Tjalling C. Koopmans Research Institute

Discussion Paper Series nr: 08-27

Economic Laboratory Experiment on Horn's Rule

Kris De Jaegher
Stephanie Rosenkranz
Utz Weitzel

Universiteit Utrecht
Utrecht School of Economics

Tjalling C. Koopmans Research Institute
Utrecht School of Economics
Utrecht University

Janskerkhof 12
3512 BL Utrecht
The Netherlands
telephone +31 30 253 9800
fax +31 30 253 7373
website www.koopmansinstitute.uu.nl

The Tjalling C. Koopmans Institute is the research institute and research school of Utrecht School of Economics.
It was founded in 2003, and named after Professor Tjalling C. Koopmans, Dutch-born Nobel Prize laureate in economics of 1975.

In the discussion papers series the Koopmans Institute publishes results of ongoing research for early dissemination of research results, and to enhance discussion with colleagues.

Please send any comments and suggestions on the Koopmans institute, or this series to m.vandort@econ.uu.nl

ontwerp voorblad: WRIK Utrecht

How to reach the authors

Please direct all correspondence to the first author.

Kris De Jaegher
Utrecht University
Utrecht School of Economics
Janskerkhof 12
3512 BL Utrecht, The Netherlands
E-mail: k.dejaegher@econ.uu.nl

Stephanie Rosenkranz
Utrecht School of Economics and CEPR
Utrecht University
Utrecht School of Economics
Janskerkhof 12
3512 BL Utrecht, The Netherlands
E-mail: s.rosenkranz@econ.uu.nl

Utz Weitzel
Utrecht University
Utrecht School of Economics
Janskerkhof 12
3512 BL Utrecht, The Netherlands
E-mail: u.weitzel@econ.uu.nl

Utrecht School of Economics
Tjalling C. Koopmans Research Institute
Discussion Paper Series 08-27

Economic Laboratory Experiment on Horn's rule

Kris De Jaegher[a]
Stephanie Rosenkranz[b]
Utz Weitzel[c]
[a]Utrecht School of Economics
Utrecht University
[b]Utrecht School of Economics
Utrecht University
and CEPR, London
[c]Utrecht School of Economics
Utrecht University
and MPI, Jena

October 2008

Abstract
This paper develops a framework for empirically testing several alternative game-theoretic rationales for Horn's rule. It then presents an economic laboratory experiment where these rationales are empirically tested. Subjects seem to coordinate on Horn's rule where efficiency acts as a focal point. Nevertheless, a weak interpretation of the evolutionary argument is not rejected: prior play does have an effect on future play. This suggests a hierarchy of effects, where the focal point effect dominates the evolutionary effect, with the latter being more pronounced in cheap talk situations.

Keywords: Horn's Rule, Signalling Theory, Pragmatics, Economic Laboratory Experiment.

JEL classification: C72, C92

Acknowledgements
We would like to thank Thomas Dirkmaat, Michael Franke, Robert van Rooij, Burkhard Schipper and participants of the 8th Conference on Logic and the Foundations of Game and Decision Theory, Amsterdam, July 2008, for helpful comments. Any remaining errors are our own.

1. INTRODUCTION

It is widely recognized among linguists that an economising principle has shaped human language (see Selten and Warglien, 2007, on grammar, and see the references cited in this paper). Language is thus a subject that can be studied from the point of view of economics. One instance where such an economising principle has been hypothesised to be at work in linguistics is Horn's rule (1984). This is the general rule in the branch of linguistics known as pragmatics saying that (un)marked expressions get an (un)marked interpretation. For instance, if we argue that every ten minutes, a man gets mugged in New York, then normally you will interpret this as meaning that this concerns a *different* man every ten minutes. Yet, strictly speaking, we could make such a statement to imply that every ten minutes, the same very unlucky New Yorker gets mugged. Thus, strictly speaking, our statement is ambiguous. The reason that you give our statement the first interpretation rather than the second is because you estimate it to be far more likely that the first interpretation is the correct one. But why do we not make our statement less ambiguous? Presumably, disambiguation is costly. If we have to eliminate all ambiguity, this will cost us more time. And as you are unlikely to make the second interpretation, we can leave our statement ambiguous without too much danger.

The basic game-theoretic account of Horn's rule looks as follows. The sender is denoted by symbol *S*, and the receiver by symbol *R*. Let there be two states of the world, namely a frequent state *F*, and an infrequent state *I*. The probability that the frequent state occurs is denoted p_F, so that the probability that the infrequent state occurs is $(1-p_F)$. By assumption, $p_F > 0.5$. The receiver can choose one of two actions, namely action *F*, or action *I*. Player *i*'s utility when the receiver takes action *j* in state *k* is denoted by $U_i(j|k)$, where $i = S, R$ and $j, k = I, F$. It is assumed that $U_i(j|j) > U_i(k|j)$, meaning that both the sender and the receiver prefer the receiver to take action *F* (respectively *I*) in state *F* (respectively *I*). The sender and the receiver thus have common interests. Also, it is assumed that

$$p_F U_i(F|F) + (1-p_F)U_i(F|I) > p_F U_i(I|F) + (1-p_F)U_i(I|I), \tag{1}$$

meaning that a player *i* (where $i = S, R$) who does not know what state of the world occurs, prefers action *F* to be taken. The sender knows what state of the world occurs, but the receiver does not. Finally at a cost $d > 0$, the sender can take an action (= send a signal).

The game has three Nash equilibria: one pooling equilibrium, where the sender does not send a message; one efficient separating equilibrium meeting Horn's rule, where the sender sends a message in the infrequent state; and one inefficient separating equilibrium that does *not* meet Horn's rule, where the sender sends a message in the *frequent* state (thus incurring a higher expected cost than in the equilibrium meeting Horn's rule.

Our game is similar to a *cheap-talk signalling game*, in that there is more than one separating equilibrium, and that the meaning of the signal is a matter of convention. In experimental signalling games where signals are completely costless, Blume et al. (1998) show that subjects can learn to give meaning to random signals. Horn's game is somewhat different, however, in that there is still a cost involved when signalling, and in that the separating equilibria are Pareto-ranked. Another class of signalling games is formed by the *costly signalling games* (Spence, 1974), where there is a costly signal that only one signaller type would ever prefer to send. This class of games is again different from the game in our paper, as in our game, in different separating equilibria, both types of signallers are willing to send the costly signal. A similarity between our game and costly signalling games is the fact that the different separating equilibria are Pareto-ranked. In costly signalling games, this is so because the high-quality signaller may use a signal that is more costly than is necessary to distinguish herself from the low-quality signaller. Brandts and Holt (1992) have experimentally tested the extent to which subjects playing a costly signalling game can coordinate on playing the efficient equilibrium. Selten and Warglien (2007) finally construct an experiment for a game where, unlike in costly signalling games, it is not the case that it is a dominated strategy for one or more types to send a costly signal; at the same time, unlike in cheap-talk signalling games, a cost is still attached to sending a message. This is the same setting as in our experiment. Selten and Warglien (2007), however, do not assume that the states about which signals can be sent have different probabilities. The focus here is on the use of a simple "grammar", where few signals are used to signal many states, as an application of the economising principle.

We start by reviewing the game-theoretical literature on Horn's rule (Section 2), and we show that several alternative models lead to the same predictions, suggesting that different rationales of Horn's rule cannot be tested for. Sections 3 and 4, however, develop variants on the basic Horn game that do allow one to empirically distinguish between the different rationales. Based on these results, Section 5 then proposes several hypotheses to be tested in the experiment. Section 6 describes the experiment, Section 7 describes the results. We end with a conclusion in Section 8.

2. STANDARD HORN GAME WITH COSTLY SIGNALS

From the perspective of non-cooperative game theory, there is no reason why Horn's rule should apply. It may very well be that you and we both know that costs will be saved if we do not need to disambiguate. But if we believe that you will interpret the sentence 'Every ten minutes, a man gets mugged in New York' as meaning one and the same man, then each time we want to make the more plausible statement, we will have to point out that we do not mean one and the same man. And once we talk in this way, if we do not specify that we do *not* mean one and the same man, you will interpret it as meaning one and the same man. Simply, there are multiple Nash equilibria to the signalling game, and our mutual beliefs can lock us into an inefficient separating equilibrium.

A first attempt to provide a game-theoretic rationale in favour of Horn's rule, so that only the separating equilibrium obeying Horn's rule survives, was made by Parikh (1991, 2000, 2001). Parikh argues that the fact that the signalling equilibrium that follows Horn's rule is Pareto efficient will make it a *focal point* (Schelling 1960), and will cause it to be played. Yet, it is difficult to predict what will be a salient equilibrium, and this may be another one than the Pareto-efficient one (Kim, 1996).

Van Rooij (2008) recently gives an equilibrium selection argument in favour of signalling equilibria meeting Horn's rule. Using the so-called *intuitive criterion* (Cho and Kreps 1987), he argues that, when the players are playing the pooling equilibrium, and when the signaller deviates from the equilibrium by sending a message, the receiver will always interpret this as referring to the infrequent event. This is because only the signaller who has observed the infrequent state can become better off by deviating from the pooling equilibrium. A rational receiver will therefore infer that any out-of-equilibrium signal must have been sent in the infrequent state. In turn, expecting the receiver to make such an inference, the signaller will send a message in the infrequent state. Yet in its standard form, the intuitive criterion only eliminates the pooling equilibrium, and not the inefficient signalling equilibrium. Once the players play the inefficient signalling equilibrium, the signaller who observes the infrequent state has no reason to start sending a message in the infrequent state. Van Rooij solves this problem by arguing that one should take a pooling equilibrium as one's starting point.

The intuitive criterion belongs to the type of argument for equilibrium selection known as the *forward induction* argument. An alternative version of the forward induction argument, known as *extensive form rationalisability* (Pearce, 1984; for applications to signalling games, see Battigalli, 2004), does not require the starting point of a pooling equilibrium. The technique of extensive form rationalisability is to assume common knowledge of rationality, and to put some commonly-known restrictions on players' beliefs about each other's behaviour (first-order beliefs). If the necessary restrictions on beliefs for a particular equilibrium to be played are quite weak, then it is plausible that this equilibrium will be played. In the current case, however, it is obvious that a necessary restriction for a separating equilibrium obeying Horn's rule to be played is that the receiver believes with a relatively high probability that any signal received must have been sent in state *I*. It is not clear what could generate such beliefs, unless one refers to efficiency itself – which brings us back to the focal point argument. We conclude that the forward induction argument only seems applicable in the form of the intuitive criterion, where the starting point of a pooling equilibrium is necessary.

Additionally, there have been some recent efforts to check whether evolutionary game theory selects signalling equilibria that obey Horn's rule. As pointed out by Van Rooij (2004), all signalling equilibria, including equilibria that do not obey Horn's rule, are evolutionary stable equilibria (ESS), and therefore the ESS concept does not bring one any closer to a solution. Under replicator dynamics, Benz, Jäger and Van Rooij (2005) show for a specific game that the basin of attraction of an equilibrium that obeys Horn's rule is larger than the one of an equilibrium that does not, and that starting from a strategy profile where all strategies are played with equal probability, under replicator dynamics an equilibrium obeying Horn's rule evolves.

De Jaegher (2008) argues that the evolution of an equilibrium that selects Horn's rule follows straightforwardly from the fact that a signalling equilibrium must at some point have evolved from a pooling equilibrium (otherwise, it does not make sense to study the evolution of signalling). In a pooling equilibrium, the optimal action taken by the receiver will correspond to the most likely event. The absence of a signal is therefore already interpreted as referring to a frequent event. From there on, the interpretation of a signal as referring to an infrequent event can evolve ("Horn equilibrium"). For a signalling equilibrium where a signal refers to a frequent event to evolve from the pooling equilibrium, the receiver would somehow have to completely change his initial interpretation of the absence of the fact that no signal is received. It is this necessary reversal of meaning that makes it difficult for a signalling equilibrium that does not obey Horn's rule (= "anti-Horn" equilibrium) to evolve from a pooling equilibrium. Specific about De Jaegher's model is that the roles of the players are assumed fixed; put otherwise, senders and receivers are different populations. A concise description of this model is given in Appendix A.

To conclude, in the standard Horn game, focal point theory, the intuitive criterion, and evolutionary dynamics all predict that Horn's rule applies for the standard Horn game. Both Van Rooij's interpretation of the intuitive criterion and De Jaegher's evolutionary argument, however, take the pooling equilibrium as a starting point, so that whether or not this was played first in these rationales of Horn's rule is predicted to have an influence; in focal point theory, there is no need for this. These results are summarised in Table 1, Column 1 (Standard Horn game, cost), where a "star" indicates that the theory predicts that there should be an influence from whether or not the starting point is the pooling equilibrium. As the predictions of the different theories for the standard Horn game largely coincide, the next section presents a modified Horn game for which the predictions of the different models diverge to a higher extent.

	Standard Horn game		Modified Horn game	
	Cost (Game 2)	**No cost (Game 2')**	**Cost (Game 4)**	**No cost (Game 4')**
Focal point (efficiency)	Horn's rule	/	Horn's rule	/
Intuitive criterion	Horn's rule*	/	Anti-Horn's rule*	/
Evolution	Horn's rule*	Horn's rule*	Anti-Horn's rule*	Anti-Horn's rule*
Convention	/	50/50	/	50/50

Table 1: Standard Horn game and modified Horn game: predictions (: effect of previously playing pooling equilibrium)*

3. MODIFIED HORN GAME WITH COSTLY SIGNALS

De Jaegher (2008) presents a variation on the standard Horn game, referred to here as the *modified Horn game*, which is identical to the standard Horn game except that, in the pooling equilibrium, the receiver prefers to take action I, meaning that

$$p_F U_i(F|F) + (1-p_F)U_i(F|I) < p_F U_i(I|F) + (1-p_F)U_i(I|I) \quad (2)$$

This is in spite of the fact that $p_F > 0.5$. Inequality (2) applies as long as $U_i(F|I)$ is very small. Evidently, focal point theory continues to predict here that the Horn equilibrium, which continues to be efficient in this game, will be obtained. Interestingly, however, both the evolutionary argument and the intuitive criterion predict that the anti-Horn equilibrium, where a signal is sent in the frequent state, and which is inefficient in this case, will be obtained.

De Jaegher's evolutionary argument applied in this context is the following. The pooling equilibrium, where I is always played, already predisposes the receiver to play I when not receiving any signal. Evolution of the Horn equilibrium would require a complete reversal of meaning, where "no signal" would come to mean "F" to the receiver. For the anti-Horn equilibrium to evolve, on the contrary, it suffices that, in the pooling equilibrium, a part of the receiver population starts to interpret a signal as referring to state F. Intuitively, consider animals with common interests, trying to warn each other of an approaching predator. With communication incapacitated, even if the probability that a predator approaches is unlikely, because of the high risk of being caught, the uninformed animal prefers to hide. An "all clear" signal is then predicted to evolve rather than an "alarm call", so that the signal is frequently sent. For a formal statement of this argument, see Appendix A. Applying the evolutionary argument of Benz et al. (2005) in this context, a similar prediction is obtained, because the basin of attraction of the anti-Horn equilibrium is larger than the basin of attraction of the Horn equilibrium.

A somewhat similar intuition is obtained when applying Van Rooij's (2008) application of the intuitive criterion to this game: given that in the pooling equilibrium, the action that is optimal in the infrequent state is taken, the receiver should interpret any signal as having been sent in the *frequent* state. Because of the form taken by the pooling equilibrium, the receiver this time reasons that only a sender who observed the *frequent* state has any incentive to deviate from the pooling equilibrium and send a costly message. It follows that both the intuitive criterion and evolution predict that Horn's rule does *not* apply in this case. The variant of the Horn game introduced by De Jaegher is thus interesting in that Parikh's focal point argument on the one hand (= efficient equilibrium is played, and Horn's rule applies), and Van Rooij's intuitive criterion/De Jaegher's evolutionary argument on the other hand, lead to different predictions. These predictions are summarised in Table 1, Column 3 (Modified Horn game, cost). Note that both intuitive criterion and the evolutionary argument again predict that it matters whether a pooling-equilibrium game was previously played. Unfortunately, in the modified Horn game, we still do not obtain different predictions for the intuitive criterion and for the evolutionary argument. This is why in the next section we introduce a second variation on the Horn game.

4. STANDARD AND MODIFIED HORN GAME WITH COSTLESS SIGNALS

Consider a variant of both the standard game in Section 1, and the modified game in Section 3, with the difference that signalling is now completely costless. In both these new variants of the game, focal point theory and the intuitive criterion do not make any predictions about the sort of separating equilibrium that will be played. (Note that in Van Rooij's application of the intuitive criterion, it is the fact that the signal is costly that allows the receiver to make inferences about when it must have been sent.) In the evolutionary argument, however, not noted by De Jaegher (2008), the cost of the signal is in fact not crucial to the argument. Instead, the fact that "no signal" takes on a particular meaning in the pooling equilibrium, and thereby predisposes the receiver towards a certain separating equilibrium, means that Horn's rule is most likely to apply in the standard game, and that anti Horn's rule is most likely to apply in the modified Horn game. Indeed, in this case, as signals are costless, starting from the pooling equilibrium, a tiny proportion of either the sender or the receiver playing a Horn strategy (respectively anti-Horn strategy) in the standard Horn game (respectively modified Horn game) suffices for the Horn equilibrium (respectively anti-Horn equilibrium) to evolve.

Nevertheless, it should be noted that, in the standard Horn game (respectively modified Horn game) there is also an evolutionary path from the pooling equilibrium to the Horn (respectively anti-Horn) equilibrium. Again, this is because signals are costless. If, with respect to the pooling equilibrium, the senders drift in the "wrong" direction, then the "wrong" separating equilibrium evolves. Nevertheless, a sizeable proportion of the sender population needs to deviate then from the pooling equilibrium, whereas for the evolution of the "right" equilibrium, a tiny deviation by either population suffices. If large deviations are less likely than small ones, the "right" equilibrium is still more likely to evolve. In terms of the evolutionary argument by Benz et al. (2005), the basin of attraction of the Horn equilibrium is relatively large in the standard Horn game with cheap signals, whereas the basin of attraction of the anti-Horn equilibrium is relatively large in the modified Horn game with cheap signals. These evolutionary arguments are formally shown in Appendix A.

The additional predictions obtained by adding costless signalling are summarised in Table 1, Columns 2 and 4 (Standard Horn game, No cost; Modified Horn game, No cost). While the intuitive criterion does not make any prediction about what happens in the games without cost, from the perspective of the intuitive criterion, there is no reason why there would be such a specific difference between the games without cost. From this perspective, one is able to empirically distinguish between the intuitive criterion and the evolutionary argument.

When we are treating cheap-talk signalling games, pioneered by Lewis (1969), we also need to pay attention to the theories that have been developed specifically for such games. Lewis himself predicts that the way in which a signal is used in a cheap-talk signalling game is a matter of convention. In this sense, both anti-Horn and Horn are equally likely in a cheap-talk signalling game. A different prediction can be found by again using Schelling's focal point argument. When both anti-Horn and Horn are efficient, players may focus on other aspects of these equilibria to coordinate on one of them. For instance, players could see "no signal" as 0, and signal as 1, and order these in the same way as the states and then consider *D* (down) as 0, and *U* (up) as 1, leading them to play anti-Horn. Alternatively, the players could consider it salient to send a message only in the infrequent state (Horn). If the focal point lies in the signal itself we cannot predict what the players focus on. We are thus not able to make any predictions here.

Finally, having analysed both costly-signalling and cheap-talk versions of our Horn games, we can look at whether the theories make any predictions about differences between costly-signalling and cheap-talk games. Let us *first* look at the intuitive criterion, where following Van Rooij (2008) we assume that the pooling equilibrium is the starting point. The intuitive criterion argues that, as long as the signal is costly, the receiver will interpret it in a particular way, and the sender will use it in a particular way. This theory therefore predicts that there is little strategic uncertainty in this case, and that players should find it easy to coordinate. When signals are cheap, on the contrary, the intuitive criterion does not make any predictions, suggesting that there is a lot of strategic uncertainty, and that players will find it harder to coordinate. *Second*, let us look at the evolutionary argument, where the pooling equilibrium is again assumed to be the starting point. Clearly, the basin of attraction of any pooling equilibrium is always larger in costly signalling games than in cheap-talk signalling games. Simply, more receivers need to make a particular interpretation before the senders are better off with sending a costly signal. Thus, the evolutionary argument predicts that the evolution of a signalling equilibrium is *more* difficult when signals are costly.

5. HYPOTHESES

Our hypotheses are divided into three sets: first, hypotheses that are specific to costly signalling games; second, hypotheses that are specific to cheap-talk signalling games; and, third, hypotheses that compare costly signalling games to cheap talk signalling games.

5.1 Costly signalling treatments.

The evolutionary argument and the intuitive criterion both say that the starting point matters. Yet, if we let subjects play a signalling game right away, we are not certain *what their starting point is*. In order to simulate such a starting point, in our experimental design, we first let participants play a corresponding pooling-equilibrium game against the computer (which we refer to as a *pooling-equilibruim game*), so without the possibility of some other participant sending a signal on which state occurs. Our first step is to test whether the preceding play of the corresponding game against the computer has any effect. Define "playing correctly" according to the intuitive criterion and the evolutionary argument as playing Horn (anti-Horn) in the Horn (modified Horn) game.

(H1) In the costly signalling treatments where the corresponding pooling-equilibrium game was played first, "playing correctly" occurs more often.

Yet, perhaps subjects start by playing a pooling equilibrium even in the signalling games, and a starting point is already built into the repeated play of the signalling game itself. If this is true, then from the perspective of both the evolutionary argument and the intuitive criterion, "playing correctly" (i.e., playing Horn in the Horn game and anti-Horn in the modified Horn game) should occur over *all* treatments with costly signalling games, including those treatments where the corresponding pooling-equilibrium game was *not* played first. This leads us to formulating Hypothesis 2.

(H2) In the treatments without prior play of the corresponding pooling-equilibrium game, Horn is played more often in the standard game (Game 2) than in the modified game (Game 4), and anti-Horn is played more often in the modified game (Game 4) than in the standard game (Game 2).

	(H2) not rejected	(H2) rejected
(H1) not rejected	Evolutionary/intuitive criterion – pooling equilibrium already in Games 2 and 4 itself	Evolutionary/intuitive criterion – pooling equilibrium *not* already in Games 2 and 4 itself
(H1) rejected	/	Compatible with focal point

Table 2: Costly signalling games - Hypotheses 1 and 2

Table 2 describes which combinations of rejection and non-rejection of (H1) and (H2) are compatible with which theories. Yet, the test of the theories is weak here, as the only issue that is tested is whether a certain strategy is played more often. Taking the intuitive criterion and the evolutionary argument literally, Horn should predominantly be played in the standard Game 2, and anti-Horn should be predominantly played in the modified Game 4.

(H3) In the standard game (Game 2), subjects predominantly play the Horn equilibrium; in the modified game (Game 4), they predominantly play the anti-Horn equilibrium.

5.2 Cheap-talk signalling games

Let (H3) not be rejected. Then this is compatible with both the evolutionary argument, *and* the intuitive criterion. Thus, while such an observation would allow us to reject focal point theory, it does not allow us to distinguish between the evolutionary argument and the intuitive criterion. In Section 4, however, we have shown that the intuitive criterion and the evolutionary argument lead to different predictions for cheap-talk and costly signalling games (at least in the sense that, for cheap-talk signalling games, only the evolutionary argument yields a clear prediction). In particular, acceptance of the evolutionary argument leads to the following hypothesis:

(H4) In the cheap-talk signalling treatments, having previously played the relevant pooling-equilibrium game has a positive effect on the probability of "playing right".

Analogue to Hypothesis 2, we also test for the cheap-talk signalling games whether the "starting point" of first playing a pooling equilibrium is already built into these games. In this case the evolutionary argument leads to the following hypothesis:

(H5) Comparing the cheap talk signalling treatments where the corresponding pooling-equilibrium game was not previously played, Horn is played more often in the standard game (Game 2') than in the modified game (Game 4'), and anti-Horn is played more often in the modified game (Game 4') than in the standard game (Game 2').

5.3 Comparing costly and signalling games to cheap-talk signalling games

As pointed out at the end of Section 4 above, the intuitive criterion and the evolutionary argument yield different predictions about the relative likelihood of coordination in costly signalling games and in cheap-talk signalling games. Specifically, the evolutionary argument predicts that communication fails more often with costly signalling, while the intuitive criterion predicts the opposite, i.e. that costly signalling facilitates communication. Again, we use the evolutionary perspective in order to formulate a hypothesis on these diametrically opposed predictions:

(H6) Communication fails more often in the costly signalling games than in the cheap-talk signalling games.

6. DESIGN OF THE EXPERIMENT AND PROCEDURAL DETAILS

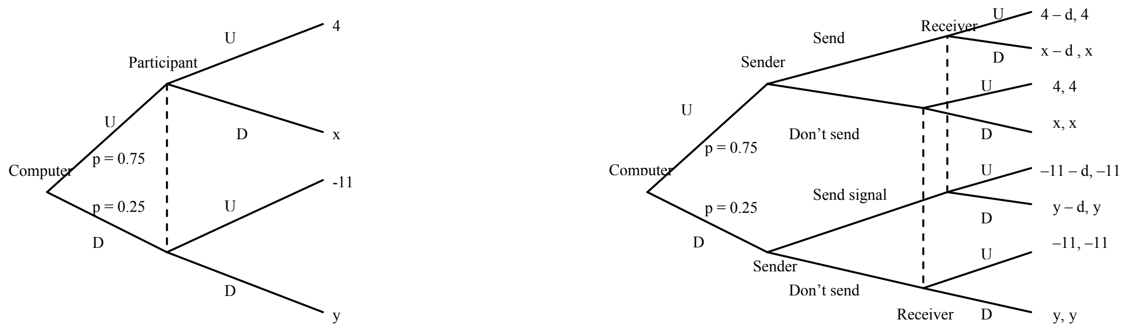
The aim of our experimental design is to test the hypotheses developed on the previous section. We designed the experiment so that both the standard Horn game of Section 1 and the modified game of Section 3 were played both in the costly signalling version and the cheap-talk version. The four resulting signalling games are summarised in the right part of Figure 1. On top of these games, in order to simulate the prior play of a pooling equilibrium, which is important for both Van Rooij's application of the intuitive criterion and for De Jaegher's evolutionary argument, in some sessions we let the relevant signalling game be preceded by a non-signalling version of that same game (referred to as a pooling-equilibrium game), where the subjects played a guessing game against the computer; subjects were then in the same situation as a receiver unable to receive a message from a sender. These games are summarised in the left part of Figure 1.

The games treated in Sections 2, 3 and 4 above differ according to three treatment variables: costs, difference in payoffs, prior pooling. This implies a 2×2×2 factorial design: each of the four games (standard with and without costly signals, modified with and without costly signals), once with a pooling-equilibrium game played before and once separately. We ran 8 sessions at the experimental laboratory ELSE at Utrecht University. The initial sessions (sessions 1 to 6) in this investigation were run on 10th and 11th December 2007 at the ELSE laboratory of Utrecht University in six different sessions each time involving between 18 and 20 participants. Two sessions (sessions 7 and 8) were run on 4th June 2008 at the same laboratory. In order to duplicate observations for our treatment variables, we administered four treatments per session, following a classical cross-over design.[1] Participating subjects came from the subject pool that mainly recruits students from Utrecht University from all faculties. The procedure during all sessions was kept the same and the sessions were computerized, using a program written with z-tree (Fischbacher, 1999). 114 subjects participated and were seated in a random order at PCs. Throughout the sessions subjects were not allowed to communicate and could not see others' screens. Before each individual treatment, and before any roles were assigned, the subjects received a written set of instructions, which was also read aloud before the start of the experiment. Questions were answered in private.

Each pooling-equilibrium game was played 10 times. In this type of games, the computer

[1] See Appendix C, Table 5, for the sequence in which the treatments were run.

first randomly decided between *U* (frequent state) and *D* (infrequent state) independently for each subject. This choice was not observed by the subject, who had to guess whether *U* or *D* had occurred (by clicking on *U* or *D* on the computer screen), after which the choice of the computer and the corresponding payoffs were announced.



Pooling-equilibrium games *Signalling games*

Figure 1: The six games played in the experiment; numbers refer to points obtained by the participants. Games 1, 2, 2': x = –11, y = 4, Games 3, 4, 4': x = y = 1; Games 2, 4: d = 2; Games 2', 4': d = 0.

Each signalling game was played 20 times. At the beginning of each signalling treatment, subjects were randomly grouped in pairs, and within each pair, the role of sender and receiver was randomly assigned. Subjects interacted anonymously through the computer. We chose a partner design (subjects stayed in the same pairs within each treatment). Thus, the evolutionary argument is interpreted as learning within a single pair.[2] In a typical signalling game, at the start of the treatment, the computer announced the subject's role. In each period, the sender was informed about whether state *U* or *D* occurred. After this, the sender could tick a box on a computer screen indicating "Send message", or indicating "Don't send message". Next, the receiver could observe the sender's choice ("the sender in your pair did/did not send you a message"), but not the state chosen by the computer. After this, the receiver had to guess whether *U* or *D* had occurred (by clicking on *U* or *D* on the computer screen), after which the choice of the computer and the corresponding payoffs were announced. With each decision subjects earned payoffs as indicated in Figure 1 and at the end of the experiment, the points were turned into Euros according to the formula 1 point = 1.65 Eurocents. This exchange rate was announced in the instructions.

This setup has some differences with Selten and Warglien's (2007) experiment, where, in each period and before observing the states, the sender has to design an encoding strategy, and the receiver a decoding strategy. After that, the state of nature is chosen randomly, and the computer executes the subjects' encoding and decoding strategies. In case of a mismatch for a particular state of nature realised in a period, the subjects are told each other's encoding and decoding strategies for this particular state of nature. While our feedback mechanism is similar (when the payoffs are realised, the receiver observes the state of nature, and the sender observes the action that the receiver took), we do not let the sender and receiver explicitly choose an encoding and decoding strategy, respectively. This is because we want to give the subjects the opportunity to play a pooling equilibrium. The starting point of a pooling equilibrium is important for the application of the intuitive criterion and of the evolutionary argument, which both need such a starting point. Note that, while the pooling equilibrium is

[2] Indeed, as shown by Börgers and Sarin (1997), replicator dynamics can be interpreted as the learning by individual players, where the percentage change in their strategy is a function of how well they did in the last period.

separately simulated in a non-signalling setting for some treatments, this is not the case of all treatments. Moreover, it is possible that the subjects initially play a pooling equilibrium even if the non-signalling game is not played first.

7. DATA DESCRIPTION

7.1 Overall frequencies

We start with some broad descriptions of the strategies played over all the sessions. Figure 1 shows the frequencies with which strategies *U* and *D* were played in the pooling-equilibrium Games 1 and 3 over the 10 periods. In Game 1, while subjects as predicted predominantly play *U*, they play *U less* often over time.[3] This can be explained by the so-called *gambler's fallacy*, which can itself be explained by Tversky and Kahneman's (1974) *representativeness heuristic* of boundedly rational decision makers facing uncertainty. After having observed three times the *U*-state (which about half of the subjects do statistically in periods 1 to 3), many subjects expect that a *D*-state will now occur, even though the probability that a *D*-state occurs is only ¼. In Game 3, while the expected value of playing *D* is larger, half of the subjects play *U*. The behaviour of the participants who play *U* can be explained by prospect theory (Kahneman and Tversky, 1979). E.g., if a subject in Game 3 thinks in terms of losses with respect to the payoff of 4 units, and if he is risk-loving with respect to losses, he indeed chooses *U*. In any case, the play in Game 3 is sufficiently different from that in Game 1 to test whether the previous playing of a pooling equilibrium influences which separating equilibrium is played.

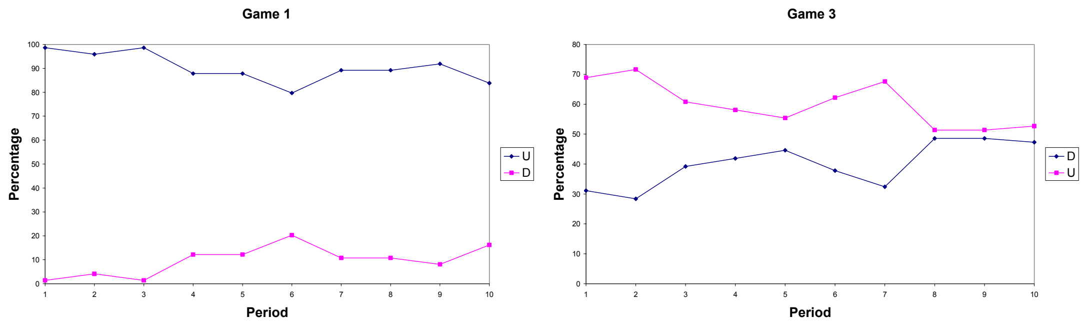


Figure 2: Frequency over periods of actions U and D in pooling-equilibrium games.

Figure 3 shows, for senders and receivers separately, the frequency of different strategies over all signalling games and per consecutive range of 5 periods. As a criterion for playing Horn's rule or anti-Horn's rule we specified consecutive play of a strategy consistent with Horn or anti-Horn over 5 periods and with at least one occurrence of the infrequent state. "Pooling, signal" and "Pooling, no signal" mean that the sender always takes the same action over 5 periods, respectively always sends a signal or never sends a signal. "Pooling frequent" and "Pooling infrequent" mean that, over a range of 5 periods, the receiver always guesses the frequent or infrequent state, respectively. The remaining sender and receiver strategies are indicated as "failure".

Figure 3 reports that in a large majority of cases subjects were able to communicate, and that both Horn and anti-Horn strategies were played, even though Horn was played more often. Interestingly, receivers failed to communicate more often than senders. Receivers also

[3] This result is also confirmed by panel probit estimations (unreported) with choice of U as the dependent and the period as the main independent variable. Unreported estimations are provided by the authors on request.

needed more learning in order to achieve coordination. In contrast, the proportion of the *sender* population playing Horn and anti-Horn changes relatively little between the initial and the final periods. It thus seems that encoding a message is more difficult for subjects than decoding it.

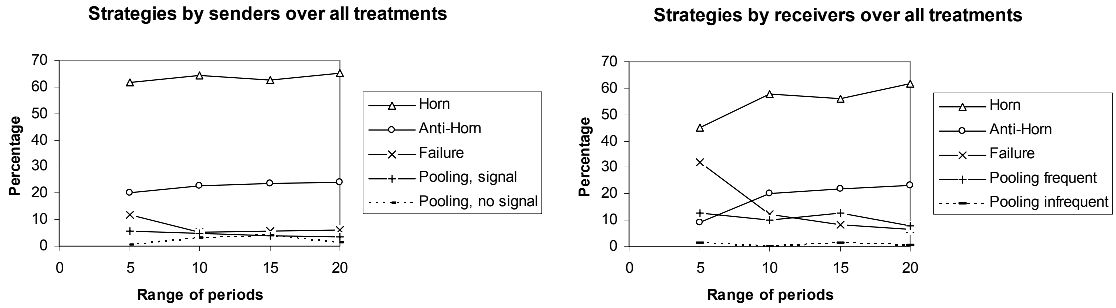


Figure 3: frequency of strategies over ranges of five periods for senders and receivers.

It is also of interest to compare the costly signalling games to the cheap-talk signalling games over the entire experiment. As shown in Figure 4, Horn is predominantly played in both types of games, but is more prominent in the costly signalling games.

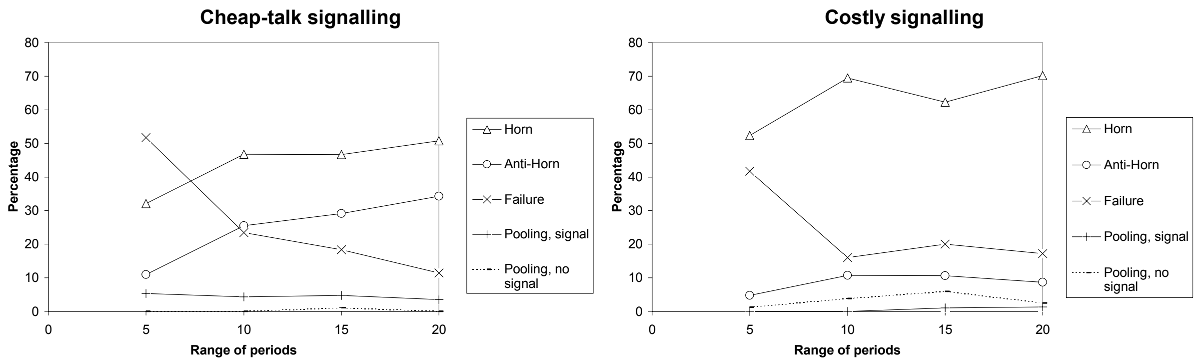


Figure 4: frequency of strategies for cheap-talk and costly signalling games.

7.2 Frequencies in individual games

Let us next look at the frequency of Horn and anti-Horn strategies for individual games, but across all sessions. Figure 5 reports how often per period (as a percentage of all decisions per period) both subjects in the randomly matched pairs decided in accordance with Horn or anti-Horn. The figure shows that subjects play Horn more often (play anti-Horn less often) in Game 2 (Game 2'), when compared to Games 4 (Game 4').

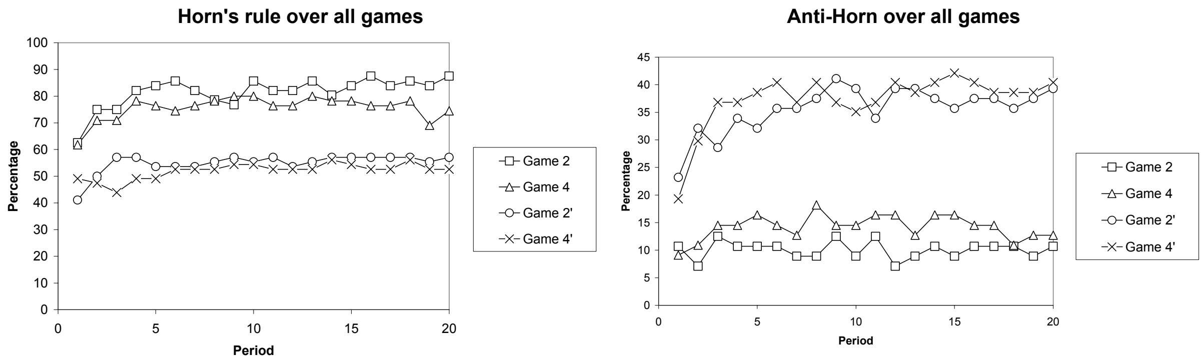


Figure 5: Frequency of Horn and anti-Horn strategies per period and game

7.3 Individual games in individual sessions

Finally, we look at play in individual treatments and sessions, as summarised in Table 4 (see Appendix B). For the pooling-equilibrium games, Table 4 shows how often *U* and *D* were played on average across 10 periods. For the signalling games, it shows how often Horn and anti-Horn was played, and how many times the subjects failed to play either of these (outcome *N*). A *first* conclusion that we can draw from Table 4 is that in any of the costly signalling games, Horn is played in a large majority of the cases. A *second* conclusion is that, if subjects in an early treatment play Horn (anti-Horn) in a majority of cases, then they continue doing this in any ensuing treatment. An exception is found in sessions 3 and 8, where costly signalling games are played after cheap-talk signalling games. In this case, the subjects predominantly switch to Horn's rule. Thus, summarising, Horn's rule is dominant in all costly signalling games. This is the case even if anti-Horn was played in a preceding cheap-talk game. Moreover, playing Horn's rule in a costly signalling game has a strong effect on the probability that Horn's rule is played in later games.

8. HYPOTHESES TESTING

(H1)

In order to test Hypothesis 1 we compare all play of Game 2 (Game 4) with preceding play of Game 1 (Game 3) to all play of Game 2 without any preceding play of Game 1 (Game 3), and this in terms of whether subjects play "right" (Horn in Game 2, anti-Horn in Game 4). Such a comparison is made in Figure 6. It reveals that the effect of playing Game 1 on Game 2 is inconclusive. At the same time, anti-Horn is played more often in Game 4 when Game 3 was played first – even though the effect gets smaller as the experiment proceeds. As already argued above, there is also an influence of the previous play of other signalling games. In order to eliminate such an effect, we also apply look at Hypothesis 1 only for the instances of Game 2 and Game 4 where there was no preceding play of another signalling game (Game 2 in Sessions 1 and 7, and Game 4 in Sessions 2 and 4 respectively). The results in these cases are very similar, confirming that the effect of previously played signalling games on the play in costly signalling games is not very strong. We conclude that, while (H1) cannot be rejected, there is only weak support for it.

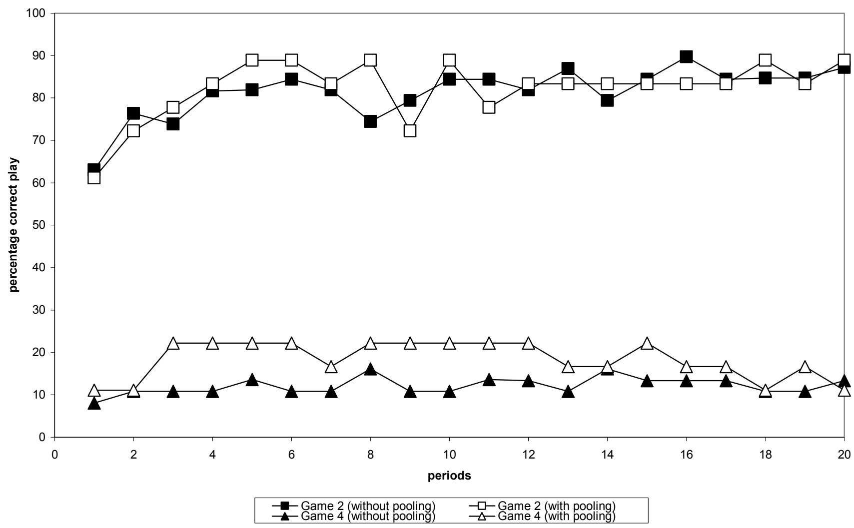


Figure 6: Frequency of "playing right" over individual periods for Games 2 and 4 with and without prior pooling-equilibrium game (all treatments)

(H2)

While this hypothesis covers only the games without prior pooling-equilibrium game, and Figure 5 reports *all* games, the patterns for the games without prior pooling-equilibrium game are nearly identical to those represented in Figure 5, and we do not separately represent them here. More Horn is played in Game 2 than in Game 4, and slightly more anti-Horn is played in Game 4 than in Game 2 (where the latter effect seems to become weaker as time progresses). (H2) cannot be rejected, as there appears to be an effect even if a pooling-equilibrium game was not played before.

Summarising, a "weak" interpretation of the evolutionary argument/intuitive criterion, in terms of a prediction that certain strategies are played more often, cannot be rejected. Nevertheless, let us look at the stronger test of these theories in (H3)

(H3)

Hypothesis 3 is clearly rejected by Figure 5. Whenever subjects play any of the costly signalling games in the experiment, they predominantly play Horn. This is indicative of Parikh's focal-point argument, saying that subjects focus on the Pareto-efficient equilibrium.[4]

(H4)

We have already pointed out that Horn's rule is predominantly played in costly signalling games. Moreover, as is clear from Table 4, in any cheap-talk game played after a costly signalling game, Horn's rule continues to be played predominantly, under the influence of the preceding costly signalling game. This is why we cannot look at all play of cheap-talk games to test (H4). Otherwise, we could be picking up the effect of a preceding costly signalling game rather than the precedence of a pooling-equilibrium game. We therefore focus on comparing the first two treatments. Figure 7 reveals that, in Game 2', with a prior pooling-equilibrium game, more Horn is played. In Game 4', with a prior pooling-equilibrium game, more anti-Horn is played. It follows that the evolutionary argument cannot be rejected. It

[4] This result is also confirmed by separate panel probit estimations (unreported), which also account for correlations within pairs (clustering) and for heteroskedasticity (robust estimation coeffcients). In all these estimations, cost has a strong and statistically significant effect on the probability that Horn is played.

should also be noted that the support for the evolutionary argument is strong compared to the one for the costly signalling games.

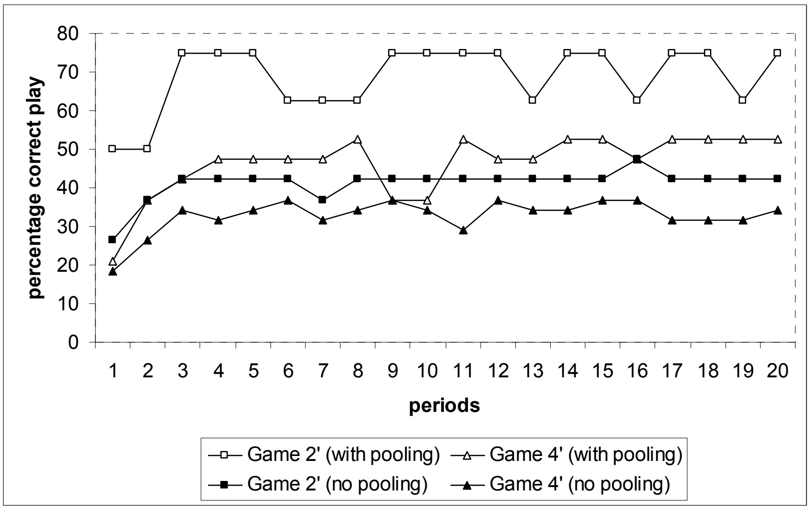


Figure 7: Frequency of Horn and anti-Horn strategies over individual periods for Games 2' and 4' with and without prior pooling (first two treatments)

(H5)
To avoid influence of one game on the other, we focus on Session 3 and 8, where Games 2' and 4' were played as first treatments, respectively. Table 4 reveals that anti-Horn is played relatively often in Game 2', and less often in Game 4'. It follows that (H5) is rejected. In combination with (H4), this suggests that the prior play of a pooling-equilibrium game is important in determining the outcome.

(H6)
Figure 4 reveals the following. Initially, there is more failure in the cheap-talk signalling games, but eventually, there is slightly more failure in the costly signalling games. Thus, the hypothesis in favour of the evolutionary argument, saying that coordination is more difficult to achieve when signals are costly (meaning that there is a risk in taking the initiative of sending a signal) cannot be rejected.

In order to further refine Hypotheses 1 and 5, we estimate a multinominal logit, with subjects' play of Horn, anti-Horn and failure to coordinate (in the pair) in the second treatment of Sessions 1, 2, 5 and 6 as dependent categorical variable, and with the absolute number of mistakes that they made (with respect to rational behaviour) in the pooling-equilibrium game in Treatment 1 as the main independent variable. We control for learning by including the period as a trend variable and correct for heteroskedasticity by computing robust estimation coefficients. Further, we account for possible correlations within pairs (clustering). The results are reported in Table 3. In Game 2', Session 5, subjects who, on average, played *D* often (high number of mistakes), play Horn significantly less often. This seems to be in line with the evolutionary argument. Yet, in Game 4, Session 2, subjects who play *U* more often (high number of mistakes) also play Horn significantly *less* often, which stands in contrast to the evolutionary argument. In general, the effect of the number of mistakes in the prior pooling-equilibrium game on the signalling game is negative, even though not all coefficients are significant. The explanation we suggest for the general negative effect is that subjects who make such mistakes are less sophisticated, and find it harder to achieve coordination through communication.

Multinominal logit with coordinated Horn, Anti-Horn, or Failure as dependent categorical variable:

	Game 2, Session 1		Game 2', Session 5		Game 4, Session 2		Game 4', Session 6	
	Horn	Anti-Horn	Horn	Anti-Horn	Horn	Anti-Horn	Horn	Anti-Horn
No. of prior pooling mistakes	-0.014	-0.041	-0.682*	-0.433^	-0.317**	-0.085	-0.118	-0.065
	[-0.238]	[-0.443]	[-1.969]	[-1.906]	[-2.707]	[-0.904]	[-1.107]	[-0.939]
period	0.086^	-0.337***	0.146	0.162	0.086	0.049	0.205**	0.212***
	[1.738]	[-7.672]	[1.470]	[1.534]	[0.874]	[0.477]	[3.115]	[3.533]
constant	1.169*	-1.379	1.297	-0.191	3.479**	1.966^	-0.472	0.495
	[2.297]	[-1.395]	[1.575]	[-0.268]	[2.635]	[1.715]	[-0.589]	[1.038]
N	360		320		360		400	
Chi2	5144.574		38.386		11.344		36.812	
p	0		0		0.023		0	

^ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Failure to coordinate is base case; robust estimation with clustering at the pair level [t-values in brackets]

Table 3. Estimations with number of prior pooling mistakes in treatments 2 of session 1, 2, 5, 6

8. DISCUSSION

The results of the experiment point into the direction of the Parikh's (2001) focal point explanation of Horn's rule. Subjects seem to coordinate on Horn's rule where efficiency acts as a focal point. A weak interpretation of the evolutionary argument, however, is not rejected: prior play does have an effect on future play. Thus, many subjects seem to follow the focal-point interpretation of Horn's rule. Nevertheless, in cheap-talk signalling games, subjects look at what was previously played in the preceding pooling-equilibrium game. This suggests a hierarchy of effects, where the focal point effect dominates the evolutionary effect. As the focal point effect only applies with costly signals, in the case of cheap talk, the evolutionary effect is still at work.

REFERENCES

Battigalli, P. (2004) 'Rationalization in signaling games: theory and applications', working paper, Università Bocconi.

Benz, A., Jäger, G. and Rooij, R. van (2005) 'An introduction to game theory for linguists', in A. Benz, G. Jäger and R. van Rooij (eds) *Games and Pragmatics*, London: Palgrave Macmillan.

Blume, A., DeJong, D.V., Kim, Y.-K. & Sprinkle, G.B. (1998) 'Experimental evidence on the evolution of meaning of messages in sender-receiver games.' *American Economic Review* 88: 1323-1340.

Börgers, T. & Sarin, R. (1997) 'Learning through reinforcement and replicator dynamics.' *Journal of Economic Theory* 77, 1-14.

Brandts, J. & Holt, C.A. (1992) 'An experimental test of equilibrium dominance in signaling games.' *American Economic Review* 82: 1350-1365.

Crawford, V. (1998) 'A survey of experiments on communication via cheap talk.' *Journal of Economic Theory* 78: 286-298.

De Jaegher, K. (2008). 'The evolution of Horn strategies.', *Journal of Economic Methodology*, forthcoming.

Fischbacher, U. (1990), z-tree 1.1.0: Experimenter's Manual, University of Zurich, Institute for Empirical Research in Economics, http://www.iew.unizh.ch/ztree/index.php

Horn, L. (1984) ‘Towards a new taxonomy of pragmatic inference : Q-based and R-based implicature’, in Deborah Schiffrin (ed.) *Meaning, Form, and Use in Context : Linguistic Applications* GURT84, Washington: Georgetown University Press, pp.11-42.

Kahneman, D. & A. Tversky (1979) ‘Prospect theory: an analysis of decision under risk.’, *Econometrica* 47, 263-291.

Kim, Y. (1996) ‘Equilibrium selection in n-person coordination games.’, Games and Economic Behavior 15, 203-227.

Parikh, P. (1991) ‘Communication and strategic inference’, *Linguistics and Philosophy* 14: 473-513.

Parikh, P. (2000) ‘Communication, meaning, and interpretation’, *Linguistics and Philosophy* 23: 185-212.

Parikh, P. (2001) *The Use of Language*, Stanford, California: CSLI Publications.

Pearce, D. (1984) ‘Rationalizable strategic behaviour and the problem of perfection’, *Econometrica* 52, 1029-1050.

Rooij, R. van (2004) ‘Signalling games select Horn strategies’, *Linguistics and Philosophy* 27: 493-527.

Rooij, R. van (2008) ‘Games and Quantity Implicatures’, *Journal of Economic Methodology*, forthcoming.

Schelling, T. (1960), *The Strategy of Conflict*, New York: Oxford University Press.

Selten, R. and Warglien, M. (2007) The emergence of simple languages in an experimental coordination game. *Proceedings of the National Academy of Sciences* 104, 7361-7366.

Tversky, A. & Kahneman, D. (1974) ‘Judgement under uncertainty: heuristics and biases.’, *Science* 185, 1124-1131.

Appendix A

This appendix describes and graphically represents the evolution of the pure strategies under replicator dynamics in a large population of senders on the one hand, and receivers on the other hand playing the four variants of the Horn game treated in this paper (games 2, 2’, 4, 4’ – see Table 1). In these games, each player has four pure strategies, making a graphical representation of the complete streamline diagram under replicator dynamics difficult. In order to keep a graphical representation tractable, following De Jaegher (forthcoming), we represent two of the faces of the complete streamline diagram that are relevant for our purposes.

To derive the first face, consider population states where only the strategies in the pooling equilibrium and in the Horn equilibrium are played, and where other strategies are never played, as represented in Table 1A. Note now that, given any such population state, the individual receiver and sender cannot do strictly better than the population by playing any of the strategies excluded in Table 1A. It follows that, in population states consistent with Table 1, the excluded strategies never grow in the population, so that from Table 1A, a face of the complete phase diagram can be constructed.

		Sender: $(1-p)$	Sender: p
		Don’t send any signal	Send signal in infrequent state
Receiver: $(1-r)$	Always interpret i	$(p_F\, U_R(i\vert F) + (1-p_F)\, U_R(i\vert I)$, $p_F\, U_S(i\vert F) + (1-p_F)\, U_S(i\vert I))$	$(p_F\, U_R(i\vert F) + (1-p_F)\, U_R(i\vert I)$, $p_F\, U_S(i\vert F) + (1-p_F)\, [U_S(i\vert I) - d])$
Receiver: r	Interpret frequent when no signal, and infrequent when signal	$(p_F\, U_R(F\vert F) + (1-p_F)\, U_R(F\vert I)$, $p_F\, U_S(F\vert F) + (1-p_F)\, U_S(F\vert I))$	$(p_F\, U_R(F\vert F) + (1-p_F)\, U_R(I\vert I)$, $p_F\, U_S(F\vert F) + (1-p_F)\, [U_S(I\vert I) - d])$

Table 1A: Pooling equilibrium versus Horn equilibrium (standard Horn game: i = F; modified Horn game: i = I; costly signalling games: d > 0; cheap-talk signalling games: d = 0)

The replicator dynamics corresponding to Table 1 are given by

$$ds/dt = s(1-s)\begin{Bmatrix} p_F\{q[U_R(F|F)-U_R(i|F)]-(1-q)[U_R(i|F)-U_R(I|F)]\}+ \\ (1-p_F)[U_R(I|I)-U_R(i|I)] \end{Bmatrix} \tag{A1}$$

$$dq/dt = q(1-q)p_F\{s[U_S(F|F)-U_S(I|F)]-d\} \tag{A2}$$

where t denotes time. Dividing the right-hand sides of (A1) and (A2), and applying partial integration, one now obtains that

$$p^{-[U_R(i|I)-U_R(F|I)]}(1-p)^{-p_F[U_R(F|F)-U_R(i|F)]-(1-p_F)[U_R(I|I)-U_R(i|I)]}(1-r)^{(1-p_F)[U_S(I|I)-U_S(F|I)-d]}r^{(1-p_F)d} = K \tag{A3}$$

where K is a constant. To each level K in (A3) corresponds a different streamline, describing the evolutionary path followed when starting from any population state consistent with Table 1A.

To derive the second face of the phase diagram, consider population states where only the strategies in the pooling equilibrium and in the anti-Horn equilibrium are played, and where other strategies are never played, as represented in Table 2A. Note again that, given any such population state, the individual receiver and sender cannot do strictly better than the population by playing any of the strategies excluded in Table 2A. It follows that, in population states consistent with Table 2A, the excluded strategies never grow in the population, so that from Table 2A, a face of the complete phase diagram can be constructed.

		Sender: $(1-q)$	Sender: q
Receiver		Don't send any signal	Send signal in frequent state
$(1-s)$	Always interpret i	$(p_F\, U_R(i\|F) + (1-p_F)\, U_R(i\|I)$, $p_F\, U_S(i\|F) + (1-p_F)\, U_S(i\|I))$	$(p_F\, U_R(F\|F) + (1-p_F)\, U_R(i\|I)$, $p_F\, [U_S(F\|F) - d] + (1-p_F)\, U_S(i\|I))$
s	Interpret infrequent when no signal, and frequent when signal	$(p_F\, U_R(I\|F) + (1-p_F)\, U_R(I\|I)$, $p_F\, U_S(I\|F) + (1-p_F)\, U_S(I\|I))$	$(p_F\, U_R(F\|F) + (1-p_F)\, U_R(I\|I)$, $p_F\, [U_S(F\|F) - d] + (1-p_F)\, U_S(I\|I))$

Table 2A: Pooling equilibrium versus anti-Horn equilibrium (standard Horn game: i = F; modified Horn game: i = I; costly signalling games: d > 0; cheap-talk signalling games: d = 0)

The replicator dynamics corresponding to Table 2 are given by

$$ds/dt = s(1-s)\begin{Bmatrix} p_F\{q[U_R(F|F)-U_R(i|F)]-(1-q)[U_R(i|F)-U_R(I|F)]\}+ \\ (1-p_F)[U_R(I|I)-U_R(i|I)] \end{Bmatrix} \tag{A4}$$

$$dq/dt = q(1-q)p_F\{s[U_S(F|F)-U_S(I|F)]-d\} \tag{A5}$$

Dividing the right-hand sides of (A4) and (A5), and applying partial integration, one now obtains that

$$\begin{aligned} &q^{(1-p_F)[U_R(I|I)-U_R(i|I)]-p_F[U_R(i|F)-U_R(I|F)]}(1-q)^{-p_F[U_R(F|F)-U_R(i|F)]-(1-p_F)[U_R(I|I)-U_R(i|I)]} \\ &(1-s)^{p_F[U_S(F|F)-U_S(I|F)-d]}s^{p_F d} = L \end{aligned} \tag{A6}$$

where L is a constant. To each level L in (A6) corresponds a different streamline, describing the evolutionary path followed when starting from any population state consistent with Table 2A.

For the four signalling games treated in this paper, Figures A1 to A4 describe the evolutionary path under replicator dynamics between the pooling equilibrium (part (a) of each figure) and the Horn equilibrium, and between the pooling equilibrium and the anti-Horn equilibrium (part (b) of each figure). There are two ways to interpret these figures. A first way is to check whether the relevant separating equilibrium can evolve from the pooling equilibrium. One can include *evolutionary drift* in such evolutionary path, where players' strategies are subject to change when they are indifferent. A second way is to check the size of the basin of attraction around each equilibrium (= the area in the phase diagram with all the population states from which a certain equilibrium evolves). The reasoning is then that players start by choosing a strategy at random, and are more likely to end up in equilibria with a larger basin of attraction.

As is clear from Figure 1A, in the standard Horn game with costly signals, there is an evolutionary path from the pooling equilibrium to the Horn equilibrium. The black line at the bottom left part of Figure 1A represents a range of weak pooling equilibria. Starting from any such equilibrium, through evolutionary drift, some receivers can start making a Horn interpretation; if a sufficient number of them do, the Horn equilibrium evolves. Figure 1B shows that there is no evolutionary path from the pooling equilibrium to the anti-Horn equilibrium. Moreover, even if Horn strategies are excluded from the populations, as is the case Table 3, the basin of attraction of the pooling equilibrium is still relatively large. This contrasts with the Figure 1A, which shows that the basin of attraction of the pooling equilibrium is small with respect to the Horn equilibrium.

As shown in Figure 2A, the modified Horn game with costly signals is the mirror case from the Horn equilibrium. Now, there is an evolutionary path from the pooling equilibrium to the anti-Horn equilibrium (through evolutionary drift, Figure 2B), but no evolutionary path from the pooling equilibrium to the Horn equilibrium. The basin of attraction of the pooling equilibrium is smaller with respect to the anti-Horn equilibrium than with respect to the Horn equilibrium. Note, however, that the basin of attraction of the Horn equilibrium continues to be quite large in Figure 2A. This is simply because this is the efficient equilibrium.

Figure 3A shows that in the standard Horn game with cheap signals, the slightest deviation away from the pooling equilibrium towards Horn behaviour by *either* of the populations suffices for the Horn equilibrium to evolve. Because signals are costless, however, there is an evolutionary path through evolutionary drift from the pooling equilibrium to the anti-Horn equilibrium as well. Starting from the pooling equilibrium, given that signals are costless, some senders could drift to following an anti-Horn strategy; if a sufficient number of them do, the anti-Horn equilibrium evolves. Nevertheless, if we assume that a small proportion of the population changes its strategy at any given time, then it is clear that a Horn equilibrium is more likely to evolve than an anti-Horn equilibrium. Also, the basin of attraction of the pooling equilibrium is larger with respect to the Horn equilibrium than with respect to the anti-Horn equilibrium. For the modified Horn game, we again obtain the mirror case of the standard Horn game (Figure A4).

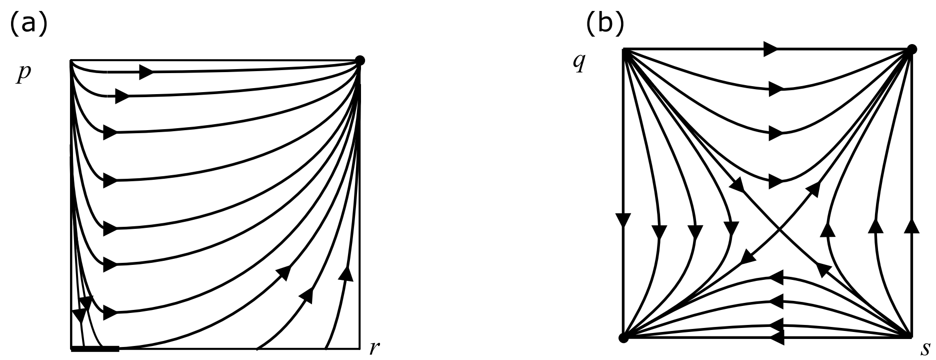


Figure A1 Standard Horn game with costly signals (i = F, d > 0; game 2) (a) Pooling equilibrium vs. Horn equilibrium (Table 1);(b) Pooling equilibrium vs. anti-Horn equilibrium (Table 2).

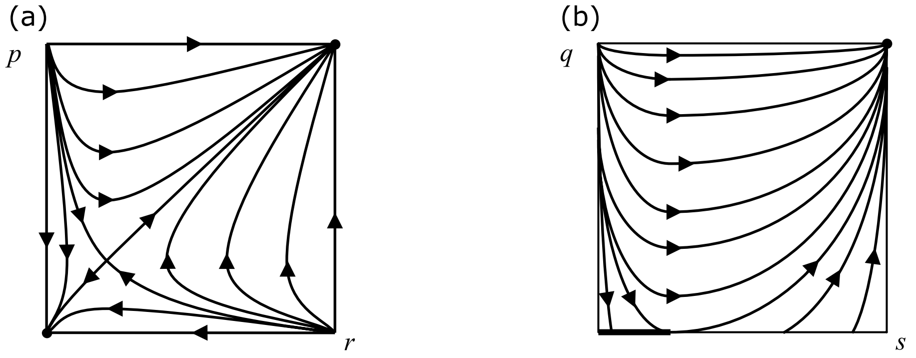


Figure A2 Modified Horn game with costly signals (i = I, d > 0; game 4). Streamlines under replicator dynamics for (a) pooling equilibrium vs. Horn equilibrium (Table 1);(b) pooling equilibrium vs. anti-Horn equilibrium (Table 2).

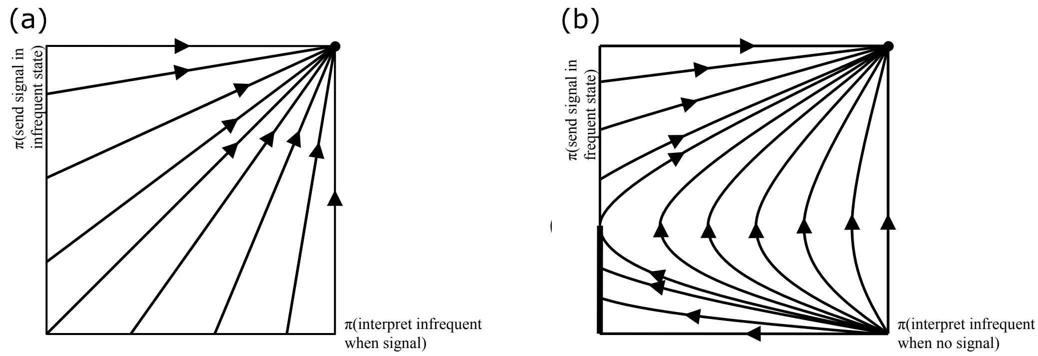


Figure A3 Standard Horn game with cheap signals (i = F, d = 0; game 2'). Streamlines under replicator dynamics for (a) pooling equilibrium vs. Horn equilibrium (Table 1); (b) pooling equilibrium vs. anti-Horn equilibrium (Table 2).

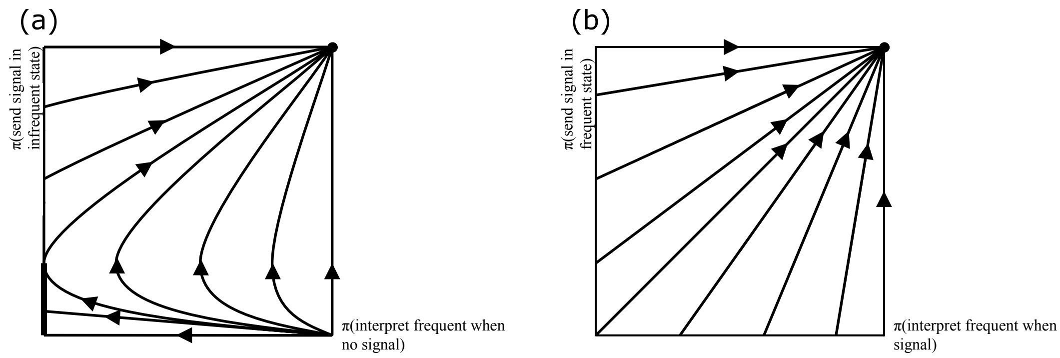


Figure A4 Modified Horn game with cheap signals (i = I, d = 0; game 4'). Streamlines under replicator dynamics for (a) pooling equilibrium vs. Horn equilibrium (Table 1); (b) pooling equilibrium vs. anti-Horn equilibrium (Table 2).

Appendix B

	Session 1				Session 2				Session 3				Session 4			
order	game	H	AH	N	game	H	AH	N	game	H	AH	N	game	H	AH	N
I	1	U : 78% ; D : 22%			3	U : 72% ; D : 27%			2’	20%	63%	7%	4	73%	14%	13%
II	2	87%	0%	13%	4	59%	35%	6%	4’	22%	63%	15%	2	85%	0%	15%
III	3	U : 56% ; D : 44%			1	U : 90% ; D : 10%			2	67%	28%	5%	4’	71%	19%	10%
IV	4	88%	2%	10%	2	77%	21%	2%	4	68%	22%	10%	2’	77%	16%	7%

	Session 5				Session 6				Session 7				Session 8			
order	game	H	AH	N	game	H	AH	N	game	H	AH	N	game	H	AH	N
I	1	U :89% ; D : 11%			3	U : 40% ; D : 60%			2	84%	7%	9%	4’	41%	38%	21%
II	2’	69%	22%	9%	4’	19%	70%	11%	4	82%	16%	2%	2’	65%	12%	23%
III	3	U : 44% ; D : 56%			1	U : 85% ; D : 15%			2’	88%	10%	2%	4	83%	7%	10%
IV	4’	76%	20%	4%	2’	19%	73%	8%	4’	88%	10%	2%	2	91%	4%	5%

Table 4: Games 1 and 3: frequency that U or D is played in 10 periods. Games 2, 2’, 4, 4’: frequency of Horn (H), Anti-Horn (AH), Failure to communicate (N, = neither Horn nor anti-Horn) over all 20 periods. Roman numbers denote the order in which the sessions were played.

Appendix C

Session	**Treatment 1**	**Treatment 2**	**Treatment 3**	**Treatment 4**
1	Game 1 (pooling)	Game 2 (standard, costs)	Game 3 (pooling)	Game 4 (modified, costs)
2	Game 3 (pooling)	Game 4 (modified, costs)	Game 1 (pooling)	Game 2 (standard, costs)
3	Game 2’ (standard, no costs)	Game 4’ (modified, no costs)	Game 2 (standard, costs)	Game 4 (modified, costs)
4	Game 4 (modified, costs)	Game 2 (standard, costs)	Game 4’ (modified, no costs)	Game 2’ (standard, no costs)
5	Game 1 (pooling)	Game 2’ (standard, no costs)	Game 3 (pooling)	Game 4’ (modified, no costs)
6	Game 3 (pooling)	Game 4’ (modified, no costs)	Game 1 (pooling)	Game 2’ (standard, no costs)
7	Game 2 (standard, costs)	Game 4 (modified, costs)	Game 2’ (standard, no costs)	Game 4’ (modified, no costs)
8	Game 4’ (modified,no costs)	Game 2’ (standard, no costs)	Game 4 (modified, costs)	Game 2 (standard, costs)

Table 5: Sequence in which the treatments were run.