Honesty and Deception in Populations of Selfish, Adaptive Individuals

DAVID CATTEEUW and BERNARD MANDERICK

Artificial Intelligence Lab, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
E-mail: {dcatteeu,bmanderi}@vub.ac.be

Abstract

Biologists have mostly studied under what circumstances honest signaling is stable. *Stability*, however, is not sufficient to explain the *emergence* of honest signaling. We study the evolution of honest signaling between selfish, adaptive individuals and observe that honest signaling can emerge through learning. More importantly, honest signaling may emerge in cases where it is not evolutionary stable. In such cases, honesty and dishonesty co-exist. Furthermore, honest signaling does not necessarily emerge in cases where it is evolutionary stable. We show that the latter is due to the existence of other, more important equilibria and that the importance of equilibria is related to Pareto-optimality.

1   Introduction

We study the emergence of honest signaling which is a major topic in both biology and economics since the 1970's. We try to understand under which circumstances signaling between individuals with *conflicting interests* will be honest.

In economics, Spence's job market model (Spence, 1973) shows that a university degree can work as an honest signal when applying for a job since there is a cost of acquiring that degree. More importantly, the degree is increasingly more costly to acquire for less skilled employees. As such, higher skilled employees invest in a higher degree than lower skilled employees and the employer, who is unable to directly observe the employees' abilities, has good reasons to believe that job candidates with higher degrees have higher abilities. Honest signaling is important in many other economic applications with private information, such as product advertisement where the seller does and the buyer does not know the quality of the product. An overview of signaling in economics is provided by Riley (2001).

The same principle was discovered independently by Zahavi (1975) and is now known as the *handicap principle*. It states that under conflict of interest honest signaling can only be stable if signals are costly. He claims that male characteristics used for sexual selection, such as a peacock's tail, the extra large antlers of a deer, or the colorful plumage of a male bird, are honest signals of the males' quality because they are a handicap. The peacock's tail, for example, makes it harder for the peacock to escape from predators. Since only the fittest can afford the largest tails, females can reliably infer which males would make better mates from the size of their tails.

Until now, researchers of both research communities have almost exclusively relied on a static analysis of honest signaling. In a static analysis, one merely verifies the stability according to some equilibrium concept. The economics literature on signaling is mostly concerned with the necessary requirements for which honest signaling is a (unique) Nash equilibrium or a refinement thereof. See for example Cho and Kreps (1987) and Cho and Sobel (1990). Similarly, in biology, game theoretic models of the handicap principle have been used to show how honest signaling

can be an equilibrium (in this case, an evolutionarily stable strategy). See for example Grafen (1990) and Maynard Smith (1991).

A *critique* on these explanations of honest signaling, is that stability is not sufficient for honest signaling to emerge, it merely says that *if* it emerges it will persist. The same critique has been formulated by Lachmann and Bergstrom (1998) and by Huttegger and Zollman (2010). The latter study the evolutionary dynamics of the Philip Sidney game (Maynard Smith, 1991) which is the standard game theoretic model of honest signaling in biology and equally representative of signaling applications in economics. Huttegger and Zollman (2010) study the evolutionary dynamics of that game. More particularly, they apply the replicator dynamics which models how strategies may spread through an infinite, well-mixed population under influence of natural selection. They discovered that in some cases honest signaling is less likely to evolve under the replicator dynamics than is otherwise suggested by the analysis of evolutionarily stable strategies. We discuss this and other related work in more detail in Section 5.

Here, we look at a more realistic model where a finite population of *adaptive* individuals repeatedly interact. We investigate whether and when honest signaling emerges. Interactions between individuals are modeled by the Philip Sidney game. We discuss the details of this game in Section 2. The selfish, adaptive individuals are modeled by a simple reinforcement learning technique which we explain in full detail in Section 3.

Our results (Section 4) are as follows.

a) Honest signaling can emerge among selfish, adaptive individuals even under conflict of interest. It was already shown that honest signaling could emerge and persist under evolution. Here, we show that honest signaling can also emerge through *learning*.

b) Honest signaling may emerge even though it is not an equilibrium and hence cannot be predicted by a static analysis. We identify such cases and show how honesty and dishonesty co-exist.

c) The cases where honest signaling does evolve are much rarer than suggested by a static analysis and this is due to the existence of other, more important, equilibria. This was previously suggested, see Section 5. Our results show a relation between the importance of equilibria and Pareto-optimality. A Nash equilibrium is a Pareto-optimal Nash equilibrium if it is not dominated by any other Nash equilibrium. We say that equilibrium a dominates equilibrium b if at least one player is better off in equilibrium a and no one is worse off.

2   The Philip Sidney game

The interactions between individuals are modeled by the Philip Sidney game. In biology it is generally accepted as the standard model of costly signaling. The Philip Sidney game, see Figure 1, is a *signaling game*, introduced by Maynard Smith (1991). It is a two-player, extensive form game of incomplete information. The first player (Sender) can be in one of two states: *healthy* or *needy*, with probability p and $1 - p$ respectively. In both cases he can either send a signal at some cost c or he can remain silent. Player 2 (Receiver) does not know the true state of Sender, but he can observe whether or not Sender signals. Furthermore, Receiver has a resource and must decide whether or not to donate his resource to Sender.

The players' chances of survival depend on the state $t \in \{healthy, needy\}$, the signal $s \in \{signal, silent\}$, and the action $a \in \{donate, keep\}$ as follows. If Receiver keeps the resource to himself, he is sure to survive. If he donates the resource to Sender his chances of survival $v_R(t, s, a)$ are reduced to $S < 1$, see Equation (1).

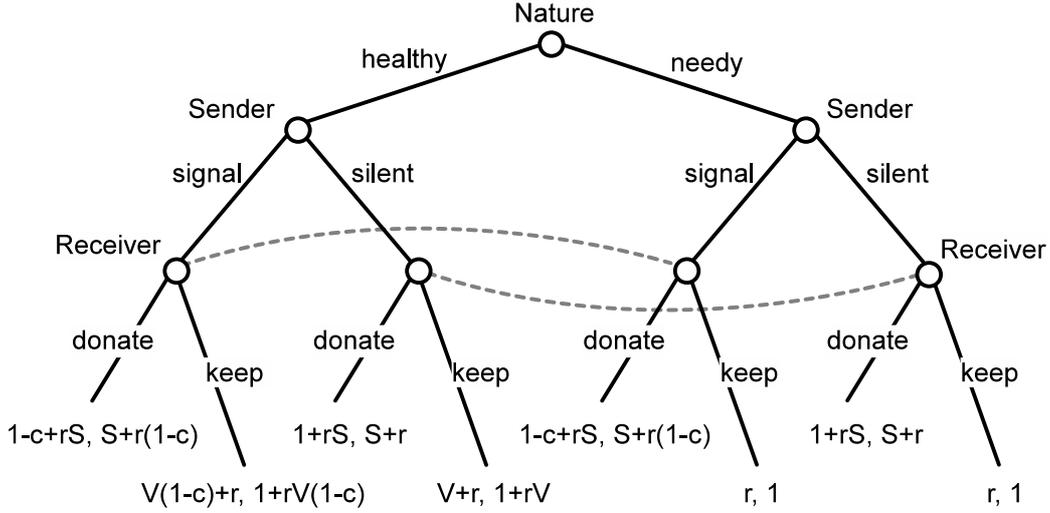$$v_R(t, s, a) = \begin{cases} S & \text{if } a = donate, \\ 1 & \text{if } a = keep. \end{cases} \tag{1}$$

Figure 1: The Philip Sidney game. Depending on the signal sent by Sender, Receiver is in one of two information states. An information state is a set of player nodes which the player cannot distinguish from each other. Decision nodes which belong to the same information state are connected by a dotted line. The utilities are given at the terminal nodes. The first utility belongs to Sender, the second to Receiver.

On the other hand, Sender is sure to survive if he receives the resource. If Sender does not receive the resource and he is *needy*, then his chance of survival is 0. If he does not receive the resource and he is *healthy*, then his chance of survival is $V < 1$. Thus, he benefits more from receiving the resource when he is *needy*, than when he is *healthy*. The survival probabilities for Sender $v_S(t, s, a)$ for state t, signal s, and action a are summarized by Equation (2):

$$v_S(t, s, a) = \begin{cases} (1 - c)w_{t,a} & \text{if } s = signal, \\ w_{t,a} & \text{if } s = silent, \end{cases} \tag{2}$$

where c is the signal cost and $w_{t,a}$ is an element of matrix w:

$$w = \begin{matrix} healthy \\ needy \end{matrix} \begin{pmatrix} \overset{donate}{1} & \overset{keep}{V} \\ 1 & 0 \end{pmatrix}.$$

Clearly, there is no reason for Receiver to donate his resource, except that the players may be related by some factor r, such that a player also benefits from the survival of the other. This is known as the principle of inclusive fitness (Hamilton, 1964). The utility u of each player is his survival probability plus a fraction r of the other player's survival probability. See Equations (3) and (4) for Sender's and Receiver's utilities, u_S and u_R respectively.

$$u_S(t, s, a) = v_S(t, s, a) + rv_R(t, s, a) \tag{3}$$

$$u_R(t, s, a) = v_R(t, s, a) + rv_S(t, s, a) \tag{4}$$

Increasing the relatedness factor r reduces the conflict between the players. When the relatedness factor $r = 1$, both players have the same utilities. The game tree and the utilities for each possible outcome of the Philip Sidney game are shown in Figure 1 and the parameters are summarized in Table 1.

Table 1 Parameters of the Philip Sidney game. Parameters c and r are in the closed interval $[0, 1]$, p, S, and V in the open interval $(0, 1)$.

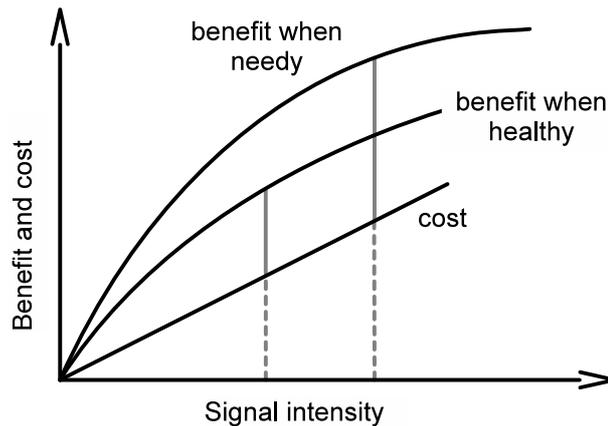|	meaning
c	signal cost
p	probability of Sender being healthy
r	relatedness factor
S	Receiver's survival probability when donating resource
V	Sender's survival probability when healthy but not getting resource

Figure 2: Visual representation of the handicap principle where signal cost and benefits increase with signal intensity. The benefit for a needy individual is higher than for a healthy one such that the optimal signaling intensity is higher for needy individuals.

We now wonder under what circumstances Sender should honestly signal his state (signal when he is needy, and remain silent when he is healthy) and Receiver should trust the Sender (donate when Sender signals, and keep otherwise).

The Philip Sidney game is a good model for the handicap principle since signals can be costly. Moreover, Sender's benefit of receiving the resource is higher when he is needy than when he is healthy. On the other hand, the cost of signaling does not depend on the state of Sender, but only on the signal intensity. Figure 2 visualizes this assuming there is a continuum of possible signals. This is not the case for the Philip Sidney game, but the reasoning remains the same. In any case, there exists an optimal signal intensity where the benefits maximally outweigh the costs (the vertical lines in Figure 2). Depending on the parameters of the game the optimal signal intensity for needy individuals may be higher than for healthy individuals. This means, that in those cases healthy individuals cannot profit from being dishonest. In other words, honest signaling is a (Nash) equilibrium.

2.1 Evolutionary Stability

The handicap principle only indicates there is a *possibility* that honest signaling is stable. To verify its stability one usually relies on the concept of evolutionarily stable strategies. A strategy is evolutionary stable when an entire population that adopts the strategy, cannot be invaded by a mutant strategy (Maynard Smith and Price, 1973). For honest signaling to be evolutionary stable Sender's strategy ("signal only when needy") must be the best response to Receiver's strategy ("donate only when Sender signals") and also the other way around: Receiver's strategy ("donate
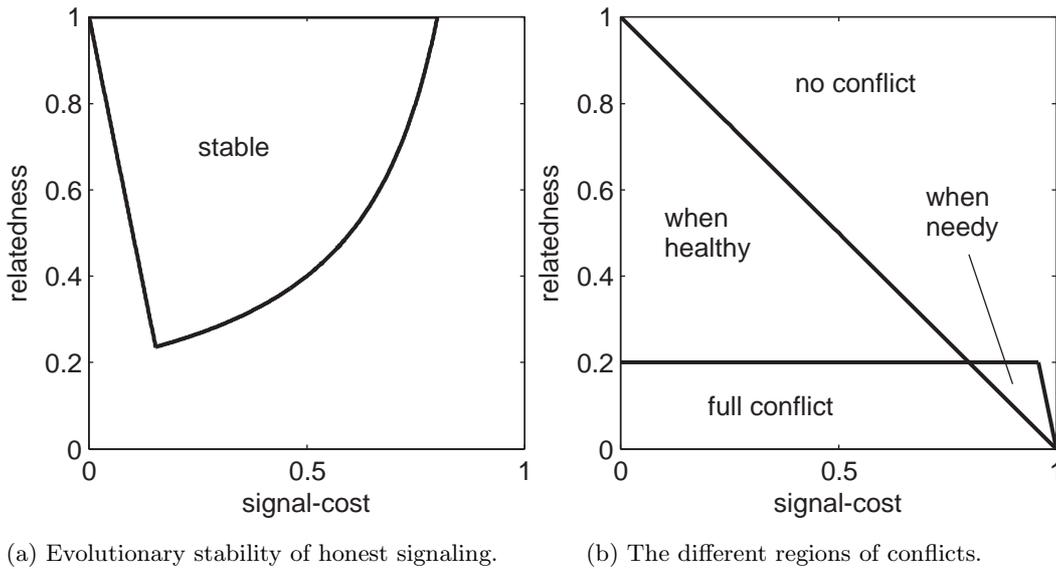
(a) Evolutionary stability of honest signaling.

(b) The different regions of conflicts.

Figure 3: Stability of honest signaling and conflict for all combinations of signal cost c and relatedness factor r, and parameters $S = V = 0.8$. The figures are independent of the probability of Sender being healthy (p).

only when Sender signals") must be the best response to Sender's strategy ("signal only when needy").

Straightforward algebra shows that "signal only when needy" is a best response to "donate only when Sender signals" whenever $r < 1 - c + rS$ and $1 - c + rS < V + r$. Similarly, "donate only when Sender signals" is a best response to "signal only when needy" whenever $1 + rV > S + r$ and $S + r(1 - c) > 1$ (Maynard Smith, 1991). Note that the probability p of Sender being healthy does not influence the evolutionary stability. For the Philip Sidney games where $S = V = 0.8$ Figure 3a shows for which combinations of signal cost c and relatedness factor r honest signaling is evolutionary stable.

The Philip Sidney game is an asymmetric two-player game since the players have different roles (Sender and Receiver) and for every asymmetric two-player game the set of evolutionarily stable strategies is equal to the set of *strict* Nash equilibria (Fudenberg and Levine, 1998) and hence, we may interchange both terms. Remember that a Nash equilibrium is a set of strategies, one for each player, such that none of the players can be better off by unilaterally deviating from the equilibrium and a strict Nash equilibrium is a set of strategies, one for each player, such that none of the players can be *strictly* better off by unilaterally deviating from the equilibrium. For honest signaling to be a strict Nash equilibrium, again, Sender's strategy ("signal only when needy") must be the best response to Receiver's strategy ("donate only when Sender signals") and Receiver's strategy must be the best response to Sender's strategy.

2.2   Conflict and Costly Signals

Remember that we try to understand under which circumstances signaling between individuals with *conflicting interests* will be honest. It is informative to verify when there is a conflict. We say there is a *conflict of interest* between the two players when there exists a state $t \in \{healthy, needy\}$ such that for all signals $s \in \{signal, silent\}$ the players prefer different actions $a, b \in \{donate, keep\}$ (Maynard Smith, 1991). More formally: there is a conflict of interest if and

only if

$$\exists\, t, \forall\, s : \arg\max_a u_S(t, s, a) \neq \arg\max_b u_R(t, s, b), \qquad (5)$$

where $u_S(t, s, a)$ and $u_R(t, s, a)$ are Sender's, respectively Receiver's, utility for outcome with state t, signal s, and action a. If Equation 5 holds for all states t, we say there is a full conflict, if it holds for none, we say there is no conflict, and otherwise we say there is a partial conflict.

Concretely, there is a conflict when Sender is healthy if $r(1 - V) < 1 - S$ and $r(1 - S) < (1 - c)(1 - V)$. Similarly, there is a conflict when Sender is needy if $r < 1 - S$ and $r(1 - S) < (1 - c)$. Note that the existence of conflict does not depend on the probability of Sender being healthy or needy. Figure 3b shows the different regions of conflict for all combinations of signal cost c and relatedness factor r when the parameters $S = V = 0.8$. Maynard Smith (1991) showed furthermore that if there is no signal cost ($c = 0$), honest signaling can only be stable if there is no conflict. When there is a conflict, honest signaling can only be stable if signals are costly ($c > 0$). The latter is the handicap principle. Note that honest signaling cannot be stable if there is a full conflict. If Sender would truthfully reveal his state, Receiver would do exactly the opposite of what Sender prefers.

Knowing whether or not there is a conflict may help in predicting the outcome of the game. When there is a full conflict, we may expect to see "never signal, never donate". When there is no conflict, we may expect to see "never signal, always donate". Both cases are states of no communication. We will see that this intuition is correct (Section 4).

3   Selfish Adaptive Behavior

Reinforcement learning algorithms are ideal as models of simple adaptive behavior of selfish individuals. A reinforcement learner repeatedly observes the state of the environment, selects an action and receives a payoff indicating the quality of his action. An individual can maintain statistics of the payoffs for all combinations of observations and actions. This way, he can learn which are the most rewarding actions in each state of the environment and hence change his behavior by selecting better actions more often.

We use Q-learning (Watkins and Dayan, 1992) with ϵ-greedy action selection, because of its ease of applicability. The parameters are easy to tune and the algorithm poses no constraints on the payoffs. Sender has a Q-value $Q_{t,s}$ for each possible state $t \in \{healthy, needy\}$ and each possible signal he can send, $s \in \{signal, silent\}$. Receiver has a Q-value $Q_{s,a}$ for each possible signal he can observe, $s \in \{signal, silent\}$, and each possible action he can take, $a \in \{donate, keep\}$.

We initialize Q-values optimistically by setting them to a value which is greater than or equal to the highest payoff of the game. This makes sure that there is sufficient exploration in the beginning of the game (Sutton and Barto, 1998, p.40). In the experiments we set initial Q-values to 2.

Now, for each game with outcome (t, s, a), both players update the Q-value that corresponds to the observation they made and the action they took during that game, according to the update rule (6). The other Q-values remain unchanged.

$$Q_{t,s} \leftarrow Q_{t,s} + \alpha\left(u_S(t, s, a) - Q_{t,s}\right)$$
$$Q_{s,a} \leftarrow Q_{s,a} + \alpha\left(u_R(t, s, a) - Q_{s,a}\right) \qquad (6)$$

In this update rule, $\alpha \in [0, 1]$ represents the learning rate. In the extreme case of $\alpha = 0$, nothing is ever learned. In the other extreme case where $\alpha = 1$, Q-values only reflect the last reward for the corresponding action. In fact, the Q-values represent the exponentially weighted moving average of the utilities (Sutton and Barto, 1998, p.38). Both players used the same learning rate $\alpha = 0.1$ in all experiments.

To balance exploration and exploitation, we used ϵ-greedy action selection: with probability $1 - \epsilon$ individuals select the action with the highest Q-value (braking ties randomly) and with

probability ϵ they select an action at random according to a uniform distribution. For all experiments we used a constant exploration rate $\epsilon = 0.01$.

Note, that we use the so-called "single-state" Q-learning. The reason for this is that each player only reaches one information state per game and takes only one action. Hence, there can be no information transfer between information states. Remember that Sender has information states *healthy* and *needy*, Receiver has information states *signal* and *silent*, and that per game the state can be either *healthy* or *needy*. In both cases Sender will either *signal* or remain *silent*.

One can prove that Q-learning will find the optimal policy, i.e. the Q-values converge to the true total expected utility provided that the environment is stationary and actions are selected infinitely often (Watkins and Dayan, 1992). Unfortunately, in a multi-agent setting the environment is non-stationary due to the presence of other agents who also learn and hence change their behavior. Whereas in a stationary environment exploration can be ignored once enough information has been collected, in a non-stationary environment the agent has to continue exploring in order to track changes in the environment. That is also why we used a constant exploration rate ϵ.

4 Experiments and Results

In all experiments reported here, initial Q-values were 2, the learning rate $\alpha = 0.1$, and the exploration rate $\epsilon = 0.01$ for all players. The population consisted of 100 individuals. We repeatedly selected two individuals at random to play a game. On average each individual was selected 10000 times as Sender and 10000 times as Receiver. During each simulation we recorded the outcomes of each game. An outcome consists of a state $t \in \{healthy, needy\}$, a signal $s \in \{signal, silent\}$, and an action $a \in \{donate, keep\}$. To get an accurate estimate of the probability of seeing a specific outcome at any timestep we used a moving average. Finally, the results were averaged over 100 simulations per experiment. An example of an experiment is shown in Figure 4 for signal cost $c = 0.1$, relatedness factor $r = 0.25$, $S = V = 0.8$, and $p = 0.5$. It shows the evolution of the frequencies of each of the outcomes over time.

In Figure 5 we show for which combinations of signal cost c and relatedness factor r honest signaling evolves. Honest signaling is the strategy pair "signal only when needy, donate only when signal", so the frequency of honest signaling is the sum of the frequencies of the outcomes ($healthy, silent, keep$) and ($needy, signal, donate$). In particular, dark red in the figure indicates honest signaling was always observed, dark blue indicates that honest signaling was never observed.

Three things are remarkable:

- First, it is particularly surprising that in a large part of the area where honest signaling is evolutionary stable (the area enclosed by the points $(0, 1)$, $(0.8, 1)$, and $(0.153, 0.236)$), honest signaling does not evolve! It only does so near the lower tip of that region at $(0.153, 0.236)$.
- Second, there is a red/orange/yellow region (to the left of the lower tip) where honest signaling is not evolutionary stable, still it evolves.
- Finally, the green area in the lower part of the figure seems to indicate honest signaling is observed 50% of the time.

Let us now try to explain each of these observations.

When honest signaling is stable In the region where honest signaling is stable, we only observed honest signaling near the lower tip (Figure 5). The reason why honest signaling not necessarily evolves even though it is an equilibrium lays in the fact that an other equilibrium may dominate honest signaling. In our example (Philip Sidney games with parameters $S = V = 0.8$ and $p = 0.5$) the strategy pair "never signal, always donate" is a Nash equilibrium for many combinations of signal cost c and relatedness r. Even more, for an important set of games where honest signaling is a Nash equilibrium, honest signaling is Pareto-dominated by the Nash
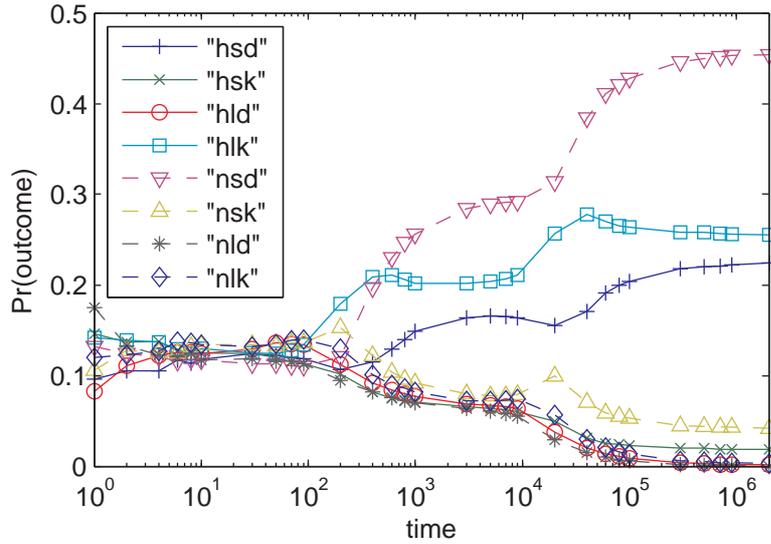
Figure 4: A typical experiment for the Philip Sidney game with parameters $S = V = 0.8$, $p = 0.5$, $c = 0.1$, and $r = 0.25$. The figure shows the evolution of the frequencies of each of the outcomes over time. Each outcome can occur at most half of the time, since the state $t \in \{healthy, needy\}$ is included in the outcome and the probability distribution over the states is fixed at $(p, 1 - p)$. In this case, when Sender is needy he almost always signals and Receiver mostly donates (this outcome is abbreviated by "nsd"). When Sender is healthy, about half of the time he remains silent and Receiver keeps the resource ("hlk"), and about half the time Sender bluffs and gets the resource ("hsd").
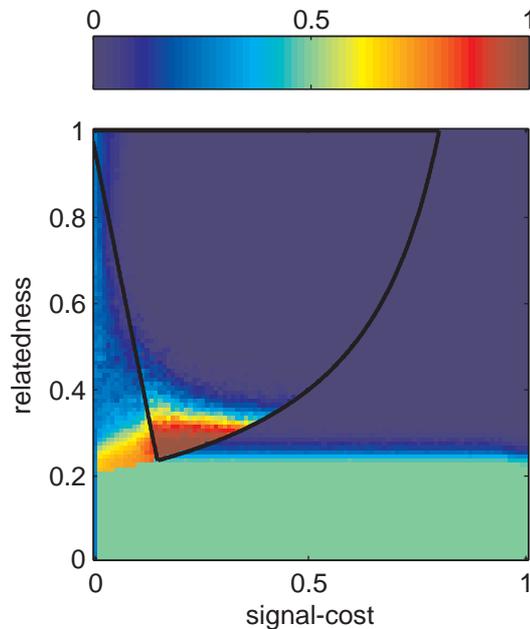

Figure 5: Frequency of honest signaling for all combinations of signal cost c and relatedness factor r. The other parameters of the Philip Sidney game are $S = V = 0.8$ and $p = 0.5$.

Figure 6: Frequency of the strategy pair "never signal, always donate" for all combinations of signal cost c and relatedness factor r. The other parameters of the Philip Sidney game are $S = V = 0.8$ and $p = 0.5$.
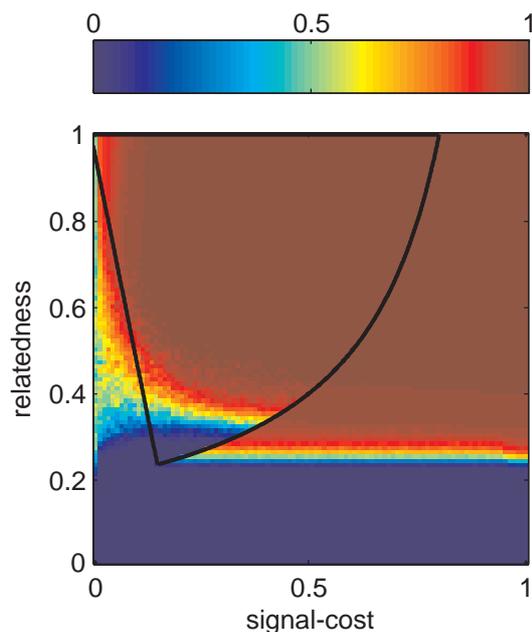
equilibrium "never signal, always donate". The frequency at which the strategy pair "never signal, always donate" was observed is shown in Figure 6. Clearly, for many games where honest signaling is a Nash equilibrium "never signal, always donate" was nevertheless observed with high frequency (> 0.9).

Figure 7 shows for which Philip Sidney games honest signaling is Pareto-dominated by another Nash equilibrium. The region where honest signaling is an unique Pareto-optimal Nash equilibrium, is also where we observed honest signaling. The region where honest signaling is a Pareto-optimal Nash equilibrium but not an unique one is where honest signaling sometimes emerged. Finally, we never observed honest signaling in the region where honest signaling is a Nash equilibrium Pareto-dominated by another Nash equilibrium. This also holds for other values of p, the probability of Sender being healthy. Note that, whether or not honest signaling is a Nash equilibrium is independent of the probability p that Sender is healthy (see Section 2.1 and Figure 3a), but that Pareto-optimality of equilibria does depend on p.

Clearly, other equilibria may be more important and we need to take this into account in order to predict the outcome of a game. By determining the evolutionary stability of honest signaling alone one overlooks other more important equilibria. One way to determine the importance of equilibria is to see whether or not they are Pareto-optimal. This may clear up some cases, but it is possible that multiple Pareto-optimal Nash equilibria exist in which case multiple equilibria may emerge.

Also note that, for a given signal cost (for example $c = 0.2$ in Figure 8), the amount of communication increases if the conflict intensifies (or, in other words, the relatedness factor r decreases) until all communication suddenly stops when relatedness $r < 1/5(1 - c)$.

Partial communication   There are cases where we observed partial communication although honest signaling is not evolutionary stable. The region is more or less located for signal cost c between 0 and 0.15, and relatedness factor r between 0.2 and 0.3 (Figure 5). In this region, it
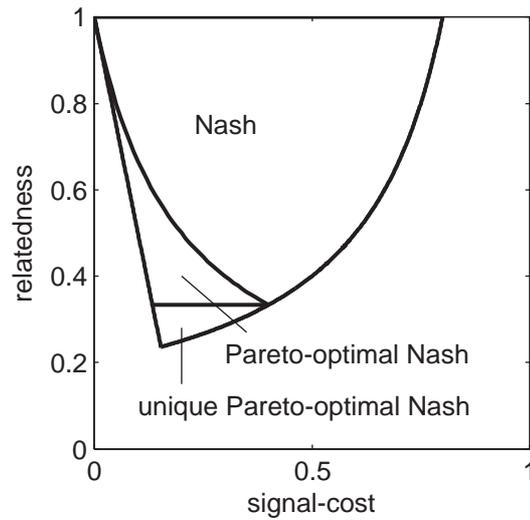
Figure 7: For which combinations of signal cost c and relatedness factor r honest signaling is a Nash equilibrium of the game. The other parameters of the Philip Sidney games are $S = V = 0.8$ and $p = 0.5$. In the biggest part of that area honest signaling is a Nash equilibrium Pareto-dominated by another equilibrium. In the region below it, it is not Pareto-dominated, but there is an other Nash equilibrium that is also not Pareto-dominated. Finally in the lower tip, honest signaling is an unique Pareto-optimal Nash equilibrium.
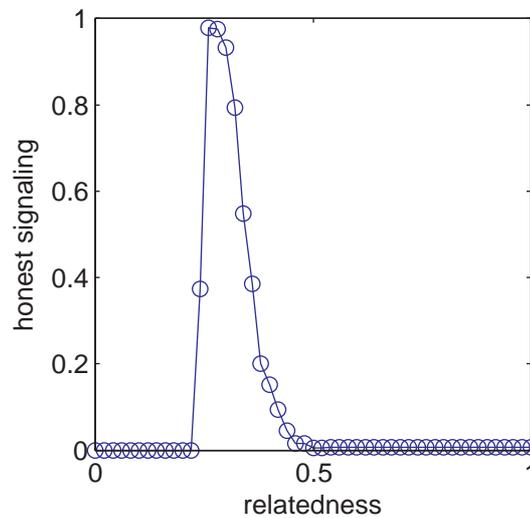

Figure 8: Frequency of honest signaling for varying relatedness r for signal cost $c = 0.2$, $S = V = 0.8$ and $p = 0.5$.

pays Receiver to donate most of the time when Sender signals. Sender is best off by signaling when needy. When healthy, he may sometimes bluff ("signal when healthy"), but not always. The same region is colored green in Figure 9, indicating that about 50% of the time when Sender is healthy he does not bluff, he remains silent. This is an interesting finding. We observe that being sometimes honest and sometimes dishonest can be an equilibrium where neither full communication (honest signaling), nor no communication is a stable outcome. The experiment
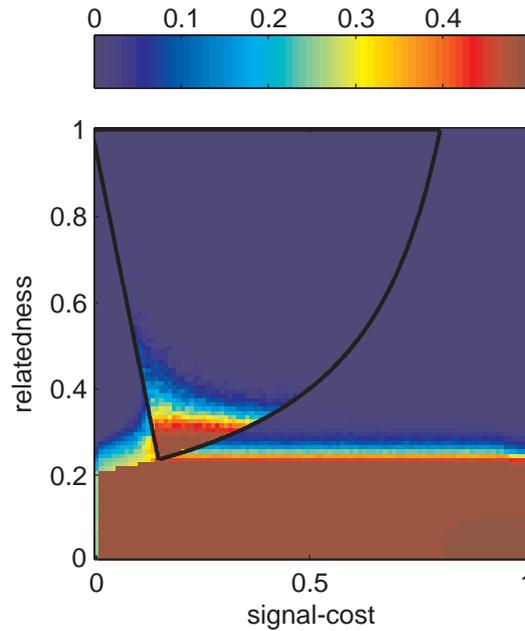
Figure 9: Frequency of outcome (*healthy*, *silent*, *keep*) for all combinations of signal cost c and relatedness factor r. The other parameters of the Philip Sidney game are $S = V = 0.8$ and $p = 0.5$.

shown in Figure 4 is an example of such a case: Sender signals always when needy, and about 50% of the time when healthy. Receiver always keeps when Sender is silent, and donates about 90% of the time.

Figure 10a shows the region where Sender bluffs and Figure 10b shows the frequency of bluffing for signal cost $c = 0.05$. For a given signal cost c, intensifying the conflict (by decreasing the relatedness factor r) makes bluffing pay off more and more. When the relatedness factor r is about 0.3 bluffing is near its maximum and decreasing the relatedness factor r even further decreases the equilibrium bluffing rate until the relatedness factor r crosses the curve $r = 1/5(1 - c)$. At that point, the conflict makes "never donate" a better strategy for Receiver. We find it interesting to see that intensifying a conflict (by decreasing the relatedness r from about 0.3 to 0.2) may decrease the amount of dishonesty and increase the amount of honesty (see Figure 10b).

For small conflicts (relatedness $r > 0.4$) bluffing does no longer pay off since Receiver is better off by always donating. For extreme conflicts ($r < 1/5(1 - c)$) it does not pay off for Sender to bluff for the opposite reason: Receiver is always doing better by keeping the resource anyway. Finally, note that, bluffing almost always payed off. This means that Sender is somehow able to learn an optimal bluffing rate, above which Receiver would no longer trust the signal.

Full conflict   The 50% of honest signaling observed in the Philip Sidney games in the lower part of Figure 5 is entirely due to the outcome (*healthy*, *silent*, *keep*), as is shown in Figure 9. The reason is that, in that region, it does not payoff for Receiver to donate whatever the state of Sender, and consequently, it does not pay off for Sender to signal. Sender will never signal unless the signal cost c is 0. The resulting strategy pair in that region is "never signal, never donate" which partially overlaps with honest signaling ("signal only when needy, donate only when signal") but is in fact a state of no communication.

(a) For varying signal cost c and relatedness r.   (b) For signal cost $c = 0.05$.
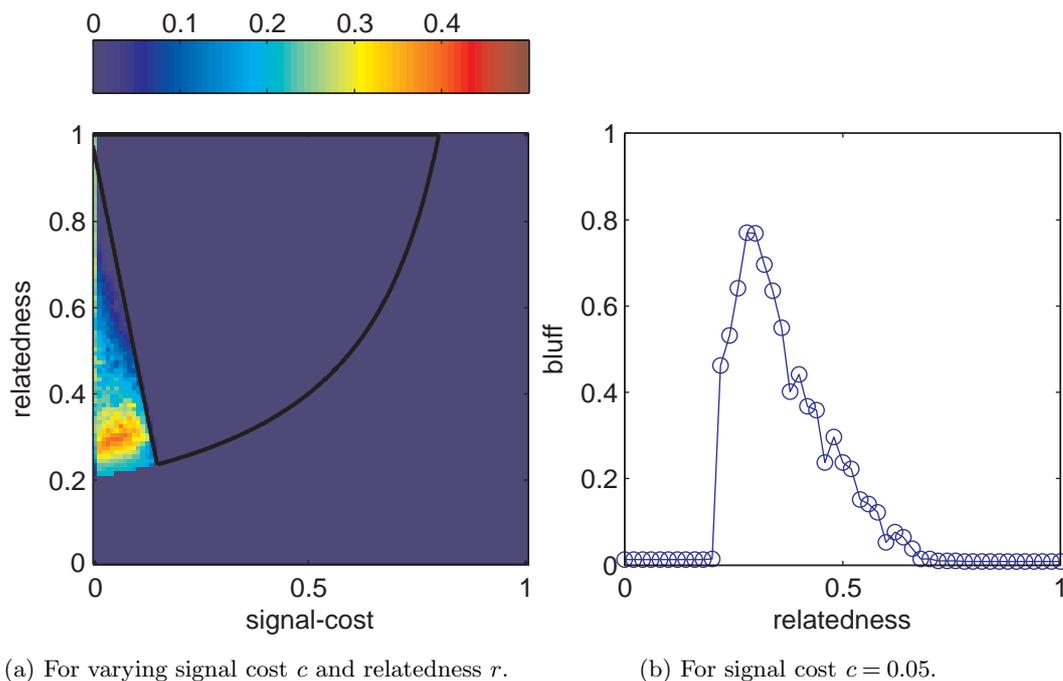
Figure 10: Frequency of bluffing for parameters $S = V = 0.8$ and $p = 0.5$.

5   Related Work

Huttegger and Zollman (2010) study evolutionary dynamics in the Philip Sidney game, and contrast their results to those obtained by calculating the evolutionarily stable strategies. They find that in many cases honest signaling has far smaller basins of attraction than other equilibria. This may mean that honest signaling is far less significant as is suggested by the existence of evolutionary stable strategies. Indeed, there is an ongoing debate about how widespread honest signaling really is (Maynard Smith and Harper, 2003; Searcy and Nowicki, 2005). Our work is exploring the same question and comes to the same conclusion by analyzing dynamics of a *learning* process.

Another part of the literature is concerned with *partially* honest signaling. Lachmann and Bergstrom (1998), for example, show the existence of many partial pooling equilibria, where Senders of different but very similar quality use the same signal. Huttegger and Zollman (2010) show the existence of mixed equilibria where the Sender is sometimes honest and sometimes dishonest. Here, we found a similar equilibrium that emerged when honest signaling is not evolutionary stable.

Closely related to the emergence of honest signaling is the emergence of signaling itself. The emergence of signaling in a game theoretic setting was first studied by Lewis (1969). His signaling games represent a communication problem where there is neither cost for signaling, nor conflict. The problem is fully cooperative and the goal for the players is to come up with a convention for the meaning of the signals without any pre-existent means of communication. Recently, both evolutionary dynamics (Hofbauer and Huttegger, 2008; Huttegger et al., 2009) and learning dynamics (Barrett and Zollman, 2009; Argiento et al., 2009; Skyrms, 2010; Catteeuw et al., 2011; Catteeuw and Manderick, 2013) received more attention and it was shown that perfect communication can evolve from initially random behavior through simple adaptive processes. While we concentrate on honest signaling in this text, we could indeed observe that initially arbitrary signals can acquire meaning through repeated interaction between learning processes.

6 Conclusion

Both in economics and biology honest signaling has mostly been studied by doing a static analysis of signaling games. Biologists mainly focused on evolutionarily stable strategies. In the process, they have been ignoring other equilibria which may be equally or even more important. Both the work of Huttegger and Zollman (2010), using evolutionary dynamics, and our work, using learning dynamics, show that honest signaling can emerge from initially random behavior through adaptive processes and more importantly it shows the existence of settings where honest signaling is an equilibrium, but where it is not (necessarily) the outcome of the dynamic process. We also observed the opposite. Partial communication may emerge in settings where honest signaling is not stable.

Economists are equally limiting the analysis of honest signaling to an equilibrium analysis. In the process they don't ignore other equilibria, rather they try to eliminate them. To do this, they impose extra requirements on the players' rationality. The research started by Cho and Kreps (1987) has lead to a whole range of Nash equilibrium refinements for signaling games, none of which seem generally applicable or agreed upon. Examples of such refinements are Condition D1, Divinity, Intuitive Criterion, etc. (Sobel, 2009). These refinements require players to be "unrealistically" rational. Behavioral economists have already provided evidence that people may not be as "rational" as is required by those equilibrium refinements. A simple adaptive process, such as the one we use here (Section 3), poses only very limited requirements on the players and can provide insights to what degree these equilibrium refinements may be reasonable, if at all. Our experiments already indicate that Pareto-optimality of an equilibrium is a good indicator for the importance of an equilibrium and may help to predict which equilibrium will emerge in a population.

We would like to mention that the example of the peacock's tail, though a classic one, is not the best example of honest signaling between selfish, adaptive individuals. Peacocks simply do not adapt their tails. A better example is stotting in gazelles. Gazelles perform an apparently weird behavior. When a gazelle spots an approaching predator, instead of running away, it jumps up and down in place. This is also called stotting. Stotting signals the predator that it is discovered and that the gazelle is quick enough to escape in case the predator would start chasing it. In the mean time, other gazelles nearby are informed of the presence of the predator (FitzGibbon and Fanshawe, 1988). A similar behavior is seen in some bird species which make alarm calls when they see a predator, or even sing while they are chased. Such behavior could be adapted during lifetime.

One may wonder, in how far the choice of learning model (Section 3) influenced our results. We argue that the choice is a reasonable one. Similar models of reinforcement learning have been applied to signaling games (Barrett and Zollman, 2009; Catteeuw and Manderick, 2013). All of them are very limited in terms of computational and informational requirements as mentioned before. They keep track of statistics summarizing the quality of their actions and keep a balance between exploiting the best actions and exploring the others. Despite their simplicity, these models have characteristics also seen in human and animal behavior such as recent outcomes affecting behavior more than older ones.

References

R. Argiento, R. Pemantle, B. Skyrms, and S. Volkov. Learning to signal: Analysis of a micro-level reinforcement model. *Stochastic Processes and their Applications*, 119(2):373–390, Feb. 2009. ISSN 03044149.

J. A. Barrett and K. J. S. Zollman. The role of forgetting in the evolution and learning of language. *Journal of Experimental & Theoretical Artificial Intelligence*, 21(4):293–309, Dec. 2009. ISSN 0952-813X.

D. Catteeuw and B. Manderick. The Limits of Reinforcement Learning in Lewis Signaling Games. In S. Devlin, D. Hennes, and E. Howly, editors, *Proceedings of the 13th Adaptive and Learning Agents workshop*, pages 22–30, Saint Paul, MN, USA, 2013.

D. Catteeuw, J. De Beule, and B. Manderick. Roth-Erev Learning in Signaling and Language Games. In P. De Causmaecker, J. Maervoet, T. Messelis, K. Verbeeck, and T. Vermeulen, editors, *Proceedings of the 23rd Benelux Conference on Artificial Intelligence*, pages 65–74, Ghent, Belgium, 2011.

I.-K. Cho and D. M. Kreps. Signaling Games and Stable Equilibria. *The Quarterly Journal of Economics*, 102(2):179–221, 1987.

I.-K. Cho and J. Sobel. Strategic stability and uniqueness in signaling games. *Journal of Economic Theory*, 50(2):381–413, Apr. 1990. ISSN 00220531. doi: 10.1016/0022-0531(90)90009-9.

C. D. FitzGibbon and J. H. Fanshawe. Stotting in Thomson's gazelles: an honest signal of condition. *Behavioral Ecology and Sociobiology*, 23(2):69–74, Aug. 1988. ISSN 0340-5443. doi: 10.1007/BF00299889.

D. Fudenberg and D. K. Levine. *The Theory Of Learning In Games*. The MIT Press, Cambridge, MA, 1998.

A. Grafen. Biological signals as handicaps. *Journal of Theoretical Biology*, 144(4):517–46, June 1990. ISSN 0022-5193.

W. D. Hamilton. The genetical evolution of social behaviour. *Journal of Theoretical Biology*, 7 (1):1–52, 1964.

J. Hofbauer and S. M. Huttegger. Feasibility of communication in binary signaling games. *Journal of theoretical biology*, 254(4):843–849, Oct. 2008. ISSN 1095-8541.

S. M. Huttegger and K. J. S. Zollman. Dynamic stability and basins of attraction in the Sir Philip Sidney game. *Proceedings of the Royal Society B: Biological Sciences*, 277(1689):1915–1922, June 2010. ISSN 1471-2954.

S. M. Huttegger, B. Skyrms, R. S. Smead, and K. J. S. Zollman. Evolutionary dynamics of Lewis signaling games: signaling systems vs. partial pooling. *Synthese*, 172(1):177–191, Feb. 2009. ISSN 0039-7857.

M. Lachmann and C. T. Bergstrom. Signalling among Relatives II. Beyond the Tower of Babel. *Theoretical Population Biology*, 54:146–160, 1998.

D. K. Lewis. *Convention: A Philosophical Study*. Harvard University Press, Cambridge, 1969.

J. Maynard Smith. Honest signalling: The Philip Sidney game. *Animal Behaviour*, 42:1034–1035, 1991.

J. Maynard Smith and D. Harper. *Animal signals*. Oxford University Press, Oxford, UK, 2003.

J. Maynard Smith and G. R. Price. The Logic of Animal Conflict. *Nature*, 246(5427):15–18, Nov. 1973. ISSN 0028-0836. doi: 10.1038/246015a0.

J. G. Riley. Silver signals: Twenty-five years of screening and signaling. *Journal of Economic Literature*, 39(2):432–478, 2001.

W. A. Searcy and S. Nowicki. *The Evolution of Animal Communication: Reliability and Deception in Signaling Systems*. Princeton University Press, Princeton, NJ, 2005.

B. Skyrms. *Signaling: Evolution, Learning and Information.* Oxford University Press, New York, 2010. ISBN 978-019-958294-5.

J. Sobel. Signaling Games. In R. A. Meyers, editor, *Encyclopedia of Complexity and Systems Science*, pages 8125–8139. Oxford University Press, New York, 2009. ISBN 978-0-387-30440-3.

M. Spence. Job Market Signaling. *Quarterly Journal of Economics*, 87(3):355–374, 1973.

R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction.* MIT Press, Cambridge, MA, 1998. ISBN 0-262-19398-1.

C. J. C. H. Watkins and P. Dayan. Q-Learning. *Machine Learning*, 8:279–292, 1992.

A. Zahavi. Mate selection - a selection for a handicap. *Journal of Theoretical Biology*, 53:205–214, Sept. 1975. ISSN 0022-5193.