Noname manuscript No.
(will be inserted by the editor)

Communication Without the Cooperative Principle

A Signaling Experiment

Abstract According to Grice's 'Cooperative Principle', human communicators are involved in a cooperative endeavor. The speaker attempts to make herself understood and the listener, in turn, assumes that the speaker is trying to maximize the ease and effectiveness of communication. While pragmatists recognize that people do not *always* behave in such a way, the Cooperative Principle is generally assumed to hold. However, it is often the case that the interests of speakers and listeners diverge, at least to some degree. Communication can arise in such situations when the cost of signaling is high enough that it aligns the interests of speaker and listener, but what happens when the cost of signaling is not sufficient to align the interests of those communicating? In these cases the theoretical prediction is that they will reach a partially informative system of communication. Using methods from experimental economics, we test whether theoretical predictions are borne out. We find that subjects do learn to communicate without the cooperative principle.

Keywords Cooperative Principle · Game Theory · Experimental Philosophy · Philosophy of Language

1 Introduction

> "Our talk exchanges...are characteristically, to some degree at least, cooperative efforts; and each participant recognizes in them, to some extent, a common purpose or set of purposes, or at least a mutually accepted direction." (Grice, 1991, p. 26)

In the quote above, Grice elaborates what he calls the 'Cooperative Principle'—that participants in human communication usually are involved in a cooperative endeavor. The speaker attempts to make herself understood and the listener, in turn, assumes that the speaker is trying to maximize the ease and effectiveness of communication. While Grice recognizes that humans do not *always* behave in such a way, he and others studying the pragmatics of human communication assume that the Cooperative Principle will generally hold.

In contrast, economists and social scientists who employ game theory to study human communication often seem to assume the opposite—that to at least some degree the interests of speakers and listeners diverge. Examples of communication under divergent interests include job hunters and companies looking to hire, or firms planning to go public and those who might buy stock. Outside the realm of economic behavior, one can find many further examples of such situations—people on first dates, teenagers and their parents, students and their teachers, bosses and lazy employees. One of the major findings of the game theoretic literature on this topic is that when signals are costly, and when those sending the signals pay differential costs to do so, honest communication can arise in spite of divergent interests (Spence, 1973). In such cases, the costs for signaling remove conflict of interest between the sender and receiver. In other words, communication is possible in such situations as long as costs ensure that actors are incentivized to adhere to the Cooperative Principle. We will say more about this later in the paper.

Recently, philosophers have begun to investigate the emergence of communication when interests diverge as well. The traditional game theoretic approach to behavior uses assumptions of rationality to predict strategic choice. Evolutionary game theory, on the other hand, looks at how actors come to behave strategically through learning (or natural selection, in the biological world). This approach has been favored by many philosophers as a more wholistic way to understand human behavior, including human communication.[1] In particular, philosophers have recently employed evolutionary game theoretic models to show that a type of partially communicative equilbria, usually ignored by economists, arises commonly when actors with divergent interests learn to communicate (Wagner, 2013; Huttegger and Zollman, 2010). Throughout the paper, we will refer to these partially communicative outcomes as hybrid equilibria.[2] Importantly, in these hybrid equilibria costs to signalers do not bring the interests of the actors in line. In other words, actors learn to communicate, even though they are not incentivized to adhere to the Cooperative Principle, and do not do so at equilibrium.[3]

In this paper, we investigate the possibility of hybrid equilibria arising in real scenarios of human communication. We look at groups of actors in experimental settings to see whether they develop such partially communicative behavior. We show that, in fact, such outcomes do occur in the lab. This result

[1] For recent examples of this sort of work on communication see Skyrms (2010), Barrett and Zollman (2009), or O'Connor (2014).

[2] This follows the use of the term by economists. 'Hybrid' is used because such equilibria have characteristics that are both communicative and non-communicative.

[3] We are careful with our wording here because, of course, humans may adhere to cooperative conversational norms even when it is not in their material interests. In such cases, the Cooperative Principle will hold, though it is not consistent with equilibrium behavior in a game. As we will argue, there are important cases where humans transfer information, but do not follow cooperative norms.

is perhaps surprising because actors learn to communicate even though their interests are misaligned.

The paper will proceed as follows. In section 2, we outline the costly signaling model that is employed here and discuss costly and hybrid equilibria in this model. In section 3, we focus on the recent exploration of hybrid equilibria in evolutionary models. Then, in section 4 we describe our experimental set-up and in section 5 we present our results. In the conclusion we briefly discuss work on the pragmatics of human communication when the Cooperative Principle does not hold.

2 The Model

Costly signaling has been studied both in economics and in evolutionary biology, starting with Spence (1973) and Zahavi (1975). Phenomena from economic interactions, to sexual selection, to predator-prey signaling, to parent-offspring conflict have been studied under this heading (Searcy and Nowicky, 2005). Theoretical models of these situations share some important features. In such partial conflict of interest signaling models, two players have the option to transfer information. A sender has a certain type, and has the option to either send a signal to an uninformed receiver about this type, or not to. The sender has an incentive to sometimes, but not always, reveal the information to the receiver, whereas the receiver would always like to be fully informed. The models show that, whenever these requirements hold, there is no reliable information transfer between sender and receiver unless signals are 'costly', meaning that the sender must pay something to send them.

The game shown in figure 1 illustrates this idea. This is taken from Zollman et al. (2013), and is also the game employed in the experiments we describe below.[4] In figure 1 you can see the extensive form representation of a situation of communication with potential conflicts of interest. The first move is made by 'nature' who chooses whether the sender is of type T_1 or type T_2. The sender can then either send a signal or abstain from doing so. The cost of the signal varies with the type of the sender: c_1 if the sender is of type T_1 and c_2 if she is of type T_2. The receiver observes the signal, but cannot observe the type of the sender. She can choose between two actions, A_1 and A_2.

Players' incentives are aligned, in this game, if the sender is of type T_1, and they are misaligned otherwise. When the sender is of type T_1, then both players prefer A_1 over A_2. When the sender is of type T_2, however, the receiver prefers choosing A_2 over choosing A_1 while the sender still wants her to choose A_1.

In table 1 we list all the pure strategies of this game. If the sender chooses strategy S_1 and the receiver chooses strategy R_1, the signal carries information about sender type. In this case, senders signal only when type T_1 and receivers only choose A_1 when they receive a signal. This strategy profile is not a Nash

[4] Though, as we will outline, we must shift the payoffs of the game to accord with experimental practice.

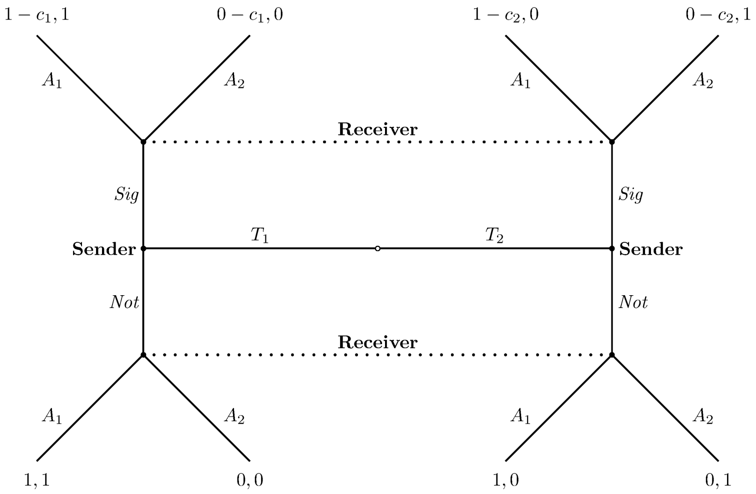

Fig. 1 A partial conflict of interest signaling game with differential costs.

Label	Description
S_1	Signal if T_1 and don't signal if T_2
S_2	Signal always
S_3	Never signal
S_4	Signal if T_2 and don't signal if T_1
R_1	A_1 if signal is observed, A_2 otherwise
R_2	A_2 always
R_3	A_1 always
R_4	A_2 if signal, A_1 otherwise

Table 1 All possible pure strategies in game pictured in figure 1 for senders, S, and receivers, R.

equilibrium when signals are cheap, e.g. $c_1 = c_2 = 0$. Since receivers choose A_1 in response to the signal, then senders also want to use the signal when type T_2 in order to obtain their preferred outcome. Thus, the sender has an incentive to choose another strategy.

For certain higher values of c, S_1 and R_1 will be a Nash equilibrium. As long as the type T_1 pays a signal cost of $c_1 < 1$, it is strictly in her interest to signal in order to ensure that the receiver takes action A_1. Also, as long as type T_2 pays a cost $c_2 > 1$ for signaling, then it is strictly in her interest not to signal; the cost of the signal outweighs the benefit obtained by getting the receiver to choose A_1. Hence, if

$$c_1 < 1 < c_2 \tag{1}$$

the strategy profile where the sender chooses S_1 and the receiver chooses R_1 is a Nash equilibrium. This is often called a 'separating equilibrium'. It is also known as a 'costly signaling equilibrium' since it is the fact that c_2 is sufficiently high that allows reliable signaling to be stable.

These considerations lead to the 'costly signaling hypothesis' mentioned above: In situations of partial conflict of interest, informative signaling is possible only if there are signals of sufficiently high costs for some types. There is an important connection between this hypothesis and Grice's 'Cooperative Principle', as discussed in the introduction. The effect of introducing costs is to align the interests of the players. In order to see what we mean by this, consider again the game in figure 1. As long as both costs are equal to zero, there is conflict of interest between receivers and type T_2 senders. However, if (1) holds, the preferences of the two players align and so the sender wants to reveal her type. That is, the sender prefers to signal when type T_1 and not to signal when T_2 (strategy S_1). This is the sense in which the Cooperative Principle holds in costly signaling games: when costs are high senders prefer a strategy that maximizes the ease and effectiveness of communication.[5]

Besides the signaling equilibrium, there are two further types of equilibria in costly signaling games. There are always pooling equilibria where the sender never uses the costly signal. While these equilibria are interesting and important, they will not play a significant role in our experiment, which was designed to investigate the other type of equilibrium, known as a 'hybrid equilibrium'. The hybrid equilibrium for the game of figure 1 is shown in figure 2. In the hybrid equilibrium, a sender always sends the signal if she is of type T_1. Otherwise, she sends the signal with probability α and does not send the signal with probability $1-\alpha$. The receiver always chooses A_2 upon not receiving the signal. If she receives the signal, then she chooses A_1 with probability β and A_2 with probability $1-\beta$. Hence, the hybrid equilibrium is a mixed equilibrium where the sender mixes between strategies S_1 and S_2 and the receiver mixes between R_1 and R_2. It can be shown that the hybrid equilibrium exists whenever

$$0 < c_2 < 1 \quad \text{and} \quad c_1 \leq c_2. \tag{2}$$

In other words, when the cost to T_2 is less than one, but the cost to T_1 is even less than this, the hybrid equilibrium will exist. In this case, the hybrid equilibrium is located at $\beta = c_2$ and $\alpha = x/(1-x)$, where x is the prior probability of type T_1 (see Zollman et al., 2013).[6]

The idea of the hybrid equilibrium is that the sender sometimes signals reliably and sometimes she does not. In order to account for the latter part of the sender's overall strategy, the receiver does not always choose the preferred action of the sender (A_1). Thus, there is information transfer between the players at the hybrid equilibrium, but it is not perfect. Importantly, information transfer is possible even though the cost c_2 is too low to align the players' interests. The hybrid equilibrium is a case where the costly signaling

[5] Of course, as noted, sender and receiver interests are not perfectly aligned over possible receiver strategies. Senders prefer that the receiver always take action A_1 (strategy R_2), while receivers prefer to only take action A_1 when senders are type T_1 (strategy R_1). What is important is that the cost of the signal aligns sender and receiver interest over the sender's strategy, ensuring that signals perfectly communicate sender type.

[6] Prior probability here refers to the likelihood that a sender will be of type T_1. Here this probability will be hashed out as the proportion of T_1 types in the experimental population.

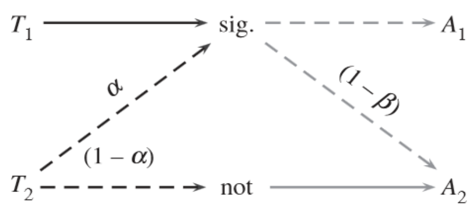

Fig. 2 Illustration of the hybrid equilibrium.

hypothesis does not hold and the Cooperative Principle fails. Yet there still is information in the signal sent by the sender.

3 Evolution and the Hybrid Equilibrium

Standard costly signaling equilibria face a number of problems. Some of them are empirical. When one measures the actual costs in biological scenarios of sending purportedly costly signals, they often turn out to be negligible.[7] There is also theoretical work that weighs against the significance of costly signaling equilibria (Huttegger and Zollman, 2010; Wagner, 2013; Zollman et al., 2013).[8] In particular, the costly signaling equilibria of many different costly signaling games does not seem to be very significant from the point of view of the evolutionary replicator dynamics. The replicator dynamics is a simple system of ordinary differential equations describing a selection process among strategies of a game. Under these dynamics, strategies are played by members of a very large population who repeatedly interact at random with each other. Strategies with an above average payoff increase in frequency, while those with a below average payoff decrease in frequency (for details, see Hofbauer and Sigmund, 1998).

For various costly signaling games (including the game of figure 1), the hybrid equilibrium fares better under the replicator dynamics (Huttegger and Zollman, 2010; Wagner, 2013; Zollman et al., 2013). This has to do with the fact that the costs required for the hybrid equilibrium can be quite small—a fact that is also relevant from an empirical standpoint, as these small costs are more in line with observed costs of real world signaling in many cases. For the replicator dynamics, the hybrid equilibrium is dynamically stable and attracts a significant portion of initial conditions of the state space. In short,

[7] See Searcy and Nowicky (2005) as well as references in Zollman et al. (2013).

[8] These theoretical issues arise from dynamical considerations and, as such, have been ignored when only the equilibrium properties of a game are analyzed. For a methodological discussion of the equilibrium analysis versus dynamical analysis see Huttegger and Zollman (2013). For an example of a case where equilibrium analyses are misleading see O'Connor (2015).

this means that many evolving populations should be expected to arrive at a state where all actors play the hybrid equilibrium. While the replicator dynamics makes specific assumptions about the underlying dynamical process, its qualitative features hold across a large number of dynamical models for evolution and learning. We therefore take these results about the hybrid equilibrium as providing theoretical predictions that will hold qualitatively across many cases of learning and evolution.

As will become clear below, in our experiment, subjects have an opportunity to learn to play the game in figure 1 with a group. We predict similar dynamical outcomes for this evolutionary scenario. While (1) holds, the interests of receivers and senders of type T_2 will be sufficiently aligned to allow for the signal to communicate sender type. When (2) holds, the interests of receivers and type T_2 senders are not aligned, but we predict partial communication of sender type.

In the remainder of the paper we study the significance of the hybrid equilibrium from an empirical perspective.

4 Experimental Set-Up

Subjects in our experiment played the differential cost game in figure 1. The payoffs shown in this figure were chosen for ease of explanation. For the experiment, as described below, these payoffs had to be modified slightly, but the structure of the game was maintained. The experiment consisted of both an experimental and control treatment which differed in the cost of signaling for type T_2 and sender's payoff for A_2. In the experimental treatment, these values were such that the interests of receivers and type T_2 senders were not aligned. In the control, these values were such that the interests of receivers and type T_2 senders were sufficiently aligned to allow for full communication. These treatments will be described in more detail below.

There were a total of 12 sessions (eight sessions of the experimental treatment and four sessions of the control treatment) each of which involved 12 participants. The subject pool consisted of undergraduate and graduate students from the University of California, Irvine who were recruited from the Experimental Social Science Laboratory subject pool via email solicitation. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

At the start of each session, experimental subjects were asked to sit at a randomly assigned computer terminal where they were presented with a set of instructions. The set of instructions provided subjects with knowledge of the game and the payment structure employed. These instructions were designed to give players only enough knowledge of the experimental set-up to make strategic decisions.[9] Deviations from complete knowledge of the game will be noted as the experimental set-up is described below.

In each session, six participants were randomly assigned to be senders (referred to as 'Role 1' in the experiment) and six to be receivers (referred to as 'Role 2'). Of the senders, two were assigned the type T_1 (referred to as 'Blue') and four were assigned the type T_2 (referred to as 'Red'). Receivers were aware that there were two possible sender types, but were unaware of the proportion of types within the sender population. Senders were aware that there may be other types within the sender population, but were not given any information about the other type.

Each session consisted of 60 rounds. In every round, each sender was randomly paired with a receiver. Each round consisted of two stages. In the first stage, each sender was asked if they would like to signal to the receiver. The signal was the "!" symbol, which was chosen so as to not have any intuitive connection with sender type. For type T_1, this signal was costless. For type T_2, the signal cost was 1 during the experimental treatment and 2 during the control. Each sender type was aware of the cost for their type, but not aware of the cost for the other type. Receivers were not aware of the signal costs.

In the second stage, receivers were told whether the sender had sent the "!" signal or not and were then asked to choose action A_1 or A_2 (described as guessing the sender was Blue or Red, respectively). Receivers got a payoff of 3 if they choose the action A_1 when matched with type T_1 sender or if they chose A_2 when matched with a type T_2 sender, and a payoff of 0 otherwise. Senders received a payoff of 3 when receivers chose A_1 and a lower payoff when receivers chose A_2. In the experimental treatment, the sender's payoff for A_2 was 1 and in the control the payoff was 2. If a type T_2 sender chose to send a signal, the cost of the signal was subtracted to yield a final payoff for the round. Each participant was only aware of their possible payoffs, not of the payoffs for other roles or types.

As noted above, these costs are slightly different than those shown in figure 1, though the structure of the game is the same. In the experimental treatment, the cost for type T_2 to signal and the payoff to sender type T_2 for A_2 were both 1. These numbers were chosen to avoid negative payoffs (type T_2's minimum payoff occurs when they signal, but the receiver takes action A_2, which ends up being 0). The same reasoning applies in the control, where the cost of signaling for type T_2 and their payoff for A_2 are both 2.

For the experimental treatment, the potential benefit for the sender of the receiver choosing A_1 rather than A_2 was 2 (a payoff of 3 verses a payoff of 1) whereas the cost of signaling for type T_2 was 1. This means that in the experimental treatment, type T_2 senders could benefit from signaling given that there was a sufficiently large chance of receivers choosing A_2 upon receipt of a signal. Since receivers would benefit from type T_2 never signaling, their interests were not aligned. For the control, the potential benefit for senders

[9] This choice is meant to induce a situation where actors are learning from experience, rather than using high rationality strategies to choose how to behave. See Bruner et al. (2015) for further justification of this choice in a similar experimental setting.

of receivers choosing A_1 rather than A_2 was 1 (3 verses 2) while the cost of signaling for type T_2 was 2. In this treatment, it was in type T_2's interest to never signal. Since it is also in the receiver's interest for type T_2 to never signal, their interests were aligned.

At the end of each round, participants were given a summary of the round. They were told the type of the sender, whether or not a signal was sent, what action the receiver chose, and their own payoff for the round. Subjects were not told the payoffs for any other participants or what occurred among any other sender-receiver pairs.

Subjects received a \$7 show-up fee for attending the experiment. In addition, they were paid for three randomly selected rounds of the experiment. Subjects earned \$1 for each point they received in these randomly selected rounds. These rounds were not chosen from the first 10 rounds in order to allow time for learning. This payment structure allowed participants to make up to \$9 in addition to the \$7 show-up fee, for a total of \$16 maximum. This method of payment was designed to minimize both risky (non-optimal) behavior and wealth accumulation effects (see Bruner et al. (2015) for details on this sort of payment structure). Subjects were paid in cash immediately following each session.

5 Results

Given the set-up described above, we expect that in the experimental treatment groups will evolve toward the hybrid equilibrium and groups in the control treatment will evolve toward the separating (costly signaling) equilibrium. We use two steps to determine whether the results are consistent with this prediction.

First, we compare the results from the experimental and control treatments. The goal here is to use the control treatment as a baseline to establish that, in fact, the experimental subjects are transferring information less perfectly than their counterparts. In particular, we will see that in the control treatment both signals and non-signals carry near perfect information about sender type, while in the experimental treatment there is near perfect information about type when the signal is absent, but not when the signal is sent.

Second, we determine if information *is* in fact being transferred when the signal is sent in the experimental treatment. To perform this second step, we compare the experimental treatment to a null hypothesis that the experimental actors are failing to transfer information at all. In particular, we check whether there is any correlation between sender types and signaling or between receipt of a signal and receiver's guess of sender type. If a sender of type T_1 is more likely to signal than type T_2, and receivers in turn are more like to take action A_1 when the signal is present than when it is absent, we can conclude that the signal is partially informative.

In making these comparisons, we use the average behavior of groups. Since we are testing whether groups will reach the hybrid equilibrium by the end of the experiment, we look at data from round 50 to 60 of the experiment. We use t-tests to evaluate the significance of our results. In a t-test, the behavior of each group is treated as a data point and the variance among these data points is taken into account.

5.1 Comparison to Control

Recall that there are two ways the hybrid equilibrium differs from the separating equilibrium. First, while type T_2 will never signal in the separating equilibrium, they will sometimes signal in the hybrid equilibrium. Second, while receivers will always take action A_1 in response to a signal in the separating equilibrium, they will sometimes take action A_2 in response to a signal in the hybrid equilibrium. Otherwise, the predictions for both treatments are the same. To test whether the results of the experimental treatment differ as predicted from the separating equilibrium, we compare to the control treatment. This gives a more accurate picture of which deviations from perfect communication are due to subjects making occasional errors and which can be attributed to the existence of the hybrid equilibrium.

	T_1 signals	T_2 does not signal (1-α)	A_1 taken after signal (β)	A_2 taken after no signal
Control	93.2	96.0	79.8	91.5
Experimental	90.4	71.9	63.4	84.8
Significance	0.24	**0.0014**	**0.052**	0.055

Table 2 A comparison of the experimental and control treatments. Percentages and p-values are shown. $1 - \alpha$ and β are defined in figure 2.

Prediction 1 (Sender Behavior): There will be no difference between the control and experimental treatments for type T_1 choosing to signal. In the control treatment type T_2 will never signal and in the experimental treatment type T_2 will sometimes signal.

We performed a one-tailed t-test to determine whether type T_1 signaled significantly less often in the experimental treatment than in the control. As table 2 shows, we find no significant difference between control and experimental treatments for type T_1 senders choosing to signal.

We again performed a one-tailed t-test to determine whether type T_2 signaled significantly more often in the experimental treatment than in the control. The result is significant at the < 0.002 level, as seen in table 2. Figure 3 shows the percentage of type T_2 signalers that do not signal in both the experimental and control treatments. Data points were calculated by determining the percentage of type T_2 signalers that fail to send the signal in the span of ten rounds. As figure 3 indicates, in the control treatment signalers of type T_2 quickly learned not to send the signal. Signalers of type T_2 in the experimental treatment, on the other hand, increased the rate at which they sent the signal until round 40, at which point they failed to send the signal approximately 70% of the time for the remainder of the experiment. To be perfectly clear, this behavior clearly accords with our predictions. We now turn our attention to the receiver's response to the signal.

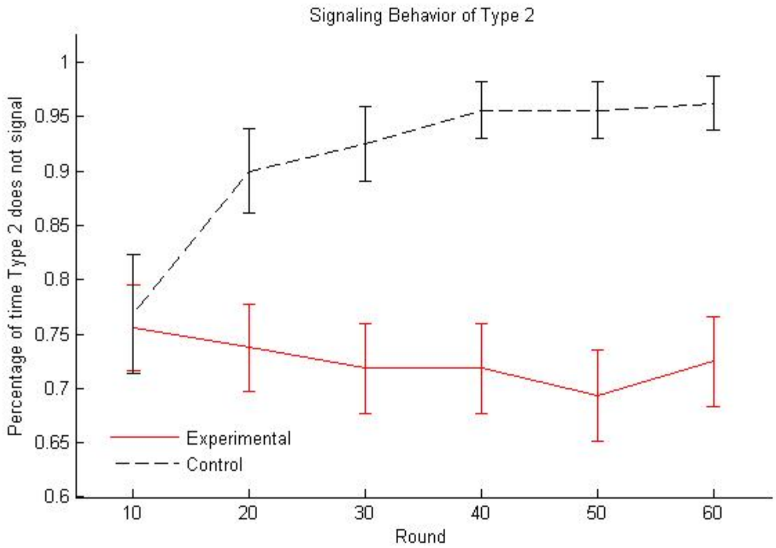

Fig. 3 Percentage of time type T_2 senders do not signal for both experimental and control treatments. Results were averaged over four runs for the control treatment and five runs for the experimental treatment. Data points are calculated for every ten rounds.

Prediction 2 (Receiver Behavior): In the control treatment receivers will always take action A_1 in response to the signal and in the experimental treatment receivers will only sometimes take action A_1 in response to the signal. There will be no difference between the control and experimental treatments for receivers taking A_2 upon receipt of the signal.

Qualitatively, the results accord with prediction in that receiver behavior differed more significantly in response to A_1 and less significantly in response to A_2 across the treatments. In particular, receivers were about 16 percentage points less likely to take A_1 in response to the signal in the experimental treatment and only about 7 percentage points less likely to take action A_2.

We performed a one-tailed t-test to determine whether receivers took action A_1 in response to the signal significantly less often in the experimental treatment than in the control. As seen in table 2, we are not able to conclude that this difference is significant, although the difference is close to being significant at $p \approx 0.052$. Similarly, a one-tailed t-test shows that difference for receivers taking A_2 when there is no signal is not significant, but is close to being significant at $p \approx 0.055$.

We cannot conclude from the significance tests that Prediction 2 is supported, but we can see from figure 4 that there are two different trends in receiver response to the signal. Figure 4 displays the percentage of the time receivers took action A_1 conditional on the sender having sent the signal. In the control treatment, the graph shows an upward trend as receivers learned to take action A_1 in response to the signal. This upward trend was not observed in the experimental treatment.

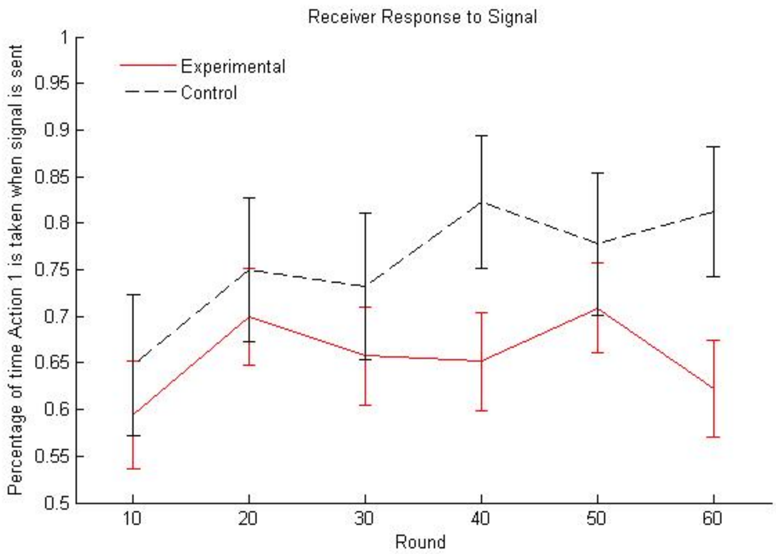

Fig. 4 Percentage of time receivers take action A_1 in response to the signal for both control and experimental treatments. Results were averaged over four runs for the control treatment and five runs for the experimental treatment. Data points are calculated for every ten rounds.

Generally, across treatments we found that senders tended to learn a signaling strategy first and receivers learned to respond more slowly. Comparing figures 3 and 4 shows that receiver behavior was much more varied than sender behavior. Thus, there is reason to think that receivers were still learning when the experiment ended.

For this reason, we provide figure 5. This figure shows the behavior we would expect the receivers to arrive at if they were to continue learning in the same fashion for another 60 rounds.[10]

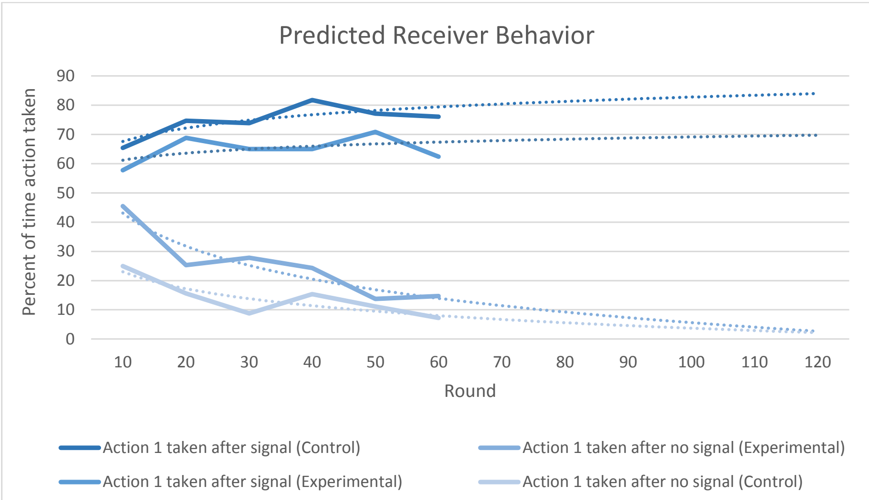

Fig. 5 Trend lines extending receiver behavior to 120 rounds.

We can see from figure 5 that, using trendlines, in the control we predict receivers will continue taking action A_1 more often in response to the signal than in the experimental treatment, while in both treatments we predict that receivers will learn to take action A_2 in absence of the signal.

To summarize, we see a significant difference in sender behavior across control and experimental treatments with the experimental treatments better conforming to the hybrid equilibrium. We do not see a significant difference in receiver behavior across treatments, though if we extrapolate observed learning trends we predict that such a difference would arise.

5.2 Comparison to Independence

The second step in determining whether results are consistent with the hybrid equilibrium predictions is to check whether there is still some information transferred when the signal is sent.

[10] A trend line is constructed by first using a regression on the data for the first 60 rounds to find the equation that best describes the receiver's learning behavior. This equation is then used to predict receiver learning for the next 60 rounds. We found that a logarithmic regression best describes receiver learning in our experiment in terms of providing the largest R^2 values. This indicates that receivers learn quickly at first, then slow down over time.

Prediction 3 (Information Transfer): The presence of a signal will contain some information about sender type in experimental treatments.

The most natural way of determining whether this prediction is confirmed is to compare the experimental results with a null hypothesis. In this case, the null hypothesis is that there is no correlation between sender type and signaling, and that there is no correlation between signal and receiver choice.[11] We predict that, in fact, the signal is sent more frequently by type T_1 and that upon receipt of the signal, receivers are more likely to take action A_1.

Again taking data from the rounds 50 to 60, we performed a one-tailed t-test to see whether type T_1 is more likely to send a signal than T_2. There is very strong evidence that sending a signal is dependent on sender type. (This result is significant at the $<< 0.0001$ level.) We can conclude that the signal contains information about sender type: receipt of the signal means it is more likely that a sender is type T_1.

We also test whether receivers are sensitive to the information contained in the signal, or in other words that there is some dependence between receipt of a signal and action taken. In order to determine whether this is the case, we compare observed receiver behavior with what a receiver would do if ignoring the signal. Since there is evidence that subjects in the laboratory setting use probability matching strategies, we assume that if receivers are ignoring the signal they take action A_1 one-third of the time.[12] We use a one-tailed t-test to determine whether receivers took action A_1 upon receipt of the signal more than a third of the time and find that the result is significant at the $<< 0.0001$ level.

6 Conclusion

To sum up, we find that under parameter values where the hybrid equilibrium exists, groups of actors do, in fact, learn to send partially communicative signals. This result is consistent with what we see in models of such scenarios—a significant portion of the time, hybrid equilibria evolve. As our results indicate, communication in humans can occur even when interests do not coincide. This gives reason to think that the Cooperative Principle is a useful idea for conceptualizing some human communication, but by no means all.

One thing that should be noted, though, is that the purpose of the Cooperative Principle in pragmatics is to provide background assumptions that allow humans to read implicature from human conversation. For example, if you ask, ‘Is Gertie a good worker?’ and I say ‘Gertie has done some good things since working here’, the implication is that Gertie has also done some less-than-good things. The reason that we can draw this conclusion has to do

[11] For more on the use of this comparison see Blume et al. (1998) and Bruner et al. (2015).

[12] For a discussion of the extent to which subjects use probability matching, see Vulkan (2000). The alternative assumption, that receivers would take action A_2 100% of the time, would only make the comparison to independence results stronger, since we would be asking if the observed frequency is greater than zero rather than one-third.

with the assumption that I am generally trying to provide the right amount and right kinds of information to you.

In recent years, though, it has become clear that pragmatic analysis can be done even in conflict of interest cases. De Jaegher and van Rooij (2014) provide an overview of work on pragmatics in conflict of interest signaling that uses game theory as a framework. Franke et al. (2012) include an analysis of a number of cases where pragmatic inference is possible in conflict of interest cases. For example, consider a case where you ask, 'What time is it?' and I reply 'Sometime after noon.' Under the Cooperative Principle, it is safe to assume that I do not know the time because otherwise I would provide you with further information. If, however, you know that I have access to the time, you can instead infer that I am unwilling to share it for some reason.

Our results also lends credence to work by previous authors arguing for the evolutionary importance of hybrid equilibria (Huttegger and Zollman, 2010; Wagner, 2013; Zollman et al., 2013). In doing so, it may give economists and biologists a reason to take this sort of signaling outcome more seriously.

This paper is part of a small but growing body of work employing the methods of experimental economics to study questions of interest to philosophers. Bruner et al. (2015) and Rubin et al. (2015) use these methods to investigate the emergence of communication in human groups. We follow these authors in thinking that these methods can be of great use to experimental philosophers, especially in cases where philosophers already employ game theory as a framework for understanding strategic interaction in humans.

References

Barrett, Jeffrey and Kevin JS Zollman (2009). "The role of forgetting in the evolution and learning of language." *Journal of Experimental & Theoretical Artificial Intelligence*, *21*(4), 293–309.

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*, 1323–1340.

Bruner, Justin, Cailin O'Connor, Hannah Rubin, and Simon Huttegger (2015). "David Lewis in the Lab." *Synthese*.

De Jaegher, Kris and Robert van Rooij (2014). "Game-theoretic pragmatics under conflicting and common interests." *Erkenntnis*, *79*(4), 769–820.

Fischbacher, Urs (2007). "z-Tree: Zurich Toolbox for Ready-made Economics Experiments." *Experimental Economics*, *10*(2), 171–178.

Franke, Michael, Tikitu De Jager, and Robert Van Rooij (2012). "Relevance in cooperation and conflict." *Journal of Logic and Computation*, *22*(1), 23–54.

Grice, Paul (1991). *Studies in the Way of Words*. Harvard University Press.

Hofbauer, Josef and Karl Sigmund (1998). *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press.

Huttegger, Simon M and Kevin J. S. Zollman (2013). "Methodology in Biological Game Theory." *The British Journal for the Philosophy of Science*, 637–658.

Huttegger, Simon M and Kevin JS Zollman (2010). "Dynamic stability and basins of attraction in the Sir Philip Sidney game." *Proceedings of the Royal Society of London B: Biological Sciences*, *277*(1689), 1915–1922.

O'Connor, Cailin (2014). "The evolution of vagueness." *Erkenntnis*, *79*(4), 707–727.

O'Connor, Cailin (2015). "Evolving to Generalize: Trading Precision for Speed." *The British Journal for the Philosophy of Science*.

Rubin, Hannah, Lindley Slipetz, and Cailin O'Connor (2015). "Structure and Signaling: a Sim-Max Game Experiment.".

Searcy, W. A. and S. Nowicky, editors. *The Evolution of Animal Communication*. Princeton: Princeton University Press.

Skyrms, Brian (2010). *Signals: Evolution, learning, and information*. Oxford University Press.

Spence, Michael (1973). "Job market signaling." *The quarterly journal of Economics*, 355–374.

Vulkan, Nir (2000). "An economist's perspective on probability matching.." *Journal of Economic Surveys*, *14*(1), 101–118.

Wagner, Elliott O (2013). "The dynamics of costly signaling." *Games*, *4*(2), 163–181.

Zahavi, A. (1975). "Mate Selection – The Selection of a Handicap." *Journal of Theoretical Biology*, *53*, 205–214.

Zollman, K. J. S., C. T. Bergstrom, and S. M. Huttegger (2013). "Between Cheap and Costly Signals: The Evolution of Partially Honest Communication." *Proceedings of the Royal Society London, B*, *280*, 20121878.