Studies in History and Philosophy of Biol & Biomed Sci 83 (2020) 101295

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biol & Biomed Sci

journal homepage: www.elsevier.com/locate/shpsc

Communication without common interest: A signaling experiment

Hannah Rubin[a,*], Justin P. Bruner[b], Cailin O'Connor[c], Simon Huttegger[c]

[a] Department of Philosophy, University of Notre Dame, United States
[b] Department of Political Economy and Moral Science, University of Arizona, United States
[c] Department of Logic and Philosophy of Science, University of California, Irvine, United States

ABSTRACT

Communication can arise when the interests of speaker and listener diverge if the cost of signaling is high enough that it aligns their interests. But what happens when the cost of signaling is not sufficient to align their interests? Using methods from experimental economics, we test whether theoretical predictions of a partially informative system of communication are borne out. As our results indicate, partial communication can occur even when interests do not coincide.

1. Introduction

David Lewis introduced a signaling game in his 1969 book *Convention*, with the goal of explaining how linguistic conventions are established. One of the basic assumptions of Lewis signaling games is common interest between senders and receivers (Lewis, 1969). Under this assumption the sender wants to convey as much information as possible to the receiver, and the receiver wants to choose acts that are beneficial for both players. The only problem to be solved here is the assignment of conventional links between states and acts, on the one hand, and signals, on the other.

The Lewisian setting is important because communicators are often involved in a cooperative endeavor. Two partners developing a school project, for example, are trying to communicate effectively to achieve mutually shared ends. However, communicative situations do not always involve this level of common interest. This raises the question of whether, or to what extent, successful communication remains feasible when the interests of senders and receivers come apart. Such situations are common, for example, between job hunters and companies looking to hire and between firms planning to go public and potential stock holders. Outside the realm of economic behavior, one can find many further examples—people on first dates, teenagers and their parents, and students and their teachers. There are many examples from non-human animals as well—interests are not always aligned in interactions between predators and prey, potential mates, and parents and offspring.

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—taking us back, effectively, to Lewis' setting. This means that rational actors will be willing to transfer information.

But are high costs always necessary to allow information transfer? Recently, scholars have been able to show that a type of partially communicative equilibrium, usually ignored in biology and economics, arises commonly when actors with divergent interests learn to communicate (Huttegger & Zollman, 2010; Wagner, 2013). Throughout the paper, we will refer to these partially communicative outcomes as "hybrid equilibria".[1] Importantly, in these hybrid equilibria costs to signalers are low, and do not bring the interests of the actors in line, yet some level of communication is still possible.

The existence of hybrid equilibria provides a partial answer to the question of what happens when Lewis' common interest assumption is dropped. Under certain conditions, the presence of divergent interests does not entail that no communication is taking place, only that communication is imperfect.

This is by now a well established theoretical finding. The aim of this paper is to 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 is perhaps surprising because actors learn to communicate even though their interests are misaligned. On the other hand, the experimental outcomes are in line with the predictions provided by evolutionary game theory. 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 the broader implications of our findings.

* Corresponding author.
E-mail address: hannahmrubin@gmail.com (H. Rubin).

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

https://doi.org/10.1016/j.shpsc.2020.101295
Received 8 September 2019; Received in revised form 7 February 2020; Accepted 1 May 2020
Available online 02 July 2020
1369-8486/

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 examined under this heading (Searcy & Nowicky, 2005). Theoretical models of these situations make use of what are called signaling games, and share some important features. In such models, two players—a sender and a receiver—can transfer information. The sender has a certain type, and can either send a signal to the receiver about this type, or not. The sender is sometimes, but not always, incentivized to reveal their type 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 at least some senders must pay something to send them.

To give an example of a case where such a model applies, imagine a population of job candidates communicating with a company. Some of them are qualified and some are not (these are their types). The company would like to know the truth about their qualifications, but all the senders want to be judged as high quality. For this reason, the company cannot necessarily trust their signals about their own quality. Suppose though that it is very difficult for low quality candidates to complete a college degree, i.e., it is costly. If it is difficult enough, they will not be willing to earn the degree, even if it would get them a job. High quality candidates, on the other hand, will be willing to pay a relatively low cost to earn the degree. The company, upon observing the degree, can then trust that a candidate is high quality.

The game shown in Fig. 1 illustrates this sort of scenario.[2] It is the extensive form of the game employed in the experiments we will describe below. (Though, as we will outline, we must shift the payoffs of the game slightly to accord with experimental practice.) This tree should be read from the central node outward. The first move is made by 'nature' who chooses whether the sender is of type T_1 or type T_2. The sender then chooses to 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. Payoffs are shown at the final nodes, with the sender listed first. The receiver gets 1 for correctly guessing the sender type, and 0 otherwise. The sender gets 1 whenever the receiver guesses A_2, minus any costs for sending the signal. Players' incentives are thus aligned, in this game, if the sender is of type T_1, and they are misaligned otherwise. Similarly, high quality candidates and companies have aligned interests, but low quality candidates' interests are misaligned.

In Table 1 we list all the pure strategies of this game, i.e., choices for senders and receivers. If the sender chooses strategy S_1 and the receiver chooses strategy R_1, the signal carries perfect information about sender type. In this case, senders signal only when they are type T_1 and receivers only choose A_1 when they receive a signal. This strategy profile is not a Nash equilibrium when signals are cheap, e.g. $c_1 = c_2 = 0$. This is for the reason described above. If receivers are choosing A_1 upon receipt of the signal, type T_2 senders will start signaling.

For certain signal costs, though, 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.

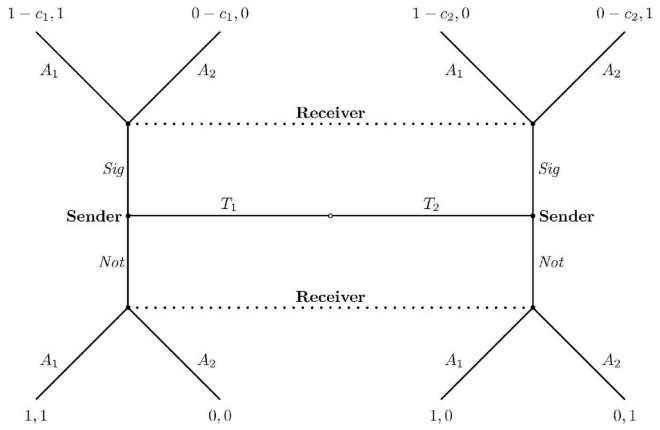

Fig. 1. A partial conflict of interest signaling game with differential costs.

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

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

This observation leads 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. Notice that the effect of introducing costs is to align the interests of the players. 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 in that the sender prefers to act in a way that reveals her type, and the receiver wants her to do so.[3] Similarly, with companies and job hunters, when college is difficult enough, low quality candidates prefer to reveal their type by not going to college. This is just what companies want them to do.

Besides the costly signaling equilibrium, there are two further types of equilibria in costly signaling games. The first are pooling equilibria where the sender never sends a signal regardless of type, meaning that no information is ever transferred. 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 the 'hybrid equilibrium'. The hybrid equilibrium for the game of Fig. 1 is shown in Fig. 2. In it, the sender always sends the signal if she is of type T_1. Otherwise, if she is type T_2 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

[2] This version of the game is taken from Zollman et al. (2013).

[3] 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.

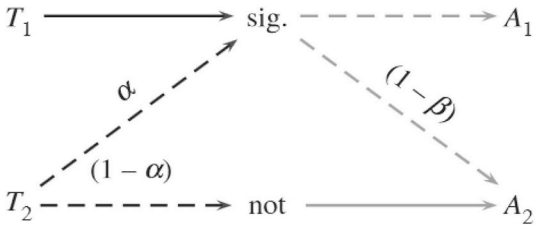

Fig. 2. Illustration of the hybrid equilibrium.

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

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

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. For this game, 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, Bergstrom, & Huttegger, 2013).[4,5]

The idea of the hybrid equilibrium is that the sender sometimes signals reliably and sometimes does not. Upon receipt of the signal, there is no clear-cut way for the receiver to infer the type of the sender. In response, 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, this information transfer is possible even though the cost c_2 is too low to align the players' interests, meaning that costly signaling hypothesis does not hold.[6]

3. Evolution and the hybrid equilibrium

The costly signaling hypothesis faces a number of challenges. Some of these 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 & Zollman, 2010; Wagner, 2013; Zollman et al., 2013).[8] In particular, costly signaling equilibria do not seem to be very significant from an evolutionary point of view. The replicator dynamics is a simple system of ordinary differential equations describing a selection process among strategies of a game. Under these dynamics, 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 this reason, they have been widely used to model both biological evolution and cultural change.

For various costly signaling games, the hybrid equilibrium is often more evolutionarily significant under the replicator dynamics (Wagner, 2013; Zollman et al., 2013). In particular, costly signaling equilibria tend to have significantly smaller basins of attraction compared to the hybrid equilibrium.[9] Basins of attraction are often taken to tell us something about the evolvability of a strategy, and so this creates a worry for costly signaling hypothesis—perhaps the high costs necessary to stabilize perfect communication prevent it from evolving.[10] For the hybrid equilibrium, on the other hand, the costs can be quite small, allowing signaling to evolve. This fact 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.

We should be more precise, though, about just what the replicator dynamics predict from this game. Because the hybrid equilibrium involves mixed strategies, there are a number of strategies close to the equilibrium that garner similar payoffs for the actors where senders mix between S_1 and S_2 and receivers mix between R_1 and R_2. The replicator dynamics prediction is that, a significant proportion of the time, the population will evolve toward the hybrid equilibrium but end up circling around it indefinitely on the plane consisting of all the possible mixtures of S_1, S_2, R_1 and R_2. So, evolution leads to a state where communication is partially informative, similar to the actual equilibrium, and stays there.

Also of interest is a perturbation of the replicator dynamics, the *selection-mutation dynamics*, which has been studied for costly signaling games in Huttegger and Zollman (2016). In the replicator dynamics the hybrid equilibrium is structurally unstable, making it prone to qualitative changes due to perturbations in the dynamics. Adding mutation to the replicator dynamics has two effects: it moves the hybrid equilibrium a bit off the boundary of state space, and it lets the system converge to it. This is compatible with the broad prediction for adaptive dynamics: we expect a population of players to be somewhere close to the hybrid equilibrium (or the part of the boundary where it's located) after a sufficient number of plays.

Another note: in section 2 we explained the hybrid equilibrium in terms of a sender using the signal depending on whether they were type T_1 or T_2. In evolutionary models, in contrast, each individual has a set type and there exists a distribution of these types in the population. The hybrid equilibrium then arises at the population, rather than the individual, level. There are a few different ways this can happen. First, it might be the case that some T_2 type senders send the signal, while some do not. Second, T_2 senders could employ a mixed strategy and send the signal probabilistically. Third, the hybrid equilibrium might arise from some combination of these first two options: some T_2 types signal

[4] 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.

[5] At the hybrid equilibrium, α is such that, after updating their beliefs about the sender's type upon receipt of the signal using Bayes Rule, a receiver has the same expected payoff from taking action A_1 and A_2. Similarly, β is such that, based on the receivers strategy, type T_2 has the same expected payoff from sending and not sending the signal.

[6] The so-called Crawford-Sobel game (Crawford & Sobel, 1982) is another conflict of interest signaling game in which some communication is possible despite the fact that signals are costless. In the original version of this game, senders are assigned a private quality-type ranging from zero to twenty. Receivers aspire to correctly identify the quality-type of their counterpart. Senders, on the other hand, prefer to somewhat inflate their underlying quality and do best when receivers incorrectly classify them as being of slightly higher quality than they actually are. Crawford and Sobel prove the existence of a partially informative signaling equilibrium at which senders coarsely partition the quality-type space, sending the same signal whenever their quality falls within a specified range. Thus at equilibrium some level of communication occurs despite the fact that interests are not aligned. Senders fail to be as informative as possible, strategically obscuring their underlying quality, and their signals are ambiguous. Experimental tests have been conducted on the Sobel-Crawford game (Blume, Douglas, Kim, & Geoffrey, 2001; Cai & Wang, 2006; Dickhaut, McCabe, & Mukherji, 1995), which find that some level of information transfer is possible in the lab. Another game that exhibits partial pooling is Lachmann and Bergstrom (1998).

[7] See Searcy & 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) .

[9] Of course, the two types of equilibria will not exist for the same game. This comparison is garnered by keeping other game features the same, changing the cost c_2, and observing what happens to evolution of the system. In signaling games with a more complex structure, both equilibria may coexist; see Kane and Zollman (2015).

[10] A basin of attraction for an equilibrium is a set of population states that will lead toward an equilibrium, given that the population starts at one of those states.

probabilistically while others either always or never send the signal.[11]

In the remainder of the paper we study the significance of the hybrid equilibrium from an empirical perspective. As will become clear, in our experiment, subjects have an opportunity to learn to play the game in Fig. 1 with a group. Before moving on, though, we should say a bit about what experimental predictions we will derive from the models presented in this section. On the basis of the replicator dynamics model, one might expect laboratory subjects to engage in mixed signaling behavior that cycles regularly around the hybrid equilibrium. Or, on the basis of selection-mutation dynamics, one might predict converge directly to it. These specific predictions from specific dynamics do not always extend to other, reasonable models. The qualitative behavior, though, holds across a large number of dynamical models for evolution and learning. As will become clear, we will stick to predictions that track more qualitative features of the model. In addition, human subjects in laboratory experiments exhibit somewhat variable, stochastic behavior. For this reason, observed behavior rarely corresponds exactly to equilibrium predictions. For this reason, our predictions will focus on differences between behavior in games with or without the hybrid equilibrium that qualitatively match what we would expect.

4. Experimental set-up

Subjects in our experiment played a version of the game in Fig. 1. The payoffs shown in the figure were chosen for ease of explanation. For the experiment 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. In the experimental (or 'hybrid') treatment, payoff values were such that the interests of receivers and type T_2 senders were not aligned. In the control (or 'separating') treatment, these values were such that the interests of receivers and type T_2 senders were sufficiently aligned to allow for full communication at equilibrium. These treatments will be described in more detail below.

There were a total of 12 sessions (eight sessions of the hybrid treatment and four sessions of the separating 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.[12] 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'). This means that the proportion of high type senders was always 1/3. Receivers were aware that there were two possible sender types, but were unaware of the proportion of types within the sender population.[13] Senders were aware that there may be other types within their own 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.[14] For type T_1, the signal was costless. For type T_2, the signal cost was 1 during the hybrid treatment and 2 during the separating treatment. 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 for a correct guess, 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 hybrid treatment, the sender's payoff for A_2 was 1 and in the separating treatment the payoff was 2. 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 Fig. 1, though the structure of the game is the same. The particular values were chosen to avoid the possibility of negative payoffs, which might influence behavior. Values are summarized below:

- Experimental (Hybrid)
  - Cost of signal for T_1: 0
  - Cost of signal for T_2: 1
  - Sender payoff for A_1: 3
  - Sender payoff for A_2: 1
  - Receiver payoff for correct guess: 3
  - Receiver payoff for incorrect guess: 0
- Control (Separating)
  - Cost of signal for T_1: 0
  - Cost of signal for T_2: 2
  - Sender payoff for A_1: 3
  - Sender payoff for A_2: 2
  - Receiver payoff for correct guess: 3
  - Receiver payoff for incorrect guess: 0

For the hybrid 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 hybrid treatment, type T_2 senders could potentially benefit from signaling. And notice that since receivers would prefer that type T_2 never signal, their interests were not aligned.[15] For the separating treatment, the potential benefit for senders 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. For this reason, in the separating treatment it was never in type T_2's interest to signal. Since it was 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

[11] This third option was the most common outcome in our experiment. For instance, often two or three out of the four Low types would send the signal with some probability between 1/3 and 2/3, while the remaining Low type(s) would not send the signal at all.

[12] 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, O'Connor, Rubin, & Huttegger (2018) for further justification of this choice in a similar experimental setting.

[13] We did not want receivers to use information about sender types to decide on a strategy before engaging with the other population.

[14] This was chosen to avoid possible salience effects. For example, if the signal was the letter "B", both senders and receivers might take it to mean "Blue".

[15] The equilibrium predictions with these payoffs are $\alpha = \beta = 1/2$.

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.[16] Subjects were paid in cash immediately following each session.

5. Results

Given the set-up described above, we expect that in the experimental (hybrid) treatment groups will learn to play strategies similar to the hybrid equilibrium and in the control (separating) treatment groups will learn to play the costly signaling equilibrium.[17] Given the payoffs in our experiment, the hybrid equilibrium is at $\alpha = 0.5$, $\beta = 0.5$. Remember, however, our prediction is not that subjects will reach these exact values, or that they will neatly cycle around them. Rather we predict that they will evolve toward them, eventually reaching some combination of S_1, S_2, R_1 and R_2. In other words, subjects in the hybrid treatment will end up with the sort of partial information transfer characteristic of the hybrid equilibrium. We use two steps to determine whether results are consistent with this prediction.

First, we compare the results from the hybrid and separating treatments. The goal here is to use the separating treatment as a baseline to establish that, in fact, the experimental subjects are transferring information less perfectly than their counterparts.[18] This baseline gives a more accurate picture of which deviations from perfect communication are due to subjects making occasional errors and experiments and which can be attributed to the structure of the underlying game. In particular, we will see that in the separating treatment near perfect information is transferred about sender type, while in the hybrid treatment there is near perfect information about type when the signal is absent, but not when the signal is sent, as expected.

Second, we determine if information is in fact being transferred when the signal is sent in the hybrid treatment. To perform this second step, we compare the hybrid treatment to a null hypothesis that the 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 focus on data from round 50 to 60 of the experiment. In particular, we are interested in the proportion of times the sender sends the signal (and, likewise, the probability the receiver responds with A_1 upon receipt of the signal). Our statistical analysis proceeds as follows. We use data from the control to pin down a beta distribution. Unlike the normal distribution (which is assumed when conducting a *t*-test), the beta distribution is contained on the unit interval. We then determine the likelihood that we would observe data from our experimental treatment given the beta distribution.[19]

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.

> Prediction 1 (Sender Behavior): There will be no difference between the hybrid and separating treatments for type T_1 choosing to signal. Type T_2 will signal more often in the hybrid treatment than in the separating treatment.

We performed a test (as described above) to determine whether type T_1 signaled significantly less often in the hybrid treatment than in the separating treatment. As Table 2 shows, we find no significant difference between treatments for type T_1 senders choosing to signal.

We perform a similar test to determine whether type T_2 signaled significantly more often in the hybrid treatment than in the separating treatment. The result is significant, as seen in Table 2. This behavior clearly accords with our predictions. Fig. 3 shows the percentage of type T_2 signalers that do not signal in both the hybrid and separating 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 Fig. 3 illustrates, in the separating treatment signalers of type T_2 quickly learned not to send the signal. Signalers of type T_2 in the hybrid treatment, on the other hand, failed to send the signal somewhere around 70%–75% of the time.

We now turn our attention to the receiver's response to the signal.

> Prediction 2 (Receiver Behavior): Receivers will take action A_1 in response to the signal more often in the separating treatment than in the hybrid. There will be no difference between treatments for receivers taking A_2 when there is no signal.

Qualitatively, the results accord with this 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 hybrid treatment and only about 7 percentage points less likely to take action A_2 in absence of the signal. We determine whether receivers took action A_1 in response to the signal significantly less often in the hybrid treatment than in the separating treatment. As seen in Table 2, this difference is significant. However, we also find that the difference for receivers taking A_2 when there is no signal in the two treatments is significant, which does not accord with our prediction.

This prediction failure may have to do with learning rates for senders and receivers. Fig. 4 displays the percentage of the time receivers took action A_1 conditional on the sender having sent the signal. In the separating treatment, it appears as if there is an upward trend as receivers learned to take action A_1 in response to the signal. This upward trend does not seem to be observed in the hybrid treatment. Generally,

[16] See Bruner, O'Connor, Rubin, & Huttegger, 2018 for details on this sort of payment structure.

[17] While pooling equilibria are also a possibility in both of these treatments, we did not observe any groups reaching anything like a pooling equilibrium.

[18] One might think that we should instead compare the results from the hybrid treatment to the specific mixed strategies theoretically predicted for the hybrid equilibrium. But this does not make sense given the dynamic modeling prediction that the population might spiral around the equilibrium (Huttegger & Zollman, 2010).

[19] One might think that we should employ a binomial test because our data (*footnote continued*) are the results of binary choices made by our subjects. Binomial tests are standardly employed for experiments with independent observations of binary outcomes, like flipping two coins with unknown biases to determine whether the biases are the same. Although our data points are similar to coin flips in that we look at proportions of binary choices, unlike coin flips the observations are not independent (e.g. what one sender chooses depends on their beliefs about the receivers strategies, which depend on the strategies employed by all the senders). For this reason, we do not employ a binomial test, similar to Blume et al. (2001).

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

	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	0.998	$<<$ **0.001**	**0.0013**	0.0017

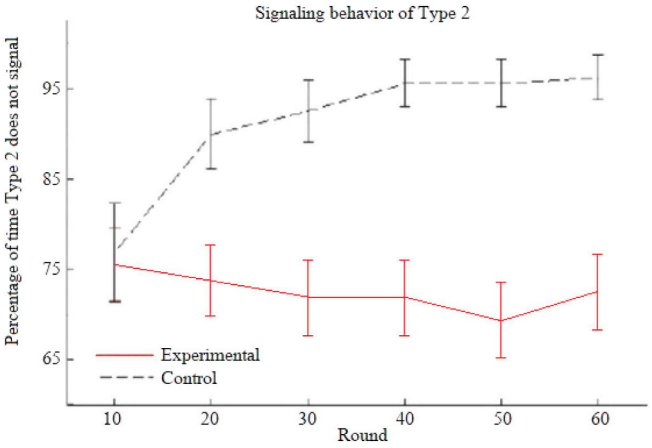

Fig. 3. Percentage of time type T_2 senders do not signal for both experimental (hybrid) and control (separating) treatments. Results were averaged over four runs for the control treatment and eight runs for the experimental treatment. Data points are calculated for every ten rounds (the current round and previous nine rounds). Error bars represent 95% confidence intervals.

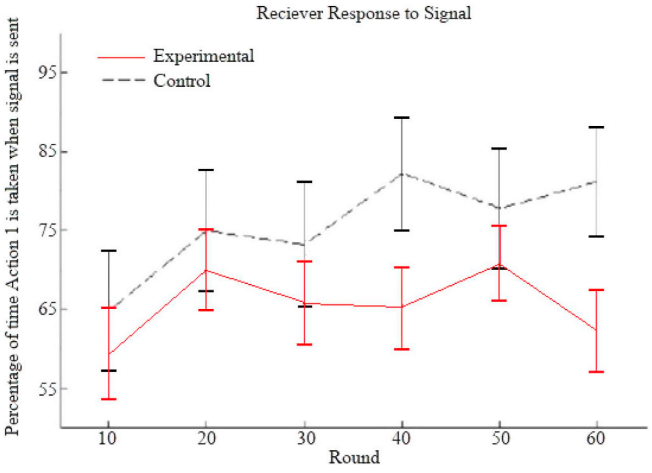

Fig. 4. Percentage of time receivers take action A_1 in response to the signal for both control (separating) and experimental (hybrid) treatments. Results were averaged over four runs for the control treatment and eight runs for the experimental treatment. Data points are calculated for every ten rounds (the current round and previous nine rounds). Error bars represent 95% confidence intervals.

across treatments we found that senders tended to learn a signaling strategy first and receivers learned to respond more slowly. In addition, if we compare Figs. 3 and 4 we see 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 Fig. 5. This figure shows the behavior we would expect the receivers to arrive at if they were to continue

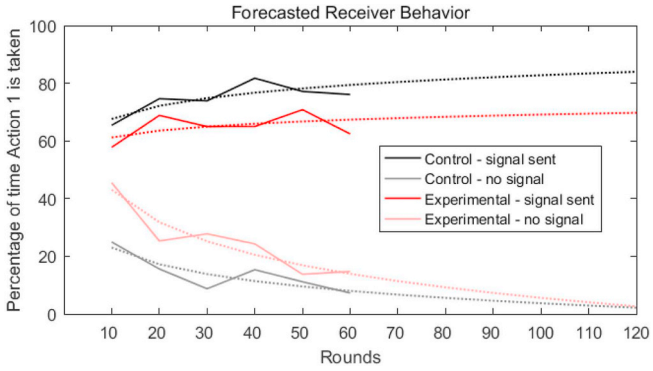

Fig. 5. Trend lines extending receiver behavior to 120 rounds.

learning in the same fashion for another 60 rounds.[20]

We can see from Fig. 5 that, using trendlines, in the separating treatment we predict receivers will continue taking action A_1 more often in response to the signal than in the hybrid 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 treatments with the hybrid 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.

> Prediction 3 (Information Transfer): The presence of a signal will contain some information about sender type in the hybrid treatment.

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.[21] 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 determine 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

[20] 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 (as compared to a linear, exponential, or polynomial regression). This indicates that receivers learn quickly at first, then slow down over time.

[21] For more on the use of this comparison see Blume, Douglas, Kim, and Geoffrey (1998) and Bruner, O'Connor, Rubin, & Huttegger (2018).

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.[22] 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

Hybrid equilibria are part of the answer to the question of what happens when one drops common interest as a basic assumption for communicative situations. In this paper, we find that under parameter values where the hybrid equilibrium exists, groups of actors do, in fact, learn to send partially communicative signals in accordance with the hybrid equilibrium. This result is consistent with what we see in models of such scenarios—a significant portion of the time, evolution leads toward the hybrid equilibrium. As our results indicate, communication in humans can occur even when interests do not coincide. Our results also lend credence to work by previous authors arguing for the evolutionary importance of hybrid equilibria (Huttegger & 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. Bicchieri and Chavez (2013) and Bicchieri and Lev-On (2007) focus on topics related to norms and ethics. Bruner, O'Connor, Rubin, & Huttegger (2018) and Rubin, O'Connor, and Bruner (2019) 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.

Grant funding

National Science Foundation grant no. EF-1038456.

Acknowledgements

The authors would like to thank Jan-Willem Romeijn, Jeff Barrett, Kevin Zollman, Brian Skyrms, Mike McBride, Ryan Kendall, Michael Caldara, and anonymous reviewers for comments, as well as the UC Irvine Experimental Social Science Laboratory staff for their help running the experiment. The paper also benefitted from feedback from audiences at: the UC Irvine Experimental Social Science Laboratory workshop; the workshop in Decisions, Games and Logic; ISHPSSB; and the Buffalo Annual Experimental Philosophy Conference.

References

Bicchieri, C., & Chavez, A. (2013). Norm manipulation, norm evasion: Experimental evidence. *Economics and Philosophy, 29*(2), 175–198.
Bicchieri, C., & Lev-On, A. (2007). Computer-mediated communication and cooperation in social dilemmas: An experimental analysis. *Politics, Philosophy & Economics, 6*(2), 139–168.
Blume, A., Douglas, V. D. J., Kim, Y.-G., & Geoffrey, B. S. (1998). Experimental evidence on the evolution of meaning of messages in sender-receiver games. *The American Economic Review,* 1323–1340.
Blume, A., Douglas, V. D. J., Kim, Y.-G., & Geoffrey, B. S. (2001). Evolution of communication with partial common interest. *Games and Economic Behavior, 37*(1), 79–120.
Bruner, J., O'Connor, C., Rubin, H., & Huttegger, S. M. (2018). David Lewis in the lab: An experimental study of signaling conventions. *Synthese, 195*(2), 603–621.
Cai, H., & Wang, J. (2006). Overcommunication in strategic information transmission games. *Games and Economic Behavior, 95*, 384–394.
Crawford, V., & Sobel, J. (1982). Strategic information transmission. *Econometrica, 50*, 1431–1451.
Dickhaut, J., McCabe, K., & Mukherji, A. (1995). "An experimental study of strategic information transmission." Econ. *Theory, 6*, 389–403.
Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economics experiments. *Experimental Economics, 10*(2), 171–178.
Hofbauer, J., & Sigmund, K. (1998). *Evolutionary games and population dynamics.* Cambridge: Cambridge University Press.
Huttegger, S. M., & Zollman, K. J. S. (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.
Huttegger, S. M., & Zollman, K. J. S. (2013). Methodology in biological game theory. *The British Journal for the Philosophy of Science,* 637–658.
Huttegger, S. M., & Zollman, K. J. S. (2016). The robustness of hybrid equilibria in costly signaling games. *Dynamic Games and Applications, 6*, 347–358.
Kane, P., & Zollman, K. J. S. (2015). An evolutionary comparison of the handicap principle and hybrid equilibrium theories of signaling. *PloS One, 10*, Article e0137271.
Lachmann, M., & Bergstrom, C. T. (1998). Signalling among relatives II: Beyond the tower of babel. *Theoretical Population Biology, 54*, 146–160.
Lewis, D. (1969). *Convention. A philosophical study*. Cambridge MA: Harvard University Press.
Rubin, H., O'Connor, C., & Bruner, J. (2019). In E. Fisher, & M. Curtis (Eds.). *"Experimental economics for philosophers." methodological advances in experimental philosophy* (pp. 175–206). London, UK: Bloomsbury Publishing.
Searcy, W. A., & Nowicky, S. (2005). *The evolution of animal communication*. Princeton: Princeton University Press.
Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics,* 355–374.
Vulkan, N. (2000). "An economist's perspective on probability matching. *Journal of Economic Surveys, 14*(1), 101–118.
Wagner, E. 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., Bergstrom, C. T., & Huttegger, S. M. (2013). Between cheap and costly signals: The evolution of partially honest communication. *Proceedings of the Royal Society of London B, 280*, 20121878.

[22] 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.