meaningful communication. In equilibrium types will be sorted into intervals with types belonging to the same interval sending a common message and types from different intervals sending distinct messages. But which message will they send? Consider the highest interval of types who send a common message in equilibrium, $[\theta_n, 1]$. A message "the state lies in the interval $[\theta_n, 1]$" might be truthful, but seems stilted from the view of natural language. What would prevent the highest type, type 1, from simply declaring his type? Wouldn't that be the most straightforward way to induce the maximal equilibrium action? This is an empirical question and can be addressed in the lab. But assume for the moment that all types in the interval $[\theta_n, 1]$ send the common message "the state is 1." Then almost all types in that interval are lying.

In the last example, is the receiver deceived? On one hand, one might argue that he is not because he understands the sender's strategy and responds optimally. On the other hand for almost all states for which the message "the state is 1" is sent he does not take an *ex post* optimal action. Finally, and this is a bit more subtle, types close to the lowest type in the interval $[\theta_n, 1]$, that is types in the interval $[\theta_n, \theta_n + \epsilon]$ for sufficiently small ϵ, would make the receiver better off by switching to the (equilibrium) message that is sent by types in the second highest interval $[\theta_{n-1}, \theta_n]$. Blume and Board [16] say that the sender is "intentionally vague" when, as in this example, his equilibrium message differs from the message the receiver would have preferred the sender to use, given the receiver's equilibrium strategy. Whenever the sender is intentionally vague and in addition the equilibrium message he uses is inconsistent with its prior meaning (e.g. all types in the interval $[\theta_n, 1]$ send the common message "the state is 1"), it is hard to argue that he is not lying and deceiving the receiver.

Kartik, Ottaviani and Squintani [59] (KOS) take a step further away from cheap talk in sender-receiver games. In addition to assigning prior meanings to messages, they modify the CS model by introducing lying costs for senders and receiver naiveté. Once messages have preexisting meanings both lying and naiveté can be expressed by referring to that meaning (similar to Crawford [27], Cai and Wang [21] and Wang, Spezio and Camerer [100]). The message space in KOS's model is identical to the state space, a naive receiver interprets a message t as the (true) statement "my type is t" and a sender for whom it is costly to misrepresent the truth bears a lying cost that is increasing in the distance between the true type t and his message m.

With upwardly biased senders and a type space that is unbounded above, KOS establish existence of separating equilibria in three environments: communication with a pool of receivers some of whom are partially naive, communication with a single receiver who may or may not be partially naive, and communication by a sender for whom lying is costly. In equilibrium there is "language inflation." Senders send messages whose literal meaning indicates a higher type than their true type. Since the equilibrium is separating, sophisticated receivers are able to invert the sender message and learn the sender's true type. Naive receiver's are deceived. Separation is achieved by senders balancing the costs and benefits from exaggeration. With explicit convex lying costs and sophisticated receivers, the gain from sending a message higher than the already inflated equilibrium message would be to move the receiver action closer to the sender's ideal point. The loss would be the increase in lying costs.

2.5 Equilibrium selection

In sender-receiver games talk is cheap. An implication is that, unlike in other signaling games, the ability to send messages does not vary with the sender's type and cannot be used to differentiate types. Farrell [38] reminds us that even though sender-receiver games comprise only a tiny subset of all signaling games they are of special interest because of the importance of language in human interaction. Talk being cheap can be thought of as capturing universal command of a common language.

Sender-receiver games pose a special challenge for game theory. They generally have a great number of distinct equilibria. Furthermore, these equilibria tend to be robust to standard refinements. There are two types of equilibrium multiplicity in sender-receiver games. One is outcome multiplicity: if there is not too much conflict of interest there will be multiple outcomes – joint distributions over actions and types – that can be supported by equilibria. This is easiest to see when players' interests are perfectly aligned. Then there will be a separating equilibrium, where each type sends a distinct message, messages perfectly identify types and the receiver takes actions that maximize the common payoff in every state of the world. At the same time, regardless of incentives, there is always a pooling equilibrium in which the receiver takes the same action after all equilibrium messages. The second type of equilibrium multiplicity in sender-receiver games is permutation multiplicity. Given any equilibrium, one can obtain another equilibrium by simply permuting messages. "Yes" and "no" are perfectly exchangeable.

Refinements that operate by placing restrictions on beliefs following out-of-equilibrium messages are ineffective because for any equilibrium there is an outcome-equivalent equilibrium that uses all messages. Furthermore, even if one restricts attention to pure-strategy equilibria, it is always possible to replace the belief following an unused message by the belief following a used message and have the receiver respond to the unused message in the same way he responds to the used message.

Farrell attempts to overcome these problems by acknowledging the fact that communication happens in the context of an existing language. To this end, he identifies two constraints that might limit successful communication. One is the availability of messages with a clear meaning and the other is credibility of messages. He points out that interlocutors tend to have access to a shared rich common language. This leads him to assume away any problems due to meaning and to concentrate on credibility of messages: all meanings can be conveyed, but need not be believed. Credibility is the only remaining barrier to communication.

Formally, the assumption that there is a shared rich language is captured by requiring that for every equilibrium and for every set of types $X \subseteq T$ there is an unused message m_X with literal meaning "my type is in X." Any such message is called a "neologism." Suppose that X is exactly the set of types who would benefit from the neologism m_X being believed by the receiver, rather than getting their equilibrium actions. Then, Farrell argues, the receiver should believe m_X and, in his terminology, m_X is a "credible neologism." Any equilibrium for which there is a credible neologism, we should reject; it is not "neologism proof."

The power of this argument can be seen in the game with the payoff structure indicated in Figure 3.

	a_1	a_2	a_3
t_1	3,3	0,0	2,2
t_2	0,0	3,3	2,2

Figure 3: Common interest with a unique pooling action

If types are equally likely, then there is a (pooling) equilibrium in which the receiver takes action a_3 regardless of the sender's message. This equilibrium may strike one as implausible given that interests are perfectly aligned. Both players would receive their maximal payoff in every state of the world if the sender truthfully revealed his type and the receiver acted accordingly. Now, given this pooling equilibrium, consider the neologism "my type is t_1." Only type t_1 would benefit from having this neologism be believed. Therefore for the pooling equilibrium the neologism is credible. Type t_1 has an incentive to use it and the receiver has reason to believe it. That, according to Farrell, breaks the pooling equilibrium.

At the same time in this example separating equilibria, in which types t_1 and t_2 send distinct messages, survive the neologism-proofness test. Given that each type attains his maximal payoff *in* equilibrium no type (or set of types) could be made better off by deviating from equilibrium and sending a neologism.

Neologism proofness presents a variety of stimulating opportunities and challenges for experimenters. The most evident one is that it makes predictions about equilibrium outcomes, which can be checked in the lab. One of the flaws of neologism proofness is that there are games in which neologism-proof equilibria fail to exist; experiments may give pointers toward amending theory to make predictions in those games. Another flaw of neologism proofness is that it is silent about how messages will be used, even for the equilibria it selects; it only makes predictions about joint distributions over types and actions, and not over joint distributions over types, actions and messages. Experiments on sender-receiver games inevitably yield data about message use, with implications for how to think about equilibrium selection. Finally, the experimental laboratory forces one to make choices about the message space, some of which may give rise to implausible equilibria that leave no room for introducing neologisms.

	a_1	a_2	a_3
t_1	3,3	0,0	2,2
t_2	1,0	-1,3	2,2

Figure 4: Mimic

The game in Figure 4 (with a uniform type distribution) has a unique equilibrium outcome. The receiver takes action a_3 for every type realization. There is however a credible neologism with literal meaning "my type is t_1." Type t_1 prefers the neologism to be believed, in which case his payoff is 3, as opposed to his equilibrium payoff of 2. Type t_2 obtains his maximal payoff of 2 in equilibrium and thus would not benefit from having the neologism believed. Experiments can give us insights about whether the unique equilibrium outcome prevails or is undermined by type t_1 sending self-signaling messages.

Farrell draws our attention to the role of language in sender-receiver games. An attractive feature of the experimental laboratory is that we are able to manipulate the language available to players. A natural question that arises is what outcomes we expect if players have access to messages with literal meanings but the available language falls far short of Farrell's rich language requirement.

Consider the "I won't tell" game with the payoff structure in Figure 5. Even though there is a separating equilibrium, the sender would prefer his type not to be revealed. One

	a_1	a_2	a_3
t_1	1,3	0,0	2,2
t_2	0,0	1,3	2,2

Figure 5: I won't tell

suspects that the pooling equilibrium in which the sender sends a message to the effect of stating "I won't tell" is a more likely outcome than separation. Indeed, given a separating outcome the neologism "I won't tell" is credible, and therefore the separating equilibrium fails to be neologism proof. Furthermore, given the pooling equilibrium outcome there is no credible neologism. Thus neologism proofness makes a sensible prediction. But does it do so for the right reason? We can imagine an experimental environment in which senders have access to only two messages, one with literal meaning "my type is t_1" and another with literal meaning "I won't tell." Now there is a separating equilibrium in which type t_1 sends the message with literal meaning "my type is t_1" and type t_2 sends the message with literal meaning "I won't tell." Given the impoverished language, there are no unused messages and therefore no neologisms, credible or otherwise. Neologism proofness does not reject the separating equilibrium in this environment and yet, we suspect, that the pooling equilibrium in which the sender sends the message with literal meaning "I won't tell" has a better chance of predicting outcomes than the separating equilibrium described above. The experimental laboratory lets us explore these suspicions.

Related to the last observation is the fact that while neologism proofness selects among equilibrium outcomes, it does not select among equilibria. In the common-interest game in Figure 3 it does select the separating equilibrium outcome, but it does not reject the equilibrium in which this outcome is supported by an equilibrium in which type t_1 sends a message with literal meaning "my type is t_2" and type t_2 sends a message with literal meaning "my type is t_1." This is a consequence of neologism proofness only making use of the focal meanings of messages out of equilibrium, while saying nothing about how focal meaning might impact the use of messages in equilibrium. Blume [9] proposes a refinement criterion, perturbed message persistence (PMP), for (finite) sender-receiver games that does make a connection to an underlying language. That language is given in the form a of a completely mixed strategy of the sender that pins down the meaning of messages the sender does not use intentionally. This completely mixed strategy perturbs the original game. PMP applies Kalai and Samet's [57] persistence criterion to this perturbed game. In the common-interest game of Figure 3 PMP selects separation, with the sender using messages consistent with their preestablished meanings.

In the Mimic game of Figure 4 there is no equilibrium that would pass the neologism-proofness test. Therefore neologism proofness makes no prediction in this game. This is not an isolated phenomenon. In Crawford-Sobel games, whenever there is an influential equilibrium (an equilibrium that induces more than one equilibrium action) neologism proofness rejects all equilibria. In contrast, the NITS criterion of Chen, Kartik and Sobel [23], which rejects equilibria for which the lowest type would rather identify himself, selects the most influential equilibrium under general conditions.

2.6 Alternative communication protocols

In most of the theoretical work discussed thus far the communication protocol is simple and fixed. In the basic sender-receiver game there is one round of communication in which a single sender sends a message after which a single receiver takes an action. One can easily think of a plethora of alternative protocols in which multiple messages may be sent by both players and third parties can act as mediators. This raises the question of whether we can use economic theory to design better communication protocols and what an optimal communication protocol might look like.

For the uniform-quadratic CS model this question is addressed by Goltsman, Hörner, Pavlov and Squintani [48] (GHPS). By the revelation principle (Myerson [78]) any equilibrium outcome (joint distribution of types and actions) that can be achieved with some communication protocol can also be achieved with a mediator who takes sender messages as inputs and makes action recommendations to the receiver. GHPS establish an efficiency bound for mediated communication. When conflict is low ($b \in (0, 1/8)$) this bound can be attained by a communication protocol with two rounds of face-to-face communication, which is studied by Krishna and Morgan [66]. For low and moderate conflict ($b \in (0, 1/4)$) the GHPS bound can be attained via communication through a very simple noisy channel with replacement noise that is examined by Blume, Board and Kawamura [14]: with some probability the sender's message is replaced by a randomly chosen message and the receiver does not know whether he observes the original message or the replacement.

The Mimic game in Figure 4 can serve as a simple illustration of how communication through a simple noisy channel has the potential to improve communication, how one might think about trying to explore this potential in the experimental laboratory, and why thus far there is no experimental work that speaks directly to this issue.

The unique equilibrium outcome in Mimic (with one round of direct (unmediated) communication) has the receiver always take the pooling action a_3 in every equilibrium. There is no equilibrium in which the sender has the ability to influence the receiver's choice of action. Now, following Myerson [80], suppose that the sender communicates via sending a messenger pigeon. The message space has two elements, "send the pigeon" and "do not send the pigeon." The pigeon, once sent, arrives with probability 1/2, and therefore the communication channel is faulty. In this game there is an equilibrium in which type t_1 sends the pigeon and t_2 does not send the pigeon. Clearly, if the pigeon arrives the receiver learns that it was sent by type t_1 and takes action a_1. If no pigeon arrives, there are two possibilities. Either the pigeon was not sent or it was lost on the way. A simple calculation shows that, given the sender's strategy, if the pigeon does not arrive, the posterior probability of type t_2 is 2/3. Therefore, if no pigeon arrives the receiver is indifferent between actions a_2 and a_3, and a_3 is an optimal response. Type t_1 strictly prefers sending the pigeon and type t_2 strictly prefers not sending the pigeon. Hence we have an equilibrium and in this equilibrium the *ex ante* payoff of both players exceeds the payoff from the unique equilibrium outcome in the original Mimic game with a perfectly reliable communication channel.

The simplicity of the Mimic game with a faulty channel would appear to make it a perfect candidate for taking it to the lab: give the sender two messages, A and B; whenever the sender sends message B transmit that message to the receiver; whenever the sender sends message A faithfully transmit that message to the receiver with probability $1/2$ and otherwise replace it with message B (without informing the receiver that a replacement took place); and, publicly explain this procedure to the subjects. What might give one pause, however, is what we already know about communication in sender-receiver games in general and in games with an incentive structure like Mimic (type t_1 likes to be identified; type t_2 prefers mimicking t_1 to being identified; and, t_2's preferred action is pooling). As we will see, a persistent theme that emerges in the experimental literature on single sender-single receiver games is over-communication and for the incentive structure in question Blume, DeJong, Kim and Sprinkle [12] find that in the experimental laboratory a non-equilibrium outcome, in which type t_1 identifies himself successfully is remarkably stable when the communication channel is perfectly reliable.[1] This leaves little room for a noisy channel to improve communication.

The closest the literature comes to identifying a positive role for faulty communication channels is in examining "randomized response" (Warner [102]). Randomized response is a technique meant to alleviate privacy concerns when conducting surveys about sensitive topics, such as drug use, tax avoidance, sexual preferences, employee theft, etc. The idea is to have respondents condition their answer on a privately controlled randomizing device, rather than answering directly. For example, in a study about employee theft each employee may be given a die to roll privately and be instructed to always say "yes" if the outcome of the die roll is 6 and otherwise to respond truthfully to the question "Have you ever stolen from your employer?" Intuitively, the use of the private die roll provides some privacy protection, because if the respondent follows instructions a "yes" answer is no longer an admission of having engaged in theft. This is analogous to the possibility of the pigeon getting lost in Mimic with a faulty channel; even though only type t_2 fails to send the pigeon, if it can get lost, non-arrival of the pigeon does not prove to the receiver that he is dealing with type t_2.

Having the respondent control the randomizing device under the randomized response technique introduces incentive issues that are not present with communication through an exogenously operating faulty channel. When communicating through an external noisy channel the sender loses part of his control over the message that is received. Under the randomized response technique, the respondent (who corresponds to the sender) privately observes a random signal but has full control over which message is received. To get around this requires assuming that respondents have a minimal preference for following instructions or truth-telling. Blume, Lai and Lim [17] carry out the first complete analysis of randomized response incentives in a fully specified game and implement that game in the experimental

[1] Focus on types t_1 and t_2 in their Game 4, p.90 and p.100. Type t_3 in their game can be ignored because the sender is content to identify that type and have it not be mimicked.

laboratory.

3 Experiments with sender-receiver games

3.1 Common interests

Common-interest games offer a natural starting point for examining the impact of barriers to communication other than misaligned incentives. This includes departures from worlds with a rich common language.

3.1.1 Emergence of referential meaning

Blume, DeJong, Kim and Sprinkle [10] (BDKS98) experimentally examine the emergence of meaning in a 2 type - 2 action - 2 message variant of the game considered by Lewis [70]. They focus on the most extreme situations, with a complete absence of a shared language at the outset. Wärneryd [101] applies evolutionary game theory to the problem of emergence of meanings in sender-receiver games. He shows that in simple common-interest games of the type considered by BDKS98, only efficient equilibria are evolutionarily stable. Blume, Kim and Sobel [8] extend this approach to more general incentive structures and allow players to have fixed roles, which fits the environment of BDKS98.

In the simplest case examined by BDKS98 a population of six senders and six receivers (with fixed role assignments) are repeatedly randomly matched over twenty periods to play a sender-receiver game with the following payoff structure (where 700 corresponds to a 70% chance of earning $1.00 under the Roth-Malouf procedure [85]):

	a_1	a_2
t_1	0,0	700,700
t_2	700,700	0,0

Types are equally likely and in each period drawn independently for each sender. The available messages are A and B. All players are informed about which sender types were drawn and which messages they sent in prior rounds. While there might be a focal association of message A with type t_1, there is another potential focal association of message A with action a_1 and since the positive payoffs are on the off-diagonal these two focal associations are in conflict. This suggests that initially players may find it difficult to communicate. Indeed in the first period, averaged over 15 replications, the frequency of the separating outcome is 48% and thus indistinguishable from the no-communication outcome. Over time, however, the separating proportion steadily increases: 65% in period 5, 74% in period 10, 88% in period 15 and 95% in period 20. Furthermore, 11 of the 15 replications achieve perfect separation in the last period.

Thus, participants in the experiment manage to establish meanings over time from a starting point where messages are meaningless. This is not accounted for by (Bayesian) Nash equilibrium, belief-based refinements of Nash equilibrium, refinements of Nash equilibrium in the tradition of strategic stability or refinements tailored to cheap-talk games that appeal to a pre-existing rich language (à la neologism proofness and perturbed message persistence). At the same time it is consistent with various learning and evolutionary approaches.

It is worth noting that at the intermediate stages, before the population arrives at a convention that assigns distinct meanings to messages, there is partial communication success. The population does better than without communication but not as well as with full separation. This does not correspond to any of the equilibria of the game, and yet may frequently be a good description of the situation faced within newly formed teams and by cross-cultural interlocutors. Meaningful communication is established gradually without strong (common) knowledge requirements and in the interim there may be some communication even without a convention in the strict sense.

In addition to studying the case where focal meaning is absent because of conflict among possible focal points, BDKS98 consider the case where all possibilities of initial focal meanings have been removed. This is achieved by giving each participant his own private representation of the message space. As a result, all messages are initially symmetric and symmetry has to be broken for communication to become possible. Here, aggregated over three replications, there is no communication initially, communication improves gradually and the separating outcome is achieved 100% of the time in the final (20th) period.

The fact that learning is gradual shows that populations do not engage in optimal symmetry breaking. Typically, symmetries can be broken very early and in principle, since population information is available, a population could achieve efficient communication by a rule like "have type t_1 send the message most frequently sent by type t_1 thus far, and have type t_2 send the other message." This suggests that there are situations in which newly formed teams that need to establish a code can benefit from being instructed about general principles for using precedents to arrive at efficient codes.

Given that there is a strong tendency for a common language to emerge in the simplest case with two types, two actions, two message and history information about past play, it makes sense to make the task more challenging. There are many ways in which this can be done. BDKS98 look into adding a safe action, making separation unattractive for senders while preserving it as an equilibrium, adding an additional message, and removing history information. Adding a third safe action for the receiver has no impact on the tendency toward full separation; making the separating equilibrium unattractive for senders reduces the tendency toward separation, even though in some replications the terminal outcome remains separation. The pooling outcome prevails when the separating equilibrium is unattractive for senders and the basin of attraction of the separating equilibrium is small. There is also a strong dampening of the tendency toward achieving separation in the 20th round when history information is removed.

Bruner, O'Connor, Rubin, and Huttegger [18] (BORH) continue the study of the emergence of linguistic conventions in an experimental laboratory setting that is inspired by

Lewis [70]. All of their games are of common interest; there is a positive payoff if and only if the receiver correctly identifies the state. They provide additional evidence for the case where players are not given history information, allow for non-uniform distributions over the state space and consider games with three states. They consider the $2 \times 2 \times 2$-game (two states, two messages and two actions) with a uniform prior distribution over the state space, the $2 \times 2 \times 2$-game with non-uniform distributions and the $3 \times 3 \times 3$-game with a uniform distribution.

Other than the different games used, BORH also adopt somewhat different experimental procedures than BDKS98. While continuing to adhere to fixed roles and random matching, they have subjects interact more frequently, either over 60 rounds (for the three-state game) or 120 rounds (for the two-state games, with 60 rounds each for different state distributions). For the two-state games, they do not reveal to the subjects the actual or the possible state distributions so as to minimize the chance that the subjects develop strategies based on unequal priors rather than through experience. They also depart from BDKS98 regarding the end-of-round information feedback. No history feedback, either at the population or the individual level, is provided; the information provided concerns only the choices made within each pair and in the current round.

The more frequent interaction may explain why in the $2 \times 2 \times 2$-game with a uniform distribution they observe a more robust tendency toward separation than BDKS98 (even though complete separation is rarely achieved). The final frequency of separation in the 60th period of the $2 \times 2 \times 2$-game with a uniform distribution, averaged over four sessions, is about 80%. Separation is also achieved some of the time in the other treatments, albeit less frequently. In the $2 \times 2 \times 2$ game with a strongly non-uniform distribution separation does not emerge in four out of eight sessions. In the $3 \times 3 \times 3$-games separation is achieved in only three out of ten sessions. At the same time there is a general tendency for information transmission to improve over the time. Results tend to be "messy" and the observed communication without separation is not the result of convergence to a partial pooling equilibrium. Overall, removing history information does not prevent the emergence of meaning in simple environments. This finding suggests that increasing the number of interactions may serve as a substitute for providing population history in achieving the emergence of meaning. The substitutability is, however, not perfect. Increasing the complexity of the game makes the emergence of meaning more difficult.

3.1.2 Reaching shared understandings with natural language

The investigation of the emergence of linguistic conventions in BDKS98 and BORH is conducted in an extreme environment devoid of any pre-existing language. More common are situations where agents have access to a natural language, but that language is imperfectly shared. This includes examples where language needs to adapt to new technologies, new social circumstances, and contacts across cultural, occupational, or disciplinary boundaries. How shared understandings can be reached with an imperfectly shared language is the topic of early research by Krauss and Weinheimer [62] [63].

Krauss and Weinheimer [62] study the change over time of references phrases used to describe novel objects. Two subjects interact over multiple rounds. In each round each subject is given a set of six cards each of which displays a different arrangement of six non-standard figures. One subject's cards are labeled 1 through 6 and the other player's cards are labeled A through F. The task is to identify the correct match of numbers and letters. The labels are randomly assigned so that, e.g., matching A with 1 does not necessarily result in a correct match. Subjects engage in free-from communication without visual contact, using natural language (in recent work, Charness and Dufwenberg [20] find an important psychological impact of using free-form natural language communication, as opposed to prefabricated messages; Crawford [30] advocates attempting to incorporate richer communication, in the spirit of unrestricted natural-language messages, into formal models). The authors find that infrequently mentioned figures are referred to with relatively long phrases, whereas frequently mentioned figures are referred to with short (and frequently one-word) phrases. In addition, the length of references phrases drops over time. In one example a figure initially referred to as an "upside-down martini glass in a wire stand" eventually becomes simply "martini."

In Krauss and Weinheimer [63] the authors follow up on their earlier paper [62] by comparing the effects of two-sided and one-sided communication. While the subjects in Krauss and Weinheimer [62] perform their card-matching tasks by engaging in unrestricted conversation, the subjects in Krauss and Weinheimer [63] are divided into two treatments that differ by whether the "listener" in a group is allowed also to talk to the "speaker." The authors find that, relative to the case of one-sided communication, the reference phrases are shorter with two-sided communication.

These findings suggest that natural language sometimes needs to be adapted to the particular communication environment and genuine conversation appears to better facilitate the adaptation than do monologues. These two studies suggest that focusing on monologues, as is the case in most work on sender-receiver games to date, and thus ignoring the role played by clarifications, may limit our understanding of how meaningful communication evolves and may underestimate the capacity for developing a common language.

In a more recent study Weber and Camerer [103] revisit the phenomenon of shared meaning emerging in the presence of natural language. They are motivated by the question of whether cultural conflict in organizations may be responsible for merger failures. Four subjects participate in multiple rounds of picture identifications. Every subject is presented with the same set of 16 unique pictures depicting different office environments. In the first 20 rounds, the four subjects are assigned into two fixed groups. The two subjects in each group alternate between the roles of sender and receiver. In each round, the experimenter indicates to the sender 8 of the 16 pictures in a specific order. The sender then describes the pictures to the receiver using free-form communication. The two subjects in a group have identical payoffs. The payoffs depend on how quickly the receiver identifies the correct figures in the correct order.

After 20 rounds, the two groups are merged. They are randomly assigned the designations of acquiring group and acquired group. One subject in the acquired group is selected to join the acquiring group as a receiver, and one subject in the acquiring group is selected to be the sender. The sender communicates to the other existing subject in the group and to the acquired receiver. There are thus one sender and two receivers in the merged group. The sender communicates with the two receivers simultaneously and publicly about the selected picture. The receivers perform their identifications individually with independent rewards. The sender receives the average of the two receivers' rewards. This second part of the experiment lasts for another 10 rounds, where role alternations are no longer used.

Similar to Krauss and Weinheimer [62], Weber and Camerer [103] find that task-completion time decreases with experience. Groups develop a shared language that allows them to quickly refer to the pictures. For instance, a picture is initially referred to in a group as "The one with three people: two men and one woman. The woman is sitting on the left. They're all looking at two computers that look like they have some PowerPoint graphs or charts. The two men are wearing ties and the woman has short, blond hair. One guy is pointing at one of the charts." The description is eventually condensed to simply "PowerPoint."

The succinct descriptions that develop within groups vary considerably across groups. While these descriptions facilitate efficient communication in a group, they can make it difficult to establish a new shared language: the task-completion time increases once the groups are merged. For example, in one case a picture is referred to in a group as "Uday Rao," because a person in the picture resembles a professor by that name who taught the two students in the group. After the merger, the acquired receiver has no idea who Uday Rao is, and a different shared language has to be developed. While the completion time decreases with experience in the merged group, it never returns to the pre-merger level. This study highlights an important aspect of the emergence of common language that has heretofore be neglected in the literature: there may be a tradeoff between the efficiencies of the emergence of shared meanings within groups and across groups. If the convention established in the pre-merger groups were constrained to be less idiosyncratic and perhaps less effective, the language might be easier for the acquired receivers to pick up.

3.1.3 Semiotics

In BDKS98 and BORH there is a preset collection of symbols that can be used as messages. In contrast, experimental semiotics is concerned with the evolution of the symbols themselves. Healey, Swoboda, Umata, and Katagiri [54] conduct experiments in which subjects privately listen to music and then use drawings to communicate. Each member of a pair produces a drawing representing his or her respective piece of music. The drawings can be of anything but cannot contain letters or numbers. Communicating through their drawings, the subjects then decide whether they share the same music. It is found that drawings with higher levels of abstraction are more effective in achieving successful communication. Interactive communication, in which subjects simultaneously draw on a shared medium and can modify each other's drawings, appears to favor the use of abstraction, compared to unidirectional communication. Using a similar design, Garrod, Fay, Lee, Oberlander, and MacLeod [43] further document that interactive communication is a primary factor for graphical communication to evolve from being iconic to symbolic.

3.1.4 Grammar and pragmatic inference

The studies of emergence of meaning of BDKS98 and BOHR are concerned with a very basic conception of meaning as reference, much like the Augustinian one that Wittgenstein [104] criticizes. Two directions of departure from their setup, in which subjects learn to attach simple unitary symbols to equally simple unitary types, are to try to incorporate rudimentary versions of grammar and pragmatic inference. In both cases the form of language is governed by efficiency considerations. Rubinstein [86] and Blume [11] discuss how the structure of language facilitates language acquisition.

Selten and Warglien [89] experimentally implement sender-receiver games to investigate the emergence of grammatical structures. Like in BDKS98 and BORH there is no common language at the outset. Unlike in BDKS98 and BORH states and messages can be complex. States are figures composed of geometric shapes, which may have inserts and may differ in color. Messages are strings of letters from a limited fixed repertoire. Senders and receivers interact in fixed pairings for 60 periods. State spaces and message spaces change every 10 rounds. In each period both players in a pair are shown the same set of figures. Each player decides on a code, an assignment of a message to each figure in the set of figures. Following the choice of codes a figure is randomly chosen. Both players receive a positive payoff if their codes assign the same message to the randomly chosen figure; otherwise they receive zero. The sender, in addition, has to pay a small cost for each letter used (thus departing from BDKS98 and BORH). The players are informed whenever there is a communication success. If instead communication fails, i.e., the codes differ for the figure at hand, both are informed about their partner's message for that figure.

Note that unlike in BDKS98 and BORH, here the sender does not have private information before making a choice and the receiver does not observe the sender's message before making a choice. Instead, in the spirit of the strategy method (see Selten [88]), in each period both players pick partially specified strategies for the stage game, which are then automatically executed. The sender's stage-game strategy amounts to an encoding rule and and the receiver's to a partial decoding rule (partial because the receiver's decoding rule may not assign a figure to every possible message).

The use of fixed pairings departs from BDKS98 and BORH, who employ repeated random matching in a population of players. Fixed pairings should facilitate the emergence of a common code, a common assignment of messages to those figures which have been encountered before.

In linguistics and philosophy of language, 'compositionality' captures the idea that the meaning of a complex expression can be inferred from the structure of the expression and the meanings of its constituents; this principle is sometimes attributed to Frege [42]. In the present setting a simple compositional code assigns a letter to each possible part of a figure and encodes that figure by a concatenation of those letters. Blume [13] refers to such codes as 'modular' and shows that in certain environments modular/compositional codes can be learned with a minimal number of observations. Learnability of compositional codes is closely related to their productivity, the ability to generate novel meaningful expressions and the ability to understand complex expressions that have never been encountered before.

Selten and Warglien [89] find that in 42% of cases a common code emerges and that in 12% of cases the code is compositional. A larger repertoire of letters facilitates the emergence of common codes; this is the case despite the fact that in principle a binary code would suffice. Role asymmetry, in which one subject acts as a leader in devising code and the other tries to adjust, instead of both subjects adjusting simultaneously, enhances communicative success. Compositional codes turn out to be extendable, i.e., when new figures are introduced the figures that were already available before are denoted by the same messages in the new code as in the old one. Environments with novelty – in the experiment players see a sequence of new figures each of which appears only once – favor compositionality. This is in line with arguments in favor compositionality that appeal to the productivity and systematicity of language.

Hong, Lim, and Zhao [55] experimentally investigate the emergence of compositional grammars in a common-interest sender-receiver game. A major difference of their experimental design from that of Selten and Warglien [89] is that messages are costless and there is no constraint on message length.

De Jaegher, Rosenkranz, and Weitzel [34] (DRW) use sender-receiver games in an experimental investigation of principles of pragmatic inference. Pragmatics is concerned with the rules that govern how interlocutors discern meaning that is implied by an utterance beyond what is merely said. Pragmatic inference leverages the context of an utterance, including what interlocutors believe about each other's rationality.

DRW study sender-receiver games with two states, two actions and the option to either send a message or to abstain from doing so. Actions are guesses of the state; the message has no preestablished meaning; and, players have a common preference for guesses being correct. DRW use a between-subject design, where each subject participates in four selected treatments. For each treatment, subjects take part in 20 rounds of the stage game in fixed roles and under a fixed-matching protocol. At the end of each round, information feedback on the selected state, the sender's message decision, and the receiver's guess is provided.

DRW consider both the case where the sender bears a small cost when sending a message (Costly Signaling) and the case where sending a message is costless (Cheap Talk). The cost is small enough that it is worth sending the message if that leads to a correct guess, when not sending message would result in an incorrect guess. Importantly, one state is more likely than the other. There are three equilibrium outcomes, pooling, separation with the message indicating the more likely state and separation with the message indicating the less likely state. This is the simplest setup in which to investigate "Horn's rule" [56] that marked (more unusual) events are referred to with marked (more elaborate and irregular) expressions, whereas unmarked events are referred to with unmarked expressions. Horn's rule is inspired by Grice's [51] work on pragmatic inference, which assumes rational agents who economize on communication effort. Matching the language of Horn's rule to the experimental design, the marked event corresponds to the infrequent state, the unmarked event to the frequent state, the marked expression to sending a message and the unmarked expression to not sending a message. In the present instance, Horn's rule suggests that in the costly signaling case the sender will send a message to indicate the less likely state. Thus, Horn's rule selects the efficient separating equilibrium. This prediction is confirmed by the experimental finding that efficient separation obtains in the majority of instances the game is played. This is also true for the cheap talk version of the game, although to a lesser degree. DRW's results contrast with De Saussure's [35] view that meanings are purely conventional.

3.1.5 Key findings from experiments with common-interest games

The experimental literature on strategic information transmission in common-interest games shows that effective communication can emerge when there is no shared natural language. Subjects can learn to attach meanings to initially meaningless messages. This process is facilitated by making population history available, but also succeeds otherwise, at least over a long enough horizon.

Even with access to natural language there is no guarantee of immediate shared understanding. Over time, however, subjects learn to label objects and to economize on shared descriptions. Such common understanding is more easily reached with genuine conversation than one-sided monologues. It can be disrupted by changing conversation partners.

Elementary 'grammars,' in the form of simple compositional codes, can emerge in the lab, these codes are extendable and environments with novelty favor the emergence of compositional codes. There is also evidence in support of simple rules of pragmatic inference.

3.2 Imperfect incentive alignment

The bulk of experimental investigations of strategic information transmission with imperfect incentive alignment focuses on games with a single sender and a single receiver. Earlier studies (e.g., Dickhaut, McCabe, and Mukherji [36], and Forsythe, Lundholm, and Rietz [39]) are concerned with testing key qualitative features of the equilibrium prediction of Crawford and Sobel [25]. Some of the more recent work is interested in non-equilibrium behavior and bounded rationality ideas such as level-k reasoning and Quantal Response Equilibrium (Cai and Wang [21]; Kawagoe and Takizawa [60]; Wang, Spezio, and Camerer [100]) and evolutionary learning dynamics (Blume, Dejong, Kim, and Sprinkle [12]). The equilibrium selection issue for cheap-talk games is explored in Kawagoe and Takizawa [60] and de Groot Ruiz, Offerman, and Onderstal [32]. Lai, Lim, and Wang [68] and Vespa and Wilson [99] study multiple-sender multidimensional cheap talk inspired by Battaglini's [5] fully revealing equilibrium. Minozzi and Woon [76] experiment on games with two senders and unidimensional state spaces. Battaglini and Makarov [6] consider multiple receivers and test the predictions of Farrell and Gibbons' [37] model.

Several recent studies explore the role of vague languages as an efficiency-enhancing device. Serra-Garcia, van Damme, and Potters [90] investigate how vague language interacts with intrinsic preferences for truth-telling and affects the extent of information transmission in a two-player sequential public-good game. In an asymmetric coordination game, Agranov and Schotter [2] show that a utilitarian sender uses vague languages to limit receivers' information about payoffs and to conceal conflict that would hinder the use of focal points for coordination. Wood [105] studies the issue in a more standard communication game environment.

3.2.1 Single sender – single receiver

The first test of the central theoretical predictions from Crawford and Sobel [25] (CS) is conducted by Dickhaut, McCabe, and Mukherji [36] (DMM). Inspired by CS, they generate the payoffs for their experiment using a parameter $b > 0$ that measures the misalignment of preferences between sender and receiver. For the experiment, state and action spaces are discretized; they are identical and have four elements. The message space is the collection of all subsets of the state space that satisfy a form of convexity: for any two states included in a subset, the subset also includes any state in-between. In each of twelve rounds, four senders are repeatedly and randomly matched with four receivers. Player roles are fixed. Players receive personal history information about the frequency of types drawn and the messages sent.

DMM find that greater preference alignment, a lower value of b, corresponds to more information transmission, a better match of actions to states and higher receiver payoffs. Support for the equilibrium predictions is mixed. For a low degree of preference misalignment, multiple equilibria are consistent with the data; for higher degrees, observed behavior cannot be rationalized by any equilibrium.

DMM do not report details on message use. In their discussion of theoretical predictions they only list a subset of monotonic sender strategies, those where the sender's type is included in the message – one may call this truth consistent. When listing sender strategies that are compatible with the data, they again report only truth-consistent strategies. While these strategies are sensible, like Crawford and Sobel's focus on *ex-ante* efficient equilibria, they are not singled out by the theory.

Forsythe, Lundholm, and Rietz [39] (FLR) study strategic information transmission in a lemons market environment (Akerlof [3]). A seller who is privately informed about the quality of an asset he owns interacts with a potential buyer in a market. Asset values have the property that if the buyer knows the true quality, trade is beneficial for both parties. The quality can be low, medium, or high, and the uninformed buyer makes a price bid to the privately informed seller who then decides whether to accept or reject the bid. Preferences are so misaligned that the market generates the adverse selection outcome, in which trade does not occur because of the buyer's skepticism. In their baseline experimental treatment in which players do not engage in cheap-talk communication, theory predicts that the market fails. In their second treatment in which the seller can engage in cheap-talk, again due to the large conflict of interests, theory predicts that no information can be transmitted via cheap-talk.[2] The set of messages available to subjects is {low, medium, high, null}. In each

[2] Forsythe, Lundholm and Rietz [39] also consider a treatment in which the seller can disclose information

of the twenty two rounds, each subject is randomly paired with each other subject twice, once as a seller and once as a buyer. Subjects alternate between being a seller and a buyer each period.

The experimental result shows that the outcome from the cheap-talk treatment is considerably different from the baseline treatment in which players do not engage in cheap-talk communication. Although this observation cannot be rationalized by equilibrium, it is consistent with the main finding from DMM. FLR take a more careful look at the individual behavior to identify where the discrepancy between the theoretical prediction and their experimental result comes from. When the seller can send cheap-talk messages, buyers seem gullible and are frequently taken in by the seller's over-optimistic statements and bid too much for the asset. As a result, a higher transaction rate is reported, which generates an efficiency gain relative to the baseline treatment. However, the higher transaction rate mainly results from the high bids, often higher than their true asset values, made by gullible buyers rather than from an efficient and truthful transmission of information.

Blume, Dejong, Kim, and Sprinkle [12] (BDKS01) consider a richer set of incentive structures and compare behavior when messages have preassigned meanings with behavior when meaning needs to emerge. They examine simple incentive structures in which partial information revelation is plausible. There are three states of the world and five actions. Three of the actions are the respective receiver best replies conditional on learning the state, one action is the receiver best reply conditional on learning that either the first or second state obtains, and the remaining action is the best response conditional on prior information. Messages are either meaningless or the numbers $1, \ldots, 5$, and thus have focal meanings corresponding to the five receiver actions, which are labeled $R1, \ldots, R5$. Six senders are repeatedly and randomly matched with six receivers. There are twenty initial rounds with a common-interest game common to all treatments, followed by forty rounds for each of the three treatments (Games 2, 3 and 4). Player roles are fixed. Players receive history information about the behavior of senders at the population level which consists of the frequency of types drawn and the messages they sent.

BDKS01 test a non-equilibrium prediction of behavior. An outcome satisfies the partial common interest (*PCI*) condition if (1) it partitions the state space so that types prefer actions corresponding to "their" partition element to other best replies, (2) different partition elements do not share best replies, and (3) there is no finer partition with that property. They consider four incentive conditions. In Game 1, which is always played for the first twenty rounds, there is a separating equilibrium that is preferred by all three types and *PCI* predicts full separation. In Game 2, the fully separating equilibrium is preferred by only two types. In Game 3, the fully separating equilibrium is preferred by only one type. In Game 4, there is no fully separating equilibrium. In Games 2, 3 and 4, *PCI* predicts partial separation of type 3 from types 1 and 2.

BDKS01 observe communication in all four incentive conditions with and without *a priori* meaning. In Games 2, 3 and 4, consistent with the *PCI* prediction, there is a tendency

under the antifraud rule, the unraveling argument holds and full information is revealed in equilibrium.

for type 3 to be separated from the remaining types. This separation is cleaner when messages have *a priori* meanings. One robust and stable non-equilibrium outcome they find is "sucker behavior," where the receiver keeps following action recommendations even when those recommendations reveal a type for which following recommendations is not optimal.

One of the most striking results in BDKS01 is obtained in Game 4 with *a priori* meanings. In the final five (out of 40) rounds of Game 4, type 1 always receives his maximal payoff, even though the corresponding action is not part of any equilibrium. Type 2 frequently either mimics type 1 or reveals his type by asking for his favorite action. It is not part of any equilibrium of the game that type 2 is revealed, because type 2 prefers to mimic type 1 to being revealed. In the data, while type 2 frequently mimics type 1 it is not enough to undermine type 1 always getting his favorite action.

The observed persistent communication in Game 4 of BDKS01 is interesting for three reasons. First, Farrell's [38] neologism-proofness criterion rejects all equilibria. Second, the persistent communication takes the form of stable non-equilibrium behavior. Third, the persistent communication pattern resembles what would be predicted by level-k behavior, even though it obtains in the final five out of 40 periods, and thus does not concern initial play in a game.

Message use in BDKS01's Game 4 is sensibly anchored at focal meanings but evolves over time to reflect incentives. Recall that in treatments with *a priori* meanings messages are framed as action recommendations. Senders primarily start out sending recommendations that if followed would maximize sender payoffs and sometimes make recommendations that would maximize receiver payoffs at their own expense. By the end of forty periods almost all senders either recommend actions that if followed would maximize their own payoff or mimic another type to guard against being identified.

Cai and Wang [21] (CW) examine a similar setup as Dickhaut, McCabe, and Mukherji [36]. Inspired by Crawford and Sobel [25], they use a payoff function with a bias parameter to derive the payoffs employed in the experiment. The state space is $\{1, 3, 5, 7, 9\}$, the action space is $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, and the sender's message space is the set of all subsets of the state space. The message observed by the receiver is a draw from a uniform distribution over the set selected by the sender. CW confirm Dickhaut, McCabe, and Mukherji [36]'s comparative statics finding that less information is transmitted as preference divergence increases, in line with the prediction from Crawford and Sobel's [25] leading uniform quadratic example. Although average payoffs are also close to the ones predicted by equilibrium, CW reject equilibrium as accounting for observed behavior.

Like Forsythe, Lundholm and Rietz [39], CW's primary focus is on the case where the bias is so large that theory admits only equilibria in which the receiver action does not vary with the state of the world. They confirm prior findings from the literature that there is over-communication. Senders are excessively truthful and receivers excessively credulous. Note that if the senders' over-communication were correctly interpreted by receivers, this would increase *ex ante* payoffs of both senders and receivers. CW find, however, that average observed payoffs are close to equilibrium payoffs and thus there is no *ex ante* gain from over-communication.

In other settings there are findings that over-communication *can* increase efficiency. Valley, Thompson, Gibbons and Bazerman [98] (VTGB) experimentally study the impact of communication in double-auctions with two-sided private information. We know from Myerson and Satterthwaite [79] that in this game all equilibria violate *ex post* efficiency and from Chatterjee and Samuelson [22] that there is a (linear) equilibrium that is maximally efficient. VTGB find that communication improves not only upon the non-communication payoff in the experiment, but also upon the theoretical efficiency bound, which is the same with and without communication.

The principal innovation in CW is subjecting the information transmission game to a level-k analysis, by adapting Crawford's [27] analysis of the communication of intentions (Crawford, Costa-Gomes and Iriberri [29] explain how to adapt Crawford [27] to information transmission games and get over-communication and excessive credulity without invoking lying costs). A crucial step in any level-k analysis is the specification of level-0 behavior. In many contexts a sensible approach is to have the lowest level non-strategic players randomize uniformly over all of their actions. Recognizing that this would not help organizing the observed data, CW choose to identify level-0 behavior with truthfulness for the sender and credulousness of the receiver.[3] A level-0 sender (S-L0) is truthful; a level-0 receiver (R-L0) best responds to S-L0; S-L1 best responds to R-L0; R-L1 best responds to S-L1; S-L2 best responds to R-L1; and, R-L2 best responds to S-L2. In addition they consider sophisticated types who best respond to the empirical distribution. Subjects are classified as belonging to a particular type if their behavior conforms with that type's behavior more than 60% of the time. In the data 75% of subjects can be classified using this standard. There is a small number of L0's, the most frequent sender types are S-L1's and S-L2's and the most frequent receiver types are R-L2's and sophisticated types. In CW's interpretation of their data this type distribution reconciles the observed over-communication with the fact that average payoffs are very close to those predicted by the most informative equilibrium (here pooling): lower type subjects' behavior accounts for the over-communication but since frequently they are not matched with each other this over-communication does not translate into higher expected payoffs.

It is worth noting that the level-k analysis predicts that S-L2 sends message "9" regardless of the state. While this is consistent with equilibrium behavior of the sender, equilibrium admits many other sender strategies. Interestingly in the data "9" is by far the most frequent message and the modal message for all sender types except the lowest type. Thus, aside from capturing bounded rationality, the level-k analysis rationalizes a particular use of messages, by anchoring the belief hierarchy in the focal meaning of messages. The modal message that emerges, both in the analysis and in the data, is not itself a truthful message but obtained from iterating on truth.

Wang, Spezio, and Camerer [100] (WSC) study a similar setup as Cai and Wang [21]. In addition to observing choices in the game, the experiment uses video-based eye tracking to measure pupil dilation and which payoffs or other game parameters sender subjects are

[3] A similar approach is taken by Franke [40] in his account of pragmatic inference.

looking at. In the game, a sender privately learns the true state in the set $\{1, 2, 3, 4, 5\}$ and sends a costless message out of $\{1, 2, 3, 4, 5\}$ to a receiver who then chooses an action among $\{1, 2, 3, 4, 5\}$. The payoffs of the sender and the receiver are determined by the distance, denoted by $b \in \{0, 1, 2\}$, between the ideal action of the sender and the ideal action of the receiver. In the experiment players' roles are fixed.

WSC observe over-communication and find that the choice data are consistent with the level-k model developed by Cai and Wang [21] in which L0 sender behavior is anchored in truth telling. Eye-tracking data provide additional support for the level-k model. In particular, sender subjects focus too much on the true state payoff row. Moreover, right before and after the message is sent, senders' pupils dilate more when their deception is larger in magnitude. This observation suggests that subjects feel guilty about deceiving or that deception is cognitively demanding.

The primary objective of Kawagoe and Takizawa [60] (KT) is to compare various equilibrium refinements and bounded rationality models in explaining experimental data obtained from three sender-receiver games. All the games considered (Games 1, 2, and 3) have two equally likely states A and B, three actions X, Y, and Z, and two available messages having literal meanings, "I am type A" and "I am type B." Actions X and Y are the receiver's *ex-post* ideal action for states A and B, respectively. Action Z is the *ex-ante* ideal action for the receiver.

The three games differ in the incentive structure of the sender. In Game 1, players' preferences are fully aligned with both sender types A and B wanting to be correctly identified and inducing the receiver to choose actions X and Y, respectively. In Game 2, players' preferences are misaligned with both sender types wanting the receiver to play Z. In Game 3, there is partial conflict of interests, i.e., sender type A wants to be correctly identified, while sender type B wants to be misidentified as type A. As a result, both a babbling equilibrium and a truth-telling equilibrium exist in Games 1 and 2 while only a babbling equilibrium exists in Game 3.

Among equilibrium refinements for cheap-talk games, KT consider neologism proofness by Farrell [38], announcement proofness by Matthews, Okuno-Fujiwara, and Postlewaite [72], and "recurrent mops" by Rabin and Sobel [84]. For Game 1, all three refinements predict the truth-telling equilibrium. For Game 2, all three refinements predict babbling. For Game 3, the unique babbling equilibrium survives both announcement proofness and recurrent mops, but not neologism proofness. For the bounded-rationality models, KT consider $AQRE$ (Agent Quantal Response Equilibrium) by McKelvey and Palfrey [73] with the precision parameter λ going to infinity and a level-k model by Crawford [31] in which the L0 sender is truthful, and the L0 receiver is randomizing.

Consistent with previous findings in the literature, KT show that a substantial proportion of players tend to be truthful as a sender and credulous as a receiver. Their findings can be better explained by the level-k model than other bounded-rationality models and equilibrium refinements discussed above. It is also worth noting that the restricted message space considered in the paper {"I am type A", "I am type B"} is far from satisfying the rich language assumption of Farrell [38], a key requirement for neologism proofness, announcement proofness, and recurrent mops to be meaningful.

de Groot Ruiz, Offerman, and Onderstal [32] (DOO) create and explore an equilibrium selection criterion based on neologism proofness by Farrell [38]. Instead of rejecting an equilibrium when there is a credible neologism, their *ACDC* (Average Credible Deviation Criterion) relies on a measure of strength of deviations and selects those equilibria for which credible deviations have the least strength. By construction, it never rejects all equilibria. Like its parent, neologism proofness, it is about rejecting equilibria and thus makes no predictions about which messages will be used and does not always make sharp predictions about which outcomes will be realized. DOO document that departures from equilibrium behavior increase as the strength of credible deviations increases and interpret this result as validating the credible deviations perspective.

Lai and Lim [69] compare the predictive power of neologism proofness in environments in which messages are endowed with literal meanings with its predictive power when a common language is initially absent.

3.2.2 Multiple senders – multiple receivers

Lai, Lim, and Wang [68] (LLW) and Vespa and Wilson [99] (VW) experimentally investigate the fully revealing equilibrium constructed by Battaglini [5] for a multidimensional state space with multiple senders and a single receiver. Recall that Battaglini's [5] equilibrium construction involves the senders truthfully revealing on distinct dimensions on which the senders share a common interest with the receiver. One essential characteristic of the fully revealing equilibrium is that the dimensions of common interests are *endogenously* determined based on the preferences of the players.

LLW create discrete environments with two senders sending simultaneous messages to a receiver regarding a 2 (horizontal dimension) $\times$ 2 (vertical dimension) state space. The receiver chooses among four actions. In their baseline treatment, each sender has four available messages framed as non-binding action recommendations. The players' ideal actions never coincide. However, assuming that each sender's influence on the receiver is limited to a distinct dimension, each sender and the receiver share a common ranking over the relevant actions. The dimensional preference alignment embodied in the payoff structure supports a fully revealing equilibrium in which one sender truthfully reveals along the horizontal dimension and the other sender along the vertical dimension. In each of fifty rounds of interactions, players are randomly matched while players' roles are fixed.

LLW look for answers to three main questions. First, under what circumstances can more information be extracted with two senders than with one sender? Second, how does the size of the message spaces affect the extent of information transmission? Third, how does the empirical performance of an equilibrium depend on the specification of out-of-equilibrium beliefs? To answer the first question, they consider a one-sender control treatment as well as a treatment in which the common interest between a sender and the receiver is along a diagonal of the state space. To investigate the second issue, they consider a treatment in which the message spaces become binary. For the last question, they consider a treatment in which one state is eliminated so that only three states remain.

The experimental results in LLW confirm the main qualitative predictions in Battaglini [5]. More information is extracted via senders revealing along different dimensions. Receivers, in general, identify true states more often with two senders than with one sender. Lower adherence to the fully revealing equilibrium play is observed when the state space is restricted.

All games with two senders in LLW admit fully revealing equilibria. However, they find that the observed extent of information transmission depends on the nature of dimensional common interests. This result is in contrast with Battaglini [5] in which a dimension of common interest is endogenous. Inconsistent with the Battaglini's [5] theory, the message spaces matter in the laboratory. When the message spaces are restricted to coincide with the dimension along which the preferences between a sender and the receiver are aligned, a drastically higher adherence to the fully revealing equilibrium is observed.

VW focus on how the directions of the senders' biases affect the attainment of full revelation. They use two circles in their experimental design to represent the two-dimensional state space of Battaglini. The original continuous state space in Battaglini is discretized, where each dimensional component has a cardinality of 360 and is pinpointed by the angle of the point drawn uniformly on the circumference of the circle. The senders' messages and the receiver's actions are similarly represented.

VW find that Battaglini's [5] equilibrium predicts the laboratory behavior reasonably well when the common interest between each sender and the receiver is along the regular horizontal or vertical coordinate axes. When the establishment of common interests requires a rotation of the coordinate axes, the observation of full revelation disappears. In this case, the behavior of sender subjects is characterized by exaggeration in the directions of their interests. These experimental findings suggest that the framing of the multidimensional information transmission problems, though irrelevant in theory according to Battaglini [5], could make a difference empirically.

Minozzi and Woon [76] (MW) also study an environment in which two senders send simultaneous messages to a single receiver. Unlike LLW and VW, the state space is unidimensional, and senders' biases are private information. MW use three distinct theoretical predictions to formulate their hypotheses: babbling equilibrium, informative partition equilibria *à la* Crawford and Sobel [25], and jamming equilibria characterized by Minozzi [75]. In their experiments, MW consider relatively large state and action spaces; both are the set of integers between -100 and 100. The message space is the set of integers between -150 and 150. At the beginning of the experiment, players' roles are randomly determined and fixed. In each of either 24 or 32 rounds, three subjects are randomly matched to form a group of two senders and one receiver. They find that equilibrium predictions fare poorly and that senders over-communicate by consistently exaggerating their messages. Receivers excel at matching their actions to their ideal action simply by taking the average of the two messages from the senders. Over time, exaggeration by senders increases and communication unravels.

MW explore two models of bounded rationality to explain their findings. The first is a level-k model in which L0 senders are naively truthful (or almost equivalently, naively selfish) and L0 receivers take an action that matches the average of the two messages. The second one is the experiential best response model in which players choose best responses to their non-equilibrium beliefs, while beliefs depend on each subject's empirical observations obtained by their experiences in the game. They find that neither model dominates the other in terms of accounting for the data. The level-k predictions are better in explaining the observed behavior of one sender, and the experiential best response predictions are better in explaining the observed behavior of the other sender. However, both models predict much more exaggeration in senders' messages than is observed in the data.

Battaglini, Lai, Lim, and Wang [7] (BLLW) experimentally evaluate the informational theory of legislative committees first proposed by Gillian and Krehbiel [45] and further developed by Krishna and Morgan [65]. The two theoretical papers provide different equilibrium characterizations, and BLLW set out to investigate which of these better predicts behavior in the laboratory. The state space is unidimensional, there are two senders, and one receiver, like in MW.

Battaglini and Makarov [6] experimentally study Farrell and Gibbons' [37] theoretical model in which one sender transmits information to multiple receivers. They consider several variations of two-type (Heads and Tails), two-action (A and B), and two-message (Heads and Tails) games. Each experimental session consists of two parts, Part A and Part B. In Part A, subjects are divided into pairs and play the game with one sender and one receiver for eighteen rounds. Six games with different payoff structures are repeated three times in a random sequence. In each period, subjects are randomly assigned to a group. Players' roles are randomly determined in each round. In Part B, subjects are divided into groups of three and each of the four groups has one sender and two receivers. Each of 5 games with different payoff structures is repeated four times for a total of twenty rounds. Subjects play Part A first in 6 sessions and play Part B first in 2 sessions.

The key finding of Battaglini and Makarov [6] is that the marginal effect of having an additional receiver is as predicted by Farrell and Gibbons [37]. Specifically, this includes the following three cases. First, confirming *one-sided discipline*, they observe truthful revelation with public messages by the sender when information transmission is possible with only one but not the other receiver with private messages. Second, confirming *mutual discipline*, they observe truthful revelation with public messages despite the fact that information transmission is not possible with any of the two receivers when messages are private. Third, confirming *subversion*, adding a second receiver in the public message environment renders communication impossible when full revelation is possible with the first receiver in the private message environment. They confirm prior findings in the literature regarding over-communication, truth bias and credulity bias. Learning tends to correct departures from equilibrium predictions.

3.2.3 Vague language

A term is vague if it admits borderline cases. In a borderline case it is not clear whether or not the term applies. There is, for example, no definite number of hairs that separates those whom we would call 'bald' from those whom we would not refer to as 'bald.' Vagueness seems useful, but is not straightforward to rationalize. Equilibrium meaning in Crawford-Sobel games takes the form of referring to a specific well-defined set of payoff types. Lipman [71] points out that in common-interest sender-receiver games of the form considered by Lewis [70] there cannot be a benefit from vagueness. O'Connor [82] argues that in common-interest games that reward partial success vagueness can arise as the result of incomplete learning. Blume and Board [15] introduce *indeterminacy of meaning*, the confounding of payoff-relevant information with information about language competence. If, for example, the receiver is uncertain about which messages are available to the sender, he faces a signal extraction problem when interpreting a message because he does not know whether this is exactly the message the sender prefers to send conditional on all messages being available, or just the closest approximation within a restricted set of available messages. In some situations indeterminacy of meaning can be avoided by only using messages in equilibrium that are commonly known to be available. Blume and Board show, however, that dropping message that are not commonly available entails an *ex ante* payoff loss. Hence, with private information about language competence it is optimal to have indeterminacy of meaning in common-interest sender-receiver games. With indeterminacy of meaning, message meaning is no longer a specific well-defined set of payoff types. Instead, in the payoff-type space meaning becomes a non-degenerate distribution over payoff types. Overlapping supports of those distributions somewhat resemble borderline cases, and therefore vagueness.

Blume and Board's notion of indeterminacy of meaning can be easily extended to other confounds besides private information about language competence. There are many sources of what Quine [83] refers as "prior collateral information" that give rise to discrepancies between the interpretations different interlocutors give to words and more complex expressions. Indeterminacy of meaning is related to what Frankel and Kartik [41] refer to as "muddled information" in signaling models with multidimensional private information; communication in the dimension of interest may be adversely affected by the presence of nuisance dimensions.

Having access to a vague language may create plausible deniability and soften incentive constraints. deJaegher and van Rooij [33] and Blume and Board [16] explore the role of vagueness in sender-receiver games with imperfectly aligned interests. Blume and Board's findings suggest that there may be circumstances where vagueness is efficiency enhancing.

Serra-Garcia, van Damme, and Potters [90] (SvDP) study the role of vagueness in preserving efficiency. They examine a public goods game where efficiency can be achieved without communication, but truthful communication would undermine it. There is an informed "leader" and an uninformed "follower." The leader first learns which of three states of the world, low, intermediate or high, obtains. In the game without communication, the leader then decides whether or not to contribute. After observing the leader's contribution decision, the follower decides whether or not to contribute. It is efficient to contribute if and only if the state is intermediate or high. There is an equilibrium that supports that outcome. Part of what holds the equilibrium together is that the follower does not learn whether the state is intermediate or high. As a consequence, when communication is added to the game there is no equilibrium in which the sender adopts a separating strategy. Hence, if truth is feasible, i.e., messages with appropriate meanings are available, the full truth is not supported by an equilibrium. There is an equilibrium in which intermediate and high types pool.

In their experiment SvDP compare behavior when only precise messages are available with behavior when leaders in addition have access to vague messages. They find that with only precise messages, leaders frequently lie in the intermediate state by pretending that the state is high. As a result high and intermediate types end up pooling. When vague messages are available, leaders use them in the intermediate state. With high types indicating that they are high, low types that they are low and intermediate types using vague messages, leaders are using a separating strategy. This is not equilibrium behavior. It is stable in the experiment because followers fail to make proper inferences.

Similar to Serra-Garcia, van Damme, and Potters [90], Wood [105] investigates how the availability of imprecise messages affects the observed extent of information transmission. Wood [105] considers experimental communication games that build on the games presented in Cai and Wang [21] and Wang, Spezio, and Camerer [100], which are discretized versions of the canonical model of Crawford and Sobel [25]. The state and action spaces are $\{1, 2, 3, 4, 5\}$ and the prior is uniform. The receiver's incentive is to take an action that matches the realized state while the sender's incentive is to induce the receiver to take an action that is higher than the realized state by a fixed number. The primary treatment variable is the form of the message space. In the baseline treatment, the message space coincides with the state/action space. In the second, rich language, treatment, subjects are endowed with a message space that contains not only the five precise messages but also three imprecise messages with literal meanings that span three states (e.g., "the state is 1, 2, or 3"). In the third, noise treatment, the same message space as the rich language treatment is given to subjects. However, the observation of the state is imprecise such that there is a bijection between the message space and the set of possible observations by the sender.

The major finding from the experiment is that the availability of imprecise messages improves information transmission. Wood [105] also confirms the common, robust finding of the experimental literature on communication games that subjects over-communicate. He develops a cognitive hierarchy model augmented with a preference for truth-telling to account for the over-communication observed. The results from the structural estimation reveal that both bounded rationality and honest behavior could explain the subjects' behavior, but the former is relatively more important.

Agranov and Schotter [2] (AS) investigate communication between a utilitarian sender and two receivers. Before sending a public message the sender privately learns the state of the world. In each state, if it were commonly known, the receivers would be playing a simultaneous move game in which each receiver has two actions. There are four equally likely states. In two of these states, states 1 and 3, the receivers play a battle of the sexes game;

in state 2 they play a pure coordination game and in state 4 both have a dominant action. Coordination is difficult in the battle of the sexes states. This is true even though actions have non-neutral focal labels, as demonstrated by Crawford, Gneezy and Rottenstreich [28].

The payoff structure is such that if the receivers only know that they are not in the dominant action state the *expected* payoffs are those of a pure communication game. One may reasonably expect then that if the sender only communicates whether or not the state is state 4, whenever the state is in the set $\{1, 2, 3\}$ receivers use the focal point to coordinate their actions. AS find that this is indeed the case, both in treatments where the sender is computerized and in treatments with human senders whose payoffs are the sum of receiver payoffs. This demonstrates that there is a possible beneficial role for ambiguous or vague communication. It can conceal conflict that would hinder the use of focal points for coordination if it became publicly known.

AS consider a variety of message spaces: one that is identical to the state space (state-space messages), one that corresponds to the collection of subsets of the state space (subset messages) and one that consists of the words "high" and "low" (word messages). They interpret messages consisting of non-degenerate subsets of the state space as ambiguous and word messages as vague. With a computerized sender and subset messages it is commonly known that the sender sends $\{1, 2, 3\}$ if and only if the state is indeed in that set. With word messages the computerized sender sends "low" if and only if the state is in the set $\{1, 2, 3\}$, and receivers know that a fixed strategy is used, but not which one. Either form of coarse messaging improves coordination, with ambiguous messages being somewhat more effective in accomplishing that goal than vague messages.

Human senders who are restricted to state space messages either send truthful messages or conceal the true state with sending message 2 for states in the set $\{1, 2, 3\}$. When subset messages are available modal sender behavior is to be truthful and senders use a variety of ways of concealing the true state when the state is in the set $\{1, 2, 3\}$. Being truthful when subset messages are available is an instance of over-communicaton, which in the AS setup is detrimental to efficiency.

All message spaces in AS are composed of indicatives. An open question is what would happen with message spaces that consist either only of directives or that permit a choice between directives and indicatives. In the AS setting directives would be a way of transmitting coarse information about the state while being precise about the suggested action. It is then conceivable that precise directives might be good substitutes for ambiguous/vague indicatives.

3.2.4 Key findings from experiments with imperfect incentive alignment

From its inception the experimental literature on strategic information transmission with a single sender and a single receiver documents systematic over-communication, relative to the most informative equilibria (an exception is a recent paper by Cabrales, Feri, Gottardi, and Meléndez-Jiménez [19] who find under-communication when there is a market for cheap-talk information). Some of this over-communication can be accounted for by bounded-rationality approaches, like learning and level-k reasoning. Theories of equilibrium selection have limited success, not only because of the observed over-communication, but also because in some cases these theories are silent when there are clear patterns in the data and equally silent about how messages are used. Here also an appeal to level-k reasoning does a better job; it captures not only structure in outcomes (joint distributions of states and actions) but also makes systematic predictions about message use.

Some of the theoretical predictions for games with added senders or added receivers are supported by the data. There is some evidence that adding experts aids information transmission. Similarly there is some support for disciplining and hindering audience effects with more than one receiver. At the same time some of the more striking theoretical predictions, e.g., that under fairly general conditions there will be full revelation fail to be supported by the evidence.

Language matters. Having access to a language with preestablished meanings rather than having to learn to attach meanings to messages affects outcomes even in the long run. In one-sender environments, the availability of vague messages interacts with an intrinsic preference for truth-telling and can be efficiency enhancing. In two-sender environments, restricting message spaces can improve information transmission.

4 Looking ahead

Sobel [93] offers a stimulating list of questions to pursue in future experiments on sender-receiver games. At the risk of repetition, here we offer some of our own.

We believe that the role of language in sender-receiver games is still under-explored. There is, for example, no systematic investigation that compares the effect of message spaces in which meanings are indicative (subsets of the type space) with that of messages spaces in which meanings are imperative (suggested actions) or message spaces in which both indicative and imperative meanings are available. Likewise, while there is some work on the difficulties of combining languages (see Weber and Camerer [103]), there is none about communication with persistent language differences and persistent uncertainties about those differences.

One area in which theory is separated from experiments by a particularly wide gap is what one might call "communication design." Theory suggests various protocols and mediated schemes with the potential to improve communication outcomes (e.g. Goltsman, Hörner, Pavlov and Squintani [48]) (GHPS). With the exception of Blume, Lai and Lim [17], we are not aware of any experimental work that intersects with this theoretical literature. BLL is not a direct implementation of GHPS in the lab, since BLL rely on a preference for truth-telling (or following instructions), which they inherit from the literature on the randomized response technique (compare Warner [102]). While undoubtedly psychological considerations, like preferences for following instructions, truth-telling, and reluctance to deceive play a significant role in communication, and have attracted a great deal of recent attention (see Abeler, Nosenzo, Raymond [1] for a survey), we believe there is a continuing need for understanding communication between selfish actors who are primarily motivated by material gains.

In a world where we are invited not to take public statements literally (Peter Thiel at the National Press Club [97]) we cannot expect that the shame that Adam Smith [91] speaks of secures truthful communication. In that world, understanding how self-motivated calculating actors use communication to manipulate others is as important as ever.

References

[1] Abeler, Johannes, Daniele Nosenzo and Collin Raymond [2016] "Preferences for Truth-Telling," University of Oxford Working Paper.

[2] Agranov, Marina, and Andrew Schotter [2012], "Ignorance is Bliss: an Experimental Study of the Use of Ambiguity and Vagueness in the Coordination Games with Asymmetric Payoffs." *American Economic Journal: Microeconomics* **4**, 77–103.

[3] Akerlof, George [1970], "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* **84**, 488-500.

[4] Austen-Smith, David [1994], "Information Transmission in Debate," *American Journal of Political Science* **34**, 124–152.

[5] Battaglini, Marco [2002], "Multiple Referrals and Multidimensional Cheap Talk," *Econometrica* **70**, 1379-1401.

[6] Battaglini, Marco, and Uliana Makarov [2014], "Cheap Talk with Multiple Audiences: An Experimental Analysis," *Games and Economic Behavior* **83**, 147–164.

[7] Battaglini, Marco, Ernest K. Lai, Wooyoung Lim, and Joseph Tao-yi Wang [2016], "The Informational Theory of Legislative Committees: An Experimental Analysis," CEPR Discussion Paper No. DP11356.

[8] Blume, Andreas, Yong-Gwan Kim, and Joel Sobel [1993], "Evolutionary Stability in Games of Communication," *Games and Economic Behavior* **5**, 547–575.

[9] Blume, Andreas [1996], "Neighborhood Stability in SenderReceiver Games," *Games and Economic Behavior* **13**, 2–25.

[10] Blume, Andreas, Douglas V. DeJong, Yong-Gwan Kim and Geoffrey B. Sprinkle [1998], "Experimental Evidence on the Evolution of Meaning of Messages in Sender-Receiver Games," *American Economic Review* **88**, 1323–1340.

[11] Blume, Andreas [2000], "Coordination and Learning with a Partial Language," *Journal of Economic Theory* **95**, 1–36.

[12] Blume, Andreas, Douglas V. DeJong, Yong-Gwan Kim and Geoffrey B. Sprinkle [2001], "Evolution of Communication with Partial Common Interest," *Games and Economic Behavior* **37**, 79–120.

[13] Blume, Andreas [2004], "A Learning-Efficiency Explanation of Structure in Language," *Theory and Decision* **57**, 265–285.

[14] Blume, Andreas, Oliver J. Board and Kohei Kawamura [2007], "Noisy Talk," *Theoretical Economics* **2**, 395–440.

[15] Blume, Andreas and Oliver J. Board [2013], "Language Barriers," *Econometrica*, **81**, 781–812.

[16] Blume, Andreas and Oliver J. Board [2014], "Intentional Vagueness," *Erkenntnis* **79**, 855–899.

[17] Blume, Andreas, Ernest K. Lai, and Wooyoung Lim [2015], "Eliciting Private Information with Noise: The Case of Randomized Response," Working Paper.

[18] Bruner, Justin, Cailin O'Connor, Hannah Rubin, and Simon M. Huttegger [2014], "David Lewis in the Lab: Experimental Results on the Emergence of Meaning," *Synthese* **88**,1–19.

[19] Cabrales, Antonio, Francesco Feri, Piero Gottardi, and Miguel A. Meléndez-Jiménez [2016], "Can there be a Market for Cheap-Talk Information? Some Experimental Evidence," University College London Working Paper.

[20] Charness, Gary, and Martin Dufwenberg [2006], "Promises and Partnership," *Econometrica* **74**, 1579-1601.

[21] Cai, Hongbin, and Joseph Tao-Yi Wang [2006], "Over-Communication in Strategic Information Transmission Games," *Games and Economic Behavior* **56**, 7-36.

[22] Chatterjee, Kalyan, and William Samuelson [1983], "Bargaining under Incomplete Information," *Operations Research* **31**, 835–851.

[23] Chen, Ying, Navin Kartik, and Joel Sobel [2008], "Selecting Cheap–Talk Equilibria," *Econometrica* **76**, 117–136.

[24] Cho, In-Koo, and David M. Kreps [1987], "Signaling Games and Stable Equilibria," *The Quarterly Journal of Economics* **102**, 179-221.

[25] Crawford, V.P.and J. Sobel [1982], "Strategic Information Transmission," *Econometrica* **50**, 1431–1451.

[26] Crawford, Vincent P. [1998], "A Survey of Experiments on Communication via Cheap Talk," *Journal of Economic Theory* **78**, 286–298.

[27] Crawford, Vincent P. [2003], "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions," *American Economic Review* **93**, 133–149.

[28] Crawford, Vincent, Uri Gneezy and Yuval Rottenstreich [2008], "The Power of Focal Points Is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures," *American Economic Review* **98**, 1443-58.

[29] Crawford, Vincent P., Miguel A. Costa-Gomes, and Nagore Iriberri [2013], "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications," *Journal of Economic Literature* **51**, 5–62.

[30] Crawford, Vincent P. [2016], "New Directions for Modelling Strategic Behavior: Game-Theoretic Models of Communication, Coordination, and Cooperation in Economic Relationships," *Journal of Economic Perspectives* **30**, 131–149.

[31] Crawford, Vincent [2016], "Let's Talk it Over: Coordination via Preplay Communication with Level-k Thinking," *Research in Economics.*

[32] de Groot Ruiz, Adrian, Theo Offerman, and Sander Onderstal [2014], "For those about to Talk we Salute you: an Experimental Study of Credible Deviations and ACDC," *Experimental Economics* **17**, 173–199.

[33] De Jaegher, Kris, and Robert van Rooij [2014], "Strategic vagueness, and appropriate contexts," 40–59, in Benz, Anton, Christian Ebert, Gerhard Jäger, and Robert Van Rooij (edts.) *Language, Games, and Evolution*, Springer: Berlin, Heidelberg.

[34] De Jaegher, Kris, Stephanie Rosenkranz, and Utz Weitzel [2014], "Economic Principles in Communication: An Experimental Study," *Journal of Theoretical Biology* **363**, 62–73.

[35] de Saussure, Ferdinand [1916], *Cours de Linguistique Générale*, ed. C. Bally and A. Sechehaye, with the collaboration of A. Riedlinger, Lausanne and Paris: Payot; trans. W. Baskin, Course in General Linguistics, Glasgow: Fontana/Collins, 1977.

[36] Dickhaut, John W., Kevin A. McCabe, and Arijit Mukherji [1995], "An Experimental Study of Strategic Information Transmission." *Economic Theory* **6**, 389-403.

[37] Farrell, Joseph, and Robert Gibbons [1989], "Cheap Talk with two Audiences," *American Economic Review* **79**, 1214–1223.

[38] Farrell, Joseph [1993], "Meaning and Credibility in Cheap-Talk Games," *Games and Economic Behavior* **5**, 514531.

[39] Forsythe, Robert, Russell Lundholm, and Thomas Rietz [1999], "Cheap Talk, Fraud, and Adverse Selection in Financial Markets: Some Experimental Evidence," *Review of Financial Studies* **12**, 481-518.

[40] Franke, Michael [2009], *Signal to Act.* PhD Thesis, Institute for Logic, Language and Computation, University of Amsterdam.

[41] Frankel, Alex and Navin Kartik [2016], "Muddled Information," Columbia University Working Paper.

[42] Frege, Gottlob [1892] "Über Sinn und Bedeutung," *Zeitschrift für Philosophie und philosophische Kritik* **100**, 25–50; translated as "On Sense and Reference," in P.T. Geach and M.Black, (eds.), *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell (1952), 56–78.

[43] Garrod, Simon, Nicholas Fay, John Lee, Jon Oberlander, and Tracy MacLeod [2007], "Foundations of Representation: Where Might Graphical Symbol Systems Come From?" *Cognitive Science* **31**, 961-987.

[44] Gigler, Frank [1994], "Self-Enforcing Voluntary Disclosures," *Journal of Accounting Research* **32**, 224-240.

[45] Gilligan, Thomas W. and Keith Krehbiel [1989], "Asymmetric Information and Legislative Rules with a Heterogeneous Committee," *American Journal of Political Science* **33**, 459-490.

[46] Gneezy, Uri [2005], "Deception: The Role of Consequences," *The American Economic Review* **95**, 384–394.

[47] Goel, Anand M., and Anjan V. Thakor [2015], "Information Reliability and Welfare: A Theory of Coarse Credit Ratings," *Journal of Financial Economics* **115**, 541–557.

[48] Goltsman, Maria, Johannes Hörner, Gregory Pavlov, and Francesco Squintani [2009], "Mediation, Arbitration and Negotiation," *Journal of Economic Theory* **144**, 1397-1420.

[49] Goltsman, Maria, and Gregory Pavlov [2011], "How to Talk to Multiple Audiences," *Games and Economic Behavior* **72**, 100–122.

[50] Goltsman, Maria, and Gregory Pavlov [2014], "Communication in Cournot Oligopoly," *Journal of Economic Theory* **153**, 152–176.

[51] Grice, H. Paul [1975], "Logic and Conversation," in P. Cole and J. Morgan (editors) *Syntax and Semantics*, vol. 3, Academic Press.

[52] Grossman, Sanford J. [1981], "The Informational Role of Warranties and Private Disclosure about Product Quality," *The Journal of Law & Economics* **24**, 461–483.

[53] Hagenbach, Jeanne and Eduardo Perez-Richet [2016], "Communication with Evidence in the Lab," CNRS – Ecole Polytechnique Working Paper.

[54] Healey, Patrick G. T., Nik Swoboda, Ichiro Umata, and Yasuhiro Katagiri [2002], "Graphical Representation in Graphical Dialogue," *International Journal of Human-Computer Studies* **57**, 375-395.

[55] Hong, Fuhai, Wooyoung Lim, and Xiaojian Zhao [2017], "The Emergence of Compositional Grammars in Artificial Codes," *Games and Economic Behavior* **102**, 255-268.

[56] Horn, Laurence [1984], "Toward a New Taxonomy for Pragmatic Inference: Q-based and R-based Implicature," in *Meaning, Form, and Use in Context: Linguistic Applications*, Schiffrin, Deborah (edt.), Georgetown University Press, 11–42.

[57] Kalai, Ehud, and Dov Samet [1984], "Persistent Equilibria in Strategic Games," *International Journal of Game Theory* **13**, 129–144.

[58] Kamenica, Emir, and Matthew Gentzkow [2011], "Bayesian Persuasion," *The American Economic Review* **101**, 2590–2615.

[59] Kartik, Navin, Marco Ottaviani, and Francesco Squintani [2007], "Credulity, Lies, and Costly Talk," *Journal of Economic Theory* **134**, 93–116.

[60] Kawagoe, Toshiji, and Hirokazu Takizawa [2009], "Equilibrium Refinement vs. Level-k Analysis: An Experimental Study of Cheap-Talk Games with Private Information," *Games and Economic Behavior* **66**, 238-255.

[61] Kohlberg, Elon, and Jean-Francois Mertens [1986], "On the Strategic Stability of Equilibria," *Econometrica* **54**, 1003-1037.

[62] Krauss, Robert M. and Sidney Weinheimer [1964], "Changes in the Length of Reference Phrases as a Function of Social Interaction: A Preliminary Study," *Psochonomic Science* **1**, 113–114.

[63] Krauss, Robert M. and Sidney Weinheimer [1966], "Concurrent Feedback, Confirmation, and the Encoding of Referents in Verbal Communication," *Journal of Personality and Social Psychology* **4**, 343–346.

[64] Kreps, David M., and Robert Wilson [1982], "Sequential Equilibria," *Econometrica* **50**, 863-894.

[65] Krishna, Vijay and John Morgan [2001], "Asymmetric Information and Legislative Rules: Some Amendments," *American Political Science Review* **95**, 435-452.

[66] Krishna, Vijay and John Morgan [2004], "The Art of Conversation: Eliciting Information from Informed Parties Through Multi-Stage Communication," *Journal of Economic Theory* **117**, 147-179.

[67] Kübler, Dorothea, Wieland Müller, and Hans-Theo Normann [1966], "Job-market Signaling and Screening: An Experimental Comparison," *Games and Economic Behavior* **64**, 219–236.

[68] Lai, Ernest K., Wooyoung Lim, and Joseph Tao-yi Wang [2015], "An Experimental Analysis of Multidimensional Cheap Talk," *Games and Economic Behavior* **91**, 114–144.

[69] Lai, Ernest K. and Wooyoung Lim [2016], "Meaning, its Evolution and Credibility in Experimental Cheap-Talk Games," HKUST Working Paper.

[70] Lewis, David [1969], *Convention: A Philosophical Study*, Harvard University Press, Cambridge, MA.

[71] Lipman, Barton L. [2009], "Why is language vague?," Boston University Working Paper.

[72] Matthews, Steven, Masahiro Okuno-Fujiwara, and Andrew Postlewaite [1991], "Refining Cheap-Talk Equilibria," *Journal of Economic Theory* **55**, 247-273.

[73] McKelvey, Richard and Thomas Palfrey [1998], "Quantal Response Equilibria for Extensive Form Games," *Experimental Economics* **1**, 9-41.

[74] Milgrom, Paul R. [1981] "Good News and Bad News: Representation Theorems and Applications," *Games and Economic Behavior* **12**, 380–391.

[75] Minozzi, William [2011], "A Jamming Theory of Politics," *Journal of Politics* **73**, 301-315.

[76] Minozzi, William and Jonathan Woon [2016] "Competition, Preference Uncertainty, and Jamming: A Strategic Communication Experiment," *Games and Economic Behavior* **96**, 97-114.

[77] Morgan, John, and Phillip C. Stocken [2003], "An Analysis of Stock Recommendations," *RAND Journal of Economics* **34**, 183–204.

[78] Myerson, Roger B. [1982] "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics* **10**, 67-81.

[79] Myerson, Roger B., and Mark A. Satterthwaite [1982] "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory* **29**, 265–281.

[80] Myerson, Roger B. [1991], *Game Theory: Analysis of Conflict*, Harvard University Press, Cambridge, MA.

[81] Nguyen, Quyen [2016] "Bayesian Persuasion: Evidence from the Laboratory," Utah State University, Working Paper.

[82] O'Connor, Cailin [2014] "The Evolution of Vagueness," *Erkenntnis* **79**, 707–727.

[83] QUINE, WILLARD VAN ORMAN [1960], *Word & Object*, MIT Press, Cambridge, MA. 179–217.

[84] RABIN, MATTHEW AND JOEL SOBEL [1996] "Deviations, Dynamics and Equilibrium Refinement," *Journal of Economic Theory* **68**, 1-25.

[85] ROTH, ALVIN E., AND MICHAEL W. MALOUF [1979] "Game-theoretic Models and the Role of Information in Bargaining," *Psychological Review* **86**, 574–594.

[86] RUBINSTEIN, ARIEL [1996] "Why are Certain Properties of Binary Relations Relatively more Common in Natural Language," *Econometrica* **64**, 343–355.

[87] SCHELLING, THOMAS [1960], *The Strategy of Conflict*, Harvard University Press, Cambridge, MA.

[88] SELTEN, REINHARD [1967], "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments," in H. Sauermann (ed.), *Beiträge zur experimentellen Wirtschaftsforschung*, Tübingen: Mohr, 136–168.

[89] SELTEN, REINHARD, AND MASSIMO WARGLIEN [2007] "The Emergence of Simple Languages in an Experimental Coordination Game," *Proceedings of the National Academy of Sciences* **104**, 7361–7366.

[90] SERRA-GARCIA, MARTA, ERIC VAN DAMME, AND JAN POTTERS [2011], "Hiding an inconvenient truth: Lies and vagueness." *Games and Economic Behavior* **73**, 244-261.

[91] SMITH, ADAM [1997], *The Theory of Moral Sentiments*, Regnery Publishing, Inc., Washington, DC (originally published in 1759, this is a reproduction of the sixth and final edition from 1817).

[92] SPENCE, MICHAEL [1973] "Job Market Signaling," *The Quarterly Journal of Economics* **87**, 355–374.

[93] SOBEL, JOEL [2013a] "Ten Possible Experiments on Communication and Deception," *Journal of Economic Behavior & Organization* **93**, 408–413.

[94] SOBEL, JOEL [2013b] "Giving and Receiving Advice," in Daron Acemoglu, Manuel Arellano, and Eddie Dekel (eds.), *Advances in Economics and Econometrics: Tenth World Congress (Econometric Society Monographs) (Volume 1)*, Cambridge University Press, New York, NY, 305–341.

[95] SOBEL, JOEL [2016] "Lying and Deception in Games," University of California - San Diego, Working Paper.

[96] STEIN, JEREMY [1989], "Cheap Talk and the Fed: A Theory of Imprecise Policy Announcements," *American Economic Review* **79**, 32–42.

[97] Thiel, Peter [2016] at *National Press Club Newsmaker with Peter Thiel* on Monday, October 31, 2016.

[98] Valley, Kathleen, Leigh Thompson, Robert Gibbons, and Max H. Bazerman [2002], "How Communication Improves Efficiency in Bargaining Games," *Games and Economic Behavior* **38**, 127-155.

[99] Vespa, Emanuel and Wilson, Alistair J. [2016], "Communication with Multiple Senders: An Experiment," *Quantitative Economics* **7**, 1–36.

[100] Wang, Joseph Tao-yi, Michael Spezio, and Colin F. Camerer [2010], "Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games," *American Economic Review* **100**, 984–1007.

[101] Wärneryd, Karl [1993] "Cheap Talk, Coordination, and Evolutionary Stability." *Games and Economic Behavior* **5**, 532–546.

[102] Warner, Stanley L. [1965] "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American Statistical Association* **60**, 63-69.

[103] Weber, Roberto A. and Colin F. Camerer [2003], "Cultural Conflict and Merger Failure: An Experimental Approach," *Management Science* **49**, 400-415.

[104] Wittgenstein, Ludwig [1958], *Philosophical Investigations*, Prentice Hall, Upper Saddle River, NJ.

[105] Wood, Daniel [2016], "Communication-Enhancing Vagueness," Clemson University Working Paper.