Evolutionary Explanations of Indicatives and Imperatives

Abstract

Recently there has been some interest in studying the explanation of meaning by using signaling games. I shall argue that the meaning of signals in signaling games remains sufficiently unclear to motivate further investigation. In particular, the possibility of distinguishing imperatives and indicatives at a fundamental level will be explored. Thereby I am trying to preserve the generality of the signaling games framework while bringing it closer to human languages. A number of convergence results for the evolutionary dynamics of our models will be proved.

1 Introduction

The main goal of this study is to investigate whether a distinction between indicatives and imperatives can be drawn at a very basic level. This will be done by building on work in evolutionary game theory where signaling games serve as a point of departure for investigating reference and meaning (see, for example, Lewis 1969; Crawford and Sobel 1982; Skyrms 1996; Nowak and Krakauer 1999; Harms 2004a; Komarova and Niyogi 2004; van Rooy 2004). In signaling games information must be encoded and decoded correctly in order to facilitate social coordination. A sender observes a state of the world while a receiver responds to the sender's signal. To coordinate behavior properly the signals must have some kind of already established *meaning*.

Harms (2000, 2004a), who is further developing Millikan's (1984) teleosemantics with game theoretic tools, explains the meaning of those signals in terms of *primitive content*. A signal has primitive content if it both tracks the environment and motivates behavior. The tracking function of the signal specifies the conditions under which it is true to utter it. The motivating function characterizes which kind of behavior follows from the signal.

Primitive content enables us to characterize animal signals like warning cries semantically although they are not translateable into human languages (Harms 2004b). This is due to the fact that many forms of non-human communication cannot be explained in terms of propositional content. To see this, note that in philosophical linguistics the meaning of indicatives

and imperatives is explained in terms of propositions with subject-predicate structure. Propositions can be thought of as abstract, truth-bearing units. A proposition may be expressed in different modes like the indicative mode or the imperative mode. The indicative mode refers to (*indicates*) a state of the world. The imperative mode, however, *indicates* an action or an outcome. Thus, propositions are fundamentally indicative since the meaning of imperatives derives from indication by turning it to the purpose of commanding.

According to Harms (2000, 2004a, 2004b), primitive content is able to explain our common sense notion of correspondence truth and the meaning of normative statements. Ultimately, however, we need to understand why norms can be expressed as imperatives while knowledge-oriented systems, like science, use indicatives. Where do indicatives and imperatives part? Do some imperatives have a grounding similar to that of indicatives?[1] Studying these questions without imposing propositional structure seems to be the right place to start.

The results elaborated in this study are based on the following observation: The main difference between indicatives and imperatives is that the emphasis in the meaning of imperatives is to motivate behavior while the emphasis in the meaning of indicatives is to indicate some state of the world. Of course, an imperative may relate to the world in some way and an indicative may motivate behavior. But imperatives lead to behavior directly and refer to world states only indirectly. Indicatives may lead indirectly to behavior but relate to the world directly. "Directly" and "indirectly" are not to be understood as strict, discrete categories. It is a matter of degree how directness is judged.

Lewis (1969) suggested that the difference between indicatives and imperatives is with deliberation on the part of the sender or the receiver. I propose that by deliberation we may understand, in the context of signaling games, any mechanism that processes information inputs and eventually leads to an output (like, e.g., a decision). Deliberation, in this sense, explains whether relations between messages and states or behaviors can be called direct or indirect. If information processing has enough behavioral consequences, then it can be employed as a strategy in some suitable class of signaling games. This class of signaling games represents a further step in the evolutionary sequence of the emergence of meaningful communication.

In Section 2 we briefly review some facts about simple signaling games. Section 3 continues by discussing primitive content. In section 4 I argue informally that indicatives and imperatives can be distinguished on a fundamental level. Section 5 introduces a first formal model. Section 6 presents a general formal analysis of the class of signaling games that induce indicatives and imperatives. Section 7 develops a second model. Section 8 concludes by discussing the relation between our results and the use of imperatives in normative systems and the use of indicatives in science.

s_1	m_1 if σ_1, m_2 if σ_2
s_2	m_2 if σ_1, m_1 if σ_2
s_3	m_1 if σ_1, m_1 if σ_2
s_4	m_2 if σ_1, m_2 if σ_2

r_1	α_1 if m_1, α_2 if m_2
r_2	α_2 if m_1, α_1 if m_2
r_3	α_1 if m_1, α_1 if m_2
r_4	α_2 if m_1, α_2 if m_2

Figure 1: Sender strategies and receiver strategies

2 Signaling Games

A simple signaling game consists of two players, the sender and the receiver, n states, n acts and n signals. The sender observes the state of the world and sends one of the n signals. The receiver has to choose an act. It is assumed that each act is a proper response to exactly one state. Moreover, the sender and the receiver get the same payoff for each outcome. Their payoff is $a > 0$ if the receiver responds correctly. Otherwise both get 0. Thus, the sender and the receiver have a common interest in coordinating states and acts. To do this optimally, their combined strategies must constitute a *signaling system*. A signaling system corresponds to a combination of a one-to-one mapping, s, from the set of states S to the set of messages, M, and a one-to-one mapping, r, from M to the set of acts, A, such that the composition $r \circ s$ associates each state σ_i with the state α_i, $i = 1, \ldots, n$.

Suppose $n = 2$, $a = 1$, and that each of the two states occurs with equal probability.[2] Then there are four possible sender strategies and four possible receiver strategies as shown in Figure 1. The payoffs are as in Figure 2. (s_1, r_1) and (s_2, r_2) are two strict Nash equilibria. They are identical to the two signaling system strategies. There are also four non-strict pure Nash equilibria, (s_3, r_3), (s_3, r_4), (s_4, r_3) and (s_4, r_4). In each of these outcomes the sender sends the same signal regardless of the state and the receiver chooses the same act regardless of the signal. Thus there are two stable outcomes where the agents communicate and four stable, but less desirable outcomes where no information is transmitted. If individuals are in one of the latter states, it is hard to see how they could get to one of the signaling systems *without communication*.

	r_1	r_2	r_3	r_4
s_1	1	0	$\frac{1}{2}$	$\frac{1}{2}$
s_2	0	1	$\frac{1}{2}$	$\frac{1}{2}$
s_3	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
s_4	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

Figure 2: A simple signaling game in strategic form (the payoffs are the same for both players)

Skyrms (1996) studies this signaling game by simulating a corresponding evolutionary dynamics, the one-population replicator dynamics. Skyrms (2000) obtains analytical results for a simplified version of these dynamics where only three types—one signaling system type and two anti-signaling system types—are present. Huttegger (forthcoming) provides a more general analysis of simple signaling games with an equal number of states, acts and messages, and for signaling games involving probabilistic associations between states, signals, and acts. It can be shown that the one-population replicator dynamics will almost surely converge to a signaling system type in simple signaling games with 2 states, acts and signals if the probability for the occurrence of the first state equals the probability for the occurrence of the second state. If this condition fails or if there are more than 2 signals, states and acts, there are non-communicating or partially communicating polymorphisms of types (each pure type can be destabilized by a signaling system type, but not every mixture of types). This suggests that mutations will carry a population away from the suboptimal polymorphisms.

The evolutionary viewpoint to explain the emergence of meaning in signaling games avoids some difficulties any rational choice approach faces.[3] In particular, evolutionary accounts for the explanation of meaning do not rely on assumptions about an already existing common understanding of relevant aspects of the signaling problem and the other players. The replicator equations are agnostic about the cognitive capacities the agents of the population might possess individually. The dynamics is driven by the performance of types with respect to the average payoff in the population. Thus, one of the strengths of the replicator dynamics as an idealized model is its compatibility with diverse specifications of the individual agents.

We should expect that some kind of evolutionary dynamics will allow communication to evolve when we look at the biological evidence we have. One of the best known examples of a signaling system in animals are the predator alarm calls of vervet monkeys (Cheney and Seyfarth 1990). There is a huge number of other examples of signaling systems for various animal species (Snowdon 1990; Hauser 1997; Maynard Smith and Harper 2003). Moreover, signaling systems can already be observed on the level of microorganisms (England et al. 1999; Crespi 2001).

The structure of human languages is, of course, far more complex than the structure of the signaling systems mentioned so far. I do not claim that all aspects of human language can be captured by viewing them as signaling systems. Still, there is at least one functional aspect of human language that can fundamentally be expressed in terms of signaling systems: communication facilitates social coordination. Human languages share this function with less complex signaling systems and can thus be viewed in a similar way from the standpoint of social coordination. From this point of view, the replicator equations provide a partial explanation for the evolution of language in a very simplified, but still interesting way.[4]

3 The Meaning of Signals in Signaling Games

As long as no signaling system or convention is established in a population, signals have no meaning. E.g., we do not want to speak about the meaning of a signal if the sender sends this signal regardless of the state of the world. On the other hand, meaning may be considered as a property of signals in equilibrium. If almost all individuals play according to a signaling system, then signals are representations of parts of the world and have these parts as contents. To be more specific, signals in a signaling system refer to a state of the world and to an act that is a proper response to this state. We will say that signals in signaling systems refer to state-act pairs.[5]

Harms (2004a, 2004b) proposes that signals in simple signaling games have primitive content. This means that two sorts of conventions apply in signaling systems: extensional tracking conventions and intensional consequence conventions. The former specify the state of the world to which the signal corresponds to or which make it true. The latter specify the consequences or the behavior that is the proper response to the signal.

Harms adopts the terms "extensional" and "intensional" for signals although they are usually associated with the meaning of words, and words are, in general, not the proper analogues to signals in human languages. Moreover, something essential is also missing in any analogy between signals and sentences. Signals in simple signaling games have primitive content. They refer to state-act pairs. This is not true of sentences, however. In general, sentences indicate acts or command actions, but not both.

These considerations lead us to our main object of study. In signaling systems of simple signaling games it is not at all clear whether signals are indicative or imperative. This is due to the fact that simple signaling games are not structured in a way that would allow us to talk about indicative and imperative signals. The next four sections are devoted to the study of models that may give rise to indicative and imperative signals. Before turning our attention to those models, allow me to point out some philosophical and scientific consequences our investigation has. Those consequences also serve to motivate our models.

First, our study allows us to further elaborate the answers offered to a skeptical philosophy of language. As we have noted in the previous section, Skyrms (1996, 2000) and Huttegger (forthcoming) provide some results which indicate that signaling systems emerge with high probability under reasonable evolutionary dynamics. A skeptic might not only call the details of the model into question.[6] A skeptic might also cast doubt on the explanatory power of these results by claiming that an account of the emergence of language conventions that does not include some of the most basic features of human languages falls short of its main goal, the explanation of meaning. If the distinction between indicative and imperative sentences is taken to be a basic semantic feature of human languages, then the skeptic might be

right.

Second, we may be able to better understand how indicatives compared to imperatives relate to the world. If a signal has the function to represent either a particular state of the world σ or a particular act α, then this signal has, respectively, σ or α as its content. This is just another way to say that the signal means σ or α. This leads to a number of problems concerning (i) the explanation of how correspondence between representations and the world is established; and (ii) the problem whether statements concerning moral wrongness or rightness and concerning the justification of acts can be called true or false with respect to some objective standards. Harms' (2004a, 2004b) account of primitive content promises to resolve these problems. With the help of our results we will be able to address some issues surrounding the use of indicatives and imperatives in knowledge-oriented and behavior-oriented systems. How this might be achieved is outlined in section 8.

Third, one of the main challenges for an evolutionary account of language is the problem of filling the gaps between simple communication systems and human language (Maynard Smith and Szathmáry 1995; Maynard Smith and Harper 2003). Thus, our account might be valuable in making precise just where meaning as we find it in indicatives and imperatives departs from primitive content.

Finally, on the scientific side we will be able to reinterpret some animal signaling systems (some of which will be considered in the next section) in our model. As we shall see, our models will allow a more coherent interpretation of signals than was possible with standard signaling games.

4 Indicatives and Imperatives as Interpretations of Signals

Lewis (1969) distinguishes between signals-that, signals-to and neutral signals. More specifically, the meaning of signals in simple signaling games can sometimes reasonably be given as a signal-that (indicative), as a signal-to (imperative) or as both (neutral). Lewis also mentions criteria to draw these distinctions. They are based on whether the sender or the receiver has to deliberate in order to achieve the optimal outcome. If the sender does not have to deliberate but the receiver must deliberate, then the signal is indicative. If the receiver must not deliberate but the sender has to deliberate, then the signal is imperative. If both descriptions are compatible with the signaling behavior of the players, then the signal is neutral.

It will be useful to investigate whether interpreting signals as indicatives or imperatives can enhance our understanding of actual signaling systems. There are a number of examples where these interpretations seem to be possible.

A closer examination of the vervet signaling system (Cheney and Seyfarth

1990) suggests that we might interpret some of the signals as indicatives and others as imperatives. E.g., the proper response to the snake alarm call is to stand bipedally and to look around. Ultimately, the *receivers* of the snake alarm call have thus to decide what to do. Hence, the snake alarm call might be interpreted as indicative. The leopard alarm call, on the other hand, usually results in the vervet monkeys running up trees immediately. So we might interpret the leopard alarm call as imperative because the *sender* has to classify the situation whereas the receiver must react very fast.

The California ground squirrel seems to have different alarm calls for ground predators and aerial predators (Owings and Henessy 1984; Snowdon 1990). Terrestrial predators usually approach slowly and their approach is closely monitored by the squirrels whereas aerial predators show up fast most of the time. Owings and Henessy (1984) report that alarm calls for aerial predators are sometimes given to terrestrial predators and sometimes the alarm call for ground predators is given for aerial predators. This happens when aerial predators are spotted by the squirrels while they are still far away or when ground predators are already close before one of the squirrels observes their presence. Thus, an interpretation of the two alarm calls in terms of referring to aerial predators or ground predators is ambiguous. We might, however, interpret the alarm calls as indicative or imperative signals. An indicative signal is used if a predator approaches slowly and an imperative signal is used if it approaches fast and there is a much greater urgency to respond. (Note that the usage of indicative and imperative signals arises out of the structure of the underlying situation.)

In a study on symbolic communication, Boesch (1991) describes the signaling behavior of the chimpanzees of the Tai national park. In one of the chimp groups the alpha male, Brutus, sometimes drummed on a tree. Brutus' drumming was related to three different messages. After drumming on two trees consecutively, the group changed direction and went on in the direction between the two trees. If Brutus was drumming twice on the same tree, the group rested for about one hour. By combining these two messages, i.e. by first drumming twice on a tree and then once on a second tree, Brutus indicated a short rest and the direction the group should take after the rest.

This example is remarkable in at least two ways. First, Brutus was able to combine two signals to form a new message. This can be regarded as a very simple "syntax". And second, it seems hard not to interpret these signals as imperative. An interpretation in purely indicative terms might be possible but would turn out to be rather complicated. It seems that the emphasis of these signals is on motivating behavior.

These examples show that a distinction between indicative and imperative communication might already be possible at the level of animal signaling systems. We can also think of human examples. Your doctor usually does not tell you what is wrong with your health, but what you ought to do to get healthy again. She decides what is best for your health since you usually

cannot make this decision yourself. On the other hand, in a situation where somebody has information that can help you make a decision you don't want her to tell you what to do, but to reveal the information.

In simple signaling games a distinction between indicative and imperative signals that is not just an interpretation of the signals seems to be impossible. Simple signaling games do not have enough structure. But Lewis' considerations show us one possibility to give a signaling game more structure. The underlying intuition is that some state-act pairs require the sender to deliberate, whereas the receiver must not deliberate but just has to act. This might be because the receiver has to act fast, or because the gathering of information would be of no use for the receiver to make the right decision (like in the doctor example). Other state-act pairs require the sender not to deliberate and the receiver to decide what would be the best thing to do.

This relates in an obvious way to the main distinction we have drawn between imperatives and indicatives. Indicatives motivate behavior only indirectly because they need to be combined with other indicatives to determine efficient behavior (see also Harms 2004b). Imperatives relate to the world indirectly via the sender's information processing mechanism.[7]

The rest of this paper is devoted to developing two models that try to capture these ideas and to present some convergence results on the class of games they belong to.

5 A First Model

The class of games we shall consider is based on a combination of two coordination problems. One is a state-act coordination problem that underlies simple signaling games. In a state-act coordination problem exactly one of the acts is the proper response to each state of the world. Accordingly, individuals only get a positive payoff if the right act is chosen in response to a particular state. This scenario is illustrated in Figure 3. The second

	α_1	α_2
σ_1	a	0
σ_2	0	a

Figure 3: A state-act coordination problem with $a > 0$

coordination problem is asymmetric. Either player may deliberate or not deliberate, but if both do the same they get no payoff. The second coordination problem is shown in Figure 4.

Our first model is based on a combination of the two coordination problems. There are two states of the world, σ_1 and σ_2, and two corresponding acts, α_1 and α_2. In addition, each player has to decide if she deliberates or if she doesn't deliberate. (σ_1, α_1) is a state-act pair that requires the

	deliberate	don't deliberate
deliberate	0	b
don't deliberate	b	0

Figure 4: An asymmetric coordination problem with $b > 0$

row player to deliberate and the column player to act without deliberating. (σ_2, α_2) is a state-act pair where the row player must not deliberate and the column player is required to deliberate. This payoff matrix is illustrated in Figure 5.

	d and α_1	n and α_1	d and α_2	n and α_2
If σ_1, d	0	1	0	0
If σ_1, n	0	0	0	0
If σ_2, d	0	0	0	0
If σ_2, n	0	0	1	0

Figure 5: A coordination problem where d stands for deliberate and n for don't deliberate

So far we have specified the coordination problem which underlies a signaling game that involves deliberation. For the signaling game itself we suppose that the sender has to decide whether to deliberate or not before sending a signal. The receiver has to decide whether to choose an act deliberately or not. A sender strategy specifies, for each state of the world, whether the player spends some time deliberating and what signal is sent. A receiver strategy specifies, for each message, whether the receiver chooses an act deliberately and what act is chosen. Thus there are sixteen sender strategies and sixteen receiver strategies. There are four groups within each of these, those who never deliberate, those who always deliberate, those who switch between deliberating and not deliberating in the wrong way and those who switch in the right way. Within these four groups there are types who employ signaling system strategies and types who always send the same signal. For a list of the strategies see Figure 6. We will assume that $\mathbb{P}(\sigma_1) = \mathbb{P}(\sigma_2)$ to avoid the difficulties mentioned in Section 2 for our first analysis. Under these assumptions, the 16×16 payoff matrix can easily be computed.

Before we are going to analyze this model let us try to motivate the additional structure imposed on signaling games. For the new signaling game we assume that deliberation has enough behavioral consequences to employ it as a strategy in a game. That is, deliberation comes with costs. It may cause delays, for instance. Or it may never lead to an appropriate decision if the player lacks necessary information in principle. In a similar way, choosing not to deliberate might be costly. It may, for example, forfeit further moves because decisions relevant for those moves were not made as a result of deliberation. For a baseline model this characterization of

Sender strategies	
s_1	0000
s_2	0001
s_3	0010
s_4	0011
s_5	0100
s_6	0101
s_7	0110
s_8	0111
s_9	1000
s_{10}	1001
s_{11}	1010
s_{12}	1011
s_{13}	1100
s_{14}	1101
s_{15}	1110
s_{16}	1111

Receiver strategies	
r_1	0000
r_2	0001
r_3	0010
r_4	0011
r_5	0100
r_6	0101
r_7	0110
r_8	0111
r_9	1000
r_{10}	1001
r_{11}	1010
r_{12}	1011
r_{13}	1100
r_{14}	1101
r_{15}	1110
r_{16}	1111

Figure 6: Sender and receiver strategies in Model 1. 1 or 0 at the first place mean that the sender deliberates or doesn't deliberate if σ_1 occurs. The second place specifies the same for σ_2. At the third place, 1 means that the sender signals m_1 if σ_1 and 0 means that she sends m_2 in this case. The fourth place specifies the same for state σ_2. A receiver strategy is coded in the same way except that the first and the third place specify what happens if m_1 was sent; the second and fourth place specify the same if m_2.

the behavioral consequences of deliberation seems to be enough. In more advanced models, however, structural effects of choosing to deliberate or choosing not to deliberate should be made explicit.

In particular, we do not specify how the deliberational process might look like. It may be something that requires more or less computational capacities. It might even be a non-cognitive mechanism that makes a decision according to a number of inputs (think of a cell that reacts in a particular way after receiving some signals and in response to other environmental states). It is best to think about it in terms of which agent acquires and processes information. If the acquiring and processing of information is *entirely* on the side of one of them, then we say that the respective agent deliberates and that the other does not deliberate. We do not assume that deliberation always leads to the right decision. Regardless of the deliberation mechanism our agents employ, deliberation leads to the right decision with a certain probability. The payoffs should thus be understood as expected payoffs. What we do assume is that agents who start to deliberate or who don't deliberate when the situation requires them to do the other thing get no

payoff. We justify this again by expected payoffs. If the sender is required to deliberate after the occurrence of a particular state and before sending a signal but decides not to deliberate, we assume that the chances for the receiver to nevertheless choose the right act are very low. The same holds for the receiver and for situations that require the agent not to deliberate. This assumption will be relaxed in our second model.

For example, a California ground squirrel that spots a nearby predator may deliberate for some time whether the perception it had resembles a predator. As soon as it gives the alarm call for a nearby predator, it would be fatal for the other squirrels to start deliberating. In this situation they are supposed to react fast and hide in holes. This response would not be appropriate for predators which are far away. In this case receivers have to deliberate more than senders.

In order to report simulation results on our first model we have to introduce some concepts from dynamical system theory (Hirsch and Smale 1974; Hirsch et al. 2004). Let $\mathbf{x} \in \mathbb{R}^n$ and $f : W \to W$ where $W \subset \mathbb{R}^n$. Then a *discrete time dynamical system* on W is given by

$$\mathbf{x}' = f(\mathbf{x}),$$

where $\mathbf{x}$ is the current state of the system and $\mathbf{x}'$ is the state at the next time step. A point $\bar{\mathbf{x}}$ is called a *fixed point* of the discrete time dynamical system if $f(\bar{\mathbf{x}}) = \bar{\mathbf{x}}$. If the system reaches a fixed point, then it remains there forever. $\bar{\mathbf{x}}$ is an *attractor* or an *attracting fixed point* for f if there is a neighborhood U of $\bar{\mathbf{x}}$ such that every orbit starting in U converges to $\bar{\mathbf{x}}$. The set of all points converging to $\bar{\mathbf{x}}$ is called its *basin of attraction*. $\bar{\mathbf{x}}$ is a *source* or a *repelling fixed point* if there is a neighborhood U such that all orbits (except $\bar{\mathbf{x}}$) leave U under iteration of f. $\bar{\mathbf{x}}$ is called *neutral* if it is neither attracting nor repelling.

The *discrete time replicator dynamics for two populations* can be used as a model for biological as well as cultural evolution (like other forms of the replicator dynamics; for more on the replicator dynamics see Weibull 1995 and Hofbauer and Sigmund 1998).[8] These dynamics are given by

$$x_i' = x_i \frac{\alpha + u(x_i, \mathbf{y})}{\alpha + u(\mathbf{x}, \mathbf{y})} \quad \text{and} \quad y_j' = y_j \frac{\alpha + u(y_j, \mathbf{x})}{\alpha + u(\mathbf{y}, \mathbf{x})}, \tag{1}$$

where x_i and y_j are the frequencies of type i senders and type j receivers at a particular time and x_i' and y_j' are their frequencies at the next time step. The state of the sender population is given by the vector of frequencies $\mathbf{x} = (x_1, \ldots, x_n)$ of the sender types. Likewise, the state of the receiver population is given by $\mathbf{y} = (y_1, \ldots, y_n)$. α is the common background fitness of individuals in both populations. $u(x_i, \mathbf{y})$ and $u(y_j, \mathbf{x})$ are the payoffs to i and j when the current sender population state is $\mathbf{x}$ and the current receiver population state is $\mathbf{y}$. $u(\mathbf{x}, \mathbf{y})$ and $u(\mathbf{y}, \mathbf{x})$ are the respective average payoffs.[9]

In our first model, there are four sender strategies, s_5 to s_8, and four receiver strategies, r_1, r_6, r_{11}, and r_{16}, that never get it right. This is so because individuals of one of those sender types choose to deliberate or not to deliberate wrongly while receivers of one of those receiver types either choose to deliberate or not to deliberate wrongly or they choose the wrong act regardless of the message. These strategies can thus be ignored in the further analysis.

There are 54 Nash equilibria in pure strategies. Only two of them are strict. The two strict Nash equilibria are (s_{10}, r_{10}) and (s_{11}, r_7). Regardless of the state, these outcomes guarantee the players the maximum payoff a. Both are the only combinations of strategies that solve the problem of deliberation-coordination and are signaling systems. All other combinations yield a maximum payoff less than a. This is due either to the fact that they always or never deliberate or because they form no signaling system or because of both.

In simulations for the discrete time replicator dynamics we observe that 100% of the time populations converge to the states corresponding to the two strict Nash equilibria (s_{10}, r_{10}) or to (s_{11}, r_7). They converge to either of them approximately half of the time. These results suggest that (s_{10}, r_{10}) and (s_{11}, r_7) are the only attracting fixed points for (1) and that their basins of attraction are of equal size. The other Nash equilibria seem to correspond to non-attracting fixed points. To obtain analytical results, we will study the continuous time replicator dynamics of our model. This will be done in the next section.

6 Simple Signaling Games With Deliberation

Although we will only consider a special case of signaling games with deliberation in this paper, we will give a definition for a more general class first. The definition of simple signaling games with deliberation is based on state-act coordination problems like the one illustrated in Figure 3. More generally, $\Pi_n = \langle S, A, u\rangle$ is a *n-state-act coordination problem* if and only if $S = \{\sigma_1, \ldots, \sigma_n\}$ is a set of n distinct states of the world, $A = \{\alpha_1, \ldots, \alpha_n\}$ is a set of n distinct acts, and u is a function that determines the utility of each state-act pair such that $u(\sigma_i, \alpha_j) = \delta_{ij} a_i$ where $\delta_{ii} = 1$ and $\delta ij = 0$ for $j \neq i$. a_i is a real number that depends on the state. For simplicity, we will assume that $a_i = 1$ for all i. Let $\mathcal{P} = \{p_1, \ldots, p_n\}$ be a probability distribution over S.

Definition 1 (simple signaling game with deliberation) *Let Π_n be a n-state-act coordination problem, let $M = \{m_1, \ldots, m_n\}$ be a set of n messages, $\mathcal{P}$ a probability distribution over the states and $D = \{d, n\}$ the set of deliberation states. Then a d-signaling game Σ_n^d is a triplet $\langle I, \{S_i\}_{i\in I}, \{u_i\}_{i\in I}\rangle$ where*

1. *$I = \{1, 2\}$ is the set of players, the sender, 1, and the receiver, 2;*

2. *S_i, $i = 1, 2$, is the set of strategies generated by Π_n and D as follows: $S_1 = \{s_k | s_k : S \to D \times M\}$ and $S_2 = \{r_l | r_l : M \to D \times A\}$;*

3. *$u_i, i = 1, 2$ are the players' utility functions generated by Π_n as follows:*

$$u_i(s_k, r_l) = \sum_{j=1}^{n} p_j \cdot u\left(\sigma_j, (r_l \circ c \circ s_k)(\sigma_j)\right), \quad i = 1, 2,$$

where $c : D \times M \to M$ is the function defined by $c(\cdot, m_k) = m_k$ for $k = 1, \ldots, n$.

Thus, a d-signaling game is an asymmetric two-player game. Condition 2 states that the set of possible sender strategies consists of all functions from the set of states to the product of the set of deliberation states and the set of messages. Similarly, the set of possible receiver strategies consists of all possible functions from the set of messages to the product of the set of deliberation states and the set of actions. Condition 3 specifies the players' payoff functions. This specification employs a function c that "cuts away" the deliberational state of the sender. What the payoffs in fact are depends on the specification of how deliberational states influence the payoff structure of the game. This flexibility in choosing how deliberational states effect the players' payoffs makes it possible to model situations where deliberational states influence payoffs in varying degrees.

In the previous section we have used the discrete time two-population replicator dynamics (1) to obtain some simulation results for a signaling game Σ_2^d. To obtain analytical results it is more convenient to work with continuous time versions of the two-population replicator dynamics. There are two principal versions of them. We will employ both in our subsequent analysis.

The standard version of the two-population replicator dynamics is a coupled system of differential equations:

$$\begin{aligned} \frac{d\,x_i}{d\,t} &= x_i\left(u(x_i, \mathbf{y}) - u(\mathbf{x}, \mathbf{y})\right) \\ \frac{d\,y_j}{d\,t} &= y_j\left(u(y_j, \mathbf{x}) - u(\mathbf{y}, \mathbf{x})\right) \end{aligned} \tag{2}$$

(see Hofbauer and Sigmund 1998). The variables and the state space of this system are the same as for the discrete time system (1). There is a second version of the continuous time two-population replicator dynamics whose qualitative behavior is in general different from the qualitative behavior of the dynamics (2). It was introduced by Maynard Smith (1982) and is formally similar to the discrete time replicator dynamics (1) since it involves

normalization by the mean payoff:

$$\begin{aligned} \frac{d\,x_i}{d\,t} &= x_i \frac{u(x_i,\mathbf{y}) - u(\mathbf{x},\mathbf{y})}{u(\mathbf{x},\mathbf{y})} \\ \frac{d\,y_j}{d\,t} &= y_j \frac{u(y_j,\mathbf{x}) - u(\mathbf{y},\mathbf{x})}{u(\mathbf{y},\mathbf{x})} \end{aligned} \tag{3}$$

The appendix provides some propositions on the dynamics of d-signaling games. In particular, the following basic convergence result for the dynamics (2) and (3) is proved. (Notice that the convergence results hold for both versions of the two-population replicator dynamics.)

Theorem 1 *Let Σ_2^d be a d-signaling game where for both σ_1 and σ_2 exactly one pair of deliberational states (one for the sender and one for the receiver) is optimal. Let $\mathbb{P}(\sigma_1) = \mathbb{P}(\sigma_2)$. If initially all types are present, then almost every solution for (2) and (3) converges to a strict Nash equilibrium of Σ_2^d.*

Theorem 1 shows that the simulation results reported in the previous section hold analytically in a more general setting. The theorem continues to hold for a more general class of games than just partnership games. Games in this more general class are structurally similar to partnership games (they are called rescaled partnership games) but do not require that the players get the same payoff for each outcome.[10]

Concerning our first model we obtain a corollary by applying Theorem 1 and observing that the vector fields (2) and (3) are invariant under permutations of the points corresponding to the strict Nash equilibria.

Theorem 2 *In Σ_2^d with payoffs specified in the previous section, almost every solution converges to (s_{10}, r_{10}) or to (s_{11}, r_7) for (2) and (3). Moreover, their basins of attraction are of equal size.*

7 A Second Model

The payoffs in our first model are quite rigid. They require the players to coordinate their deliberational activities strictly. It is possible to relax this requirement while preserving enough of the original structure of the game. This will allow us to still talk meaningfully about imperatives and indicatives in many cases.

We will change our first model according to the following considerations: If the sender fails to deliberate in state σ_1 but the receiver is able to deliberate and to choose the right act, then they may still get it right. For state σ_2, if the sender starts to deliberate and the receiver is nonetheless fast or lucky enough to choose the right act, there might be some chance to get a payoff as well. This gives rise to the payoff structure that is illustrated in Figure 7.

	d and α_1	n and α_1	d and α_2	n and α_2
σ_1, d	0	1	0	0
σ_1, n	$\frac{1}{2}$	0	0	0
σ_2, d	0	0	0	$\frac{1}{2}$
σ_2, n	0	0	1	0

Figure 7: Model 2; d stands for deliberate and n for not deliberate

The game corresponding to this payoff matrix has 28 Nash equilibria in pure strategies, 6 of which are strict. The sender strategies or receiver strategies that always got it wrong in model 1 have disappeared. But still none of these strategies is part of a strict Nash equilibrium. The strict Nash equilibria are (s_{10}, r_{10}), (s_{11}, r_{10}), (s_2, r_{14}), (s_3, r_{15}), (s_{14}, r_2) and (s_{15}, r_3). The first two give a payoff of 1. The latter four give a payoff of $\frac{3}{4}$. They are either characterized by a sender strategy that never deliberates and a receiver strategy that always deliberates or the other way round. But note that the signaling parts of each of these strict Nash equilibria form signaling systems. Thus, we have as an important side result that *the signaling parts of all strict Nash equilibria constitute signaling systems.*

In simulations for the discrete time replicator dynamics (1) the population converged to each of (s_{10}, r_{10}) and (s_{11}, r_3) at a rate of about 40%, and to each of the other strict Nash equilibria about 5% of the time. This indicates that the basins of attraction of the latter ones are considerably smaller than the basins of attraction of the first two. Theorem 1 informs us again about convergence.

Theorem 3 *Let Σ_2^d be given by the above payoffs and let $\mathbb{P}(\sigma_1) = \mathbb{P}(\sigma_2)$. Then (s_2, r_{14}), (s_3, r_{15}), (s_{10}, r_{10}), (s_{11}, r_{10}), (s_{14}, r_2) and (s_{15}, r_3) are the only asymptotically stable states under (2) and (3). If initially all types are present, then almost every solution converges to one of the strict Nash equilibria.*

There are a number of interesting limiting cases we can obtain by changing the payoff structure. If the players must coordinate deliberations more precisely than in the second model, we get back to our first model. If the $\frac{1}{2}$ payoffs go to one, the distinction between indicatives and imperatives becomes less important. If the payoffs are such that sender and receiver get the same payoff no matter whether sender and receiver deliberate or do not deliberate, we are back at simple signaling games where a distinction between indicatives and imperatives is not meaningful anymore. As we vary the payoffs, the meaningfulness of this distinction corresponds to the asymptotically stable states and the size of their basins of attraction.

8 Normative and Descriptive Statements

Let us conclude with an application of the previous analysis. The goal of this application is to gain some heuristic understanding of why norms can be expressed as imperatives and why science, on the other hand, uses indicatives.

In the last part of Harms (2004a) the basic ingredients of a naturalistic theory of the meaning of descriptive and, in particular, normative statements are presented. Harms thereby attempts to describe the semantic content of normative statements by two features. First, they are part of signaling systems (which we may call *normative systems*). And second, they enforce rules of behavior. To see the advantages of this point of view, compare it to a more traditional characterization of normative imperatives in terms of propositional content. According to this characterization, norms may be expressed in the indicative or in the imperative mood. But this yields the problem of what is indicated with a normative imperative. More generally, if a proposition is used in the imperative mood, then it might not be clear what is indicated. E.g., the imperative mood "open the door" of the proposition "the door is open" looses its regular indicative function because the door is closed and may never be opened. Essentially the same happens in the context of norms. Most norms may be expressed in the indicative mood, like "stealing is wrong". They may also be turned to the purpose of commanding like in "don't steal this". But even if you steal this, the norm expressed by "stealing is wrong" does not cease to be a norm.

Harms (2004a), on the other hand, argues that normativity comes from primitive content. A normative system specifies the appropriate actions for particular circumstances. In this respect, norms, or normative intuitions, resemble animal warning cries. We may also think of a normative system as an internal control mechanism.

The primitive content of a norm specifies some historically determined standards of how to behave in certain situations. Suppose a convention prescribes to act according to a certain behavioral rule b if situation s occurs. Conventions are fallible. A response to a failure of this convention might be a signaling convention. That is, if an individual fails to act according to b in s, then another (or the same) individual might respond by sending a signal m that means something like "In a situation like s act according to b" or "In a situation like s you ought not act according to b'" if b' was the individual's behavior. This is an example of a second order convention. According to the primitive content of the signal m, there is an extension, namely "not b", that makes the behavior enforcing signal m true. Normative imperatives are, according to Harms (2004a), linguistic proxies to the intensional part of the primitive content of a signal. On this level, a normative system has the function to regulate behavior in the population via the enforcement of conventions. The primitive content of the signals in the normative system is

determined by the history of the population.

The evolution of a normative system may be described similarly to the evolution of other signaling systems (see Section 2). It presupposes special states and acts, however. For a signaling system to be called a normative system the states and acts of the underlying coordination problem must be failures of behavioral conventions and acting according to those conventions, respectively. Thus the evolution of normative systems presupposes already existing behavioral conventions. It may also be the case that a normative signaling system and behavioral conventions coevolve. Studying the coevolution of normative systems and behavioral conventions would be of considerable interest.

The emergence of indicative and imperative signals marks the departure from primitive content. A signal can be used in two different ways depending on the history of the population and the kind of situations it is confronted with. Notice that in the corresponding class of signaling games we do not have to talk about the agents' desires or beliefs. This means that we do not have to characterize an imperative in terms of an agent's desires, i.e. an agent who wants another agent to do something. Our explanation of indicative and imperative signals is thus more basic since we only require there to be agents with *some* information processing mechanism. Agents with beliefs and desires may, of course, be substituted. Our framework is compatible with that.

Often, the emphasis of a normative statement is on motivating behavior since norms enforce rules of behavior. Imperative signals as developed in the previous sections give a first, basic understanding of what the "linguistic proxies" to the primitive content of norms are. If the content of norms is captured by a normative system and if the main function of this system is to motivate behavior, then it will mostly yield imperative signals as outputs. That is, we have a simple signaling system on the one hand and a signaling system that induces imperatives on the other. Both signaling systems have to be related in some way. This situation is more complex than the ones considered in this paper and the literature so far.

To be sure, norms (i.e. a set of specific epistemic standards) are also underlying knowledge oriented systems like science. But the purpose of science, as opposed to, e.g., moral systems, is not to express its underlying epistemic standards by inducing imperatives. That is to say, as long as this underlying normative system is "silent", imperatives do not enter science. At another level, science may be viewed as a signaling system that is about the world. As such, it yields indicatives that inform us about the world.

To summarize, simple signaling games are not enough to describe normative systems completely. But the signaling games studied in the previous sections are able to give us a first explanation of why different systems may yield imperatives or indicatives.

Appendix

I will present a number of results which shed light on the dynamical properties of generalized simple signaling games and which will be used in the proof of Theorem 1. Let me introduce the concept of partnership games first. An asymmetric noncooperative two-player game Γ is a *partnership game* if the payoff matrices (A, B) corresponding to Γ are such that $B = A^t$.

Proposition 1 *Let Σ_n^d be a d-signaling game. Then Σ_n^d is a partnership game.*

Proof. Let us denote that payoff matrix for the sender's payoffs by A and the receiver's payoff matrix by B. Then obviously $B = A^t$ since sender and receiver get the same payoff for each outcome. ■

Proposition 2 *Let Σ_n^d be a d-signaling game. Then the dynamics (2) and (3) for Σ_n^d have the same qualitative behavior.*

Proof. Let Σ_n^d be a d-signaling game. By Proposition 1, Σ_n^d is a partnership game. For partnership games, the average payoffs $u(\mathbf{x}, \mathbf{y})$ and $u(\mathbf{y}, \mathbf{x})$ coincide. To see this, note that

$$u(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot A\mathbf{y} \quad \text{and} \quad u(\mathbf{y}, \mathbf{x}) = \mathbf{y} \cdot B\mathbf{x},$$

where $\cdot$ is the dot product. Since

$$\mathbf{y} \cdot B\mathbf{x} = \mathbf{y} \cdot A^t\mathbf{x} = \mathbf{x} \cdot A\mathbf{y}$$

we have

$$u(\mathbf{x}, \mathbf{y}) = u(\mathbf{y}, \mathbf{x}).$$

Thus (3) involves just a change of velocity compared to (2), but the qualitative behavior of the two is the same. ■

To formulate our next intermediate result, we have to be clear about the notion of evolutionary stability in a two-population model. As Weibull (1995) points out, every reasonably strong analogue to the concept of evolutionary stability for n-population models ($n \geq 2$) coincides with strict Nash equilibria. Accordingly we call a state $(\mathbf{x}, \mathbf{y})$ ($\mathbf{x}$ and $\mathbf{y}$ representing the states in each population, respectively) *evolutionarily stable* if $(\mathbf{x}, \mathbf{y})$ is a strict Nash equilibrium of the underlying asymmetric game. Moreover, we have to introduce the notion of a gradient system. To do this, let V be a twice continuously differentiable function from an open subset U of $\mathbb{R}^n$ to $\mathbb{R}$. Then

$$\frac{d\mathbf{x}}{dt} = \nabla V(\mathbf{x})$$

is a gradient system with *potential* V. If the gradient ∇V is defined relative to the standard inner product for $\mathbb{R}^n$, then

$$\nabla V = \left(\frac{\partial V}{\partial x_1}, \ldots, \frac{\partial V}{\partial x_n} \right) =: \frac{\partial V}{\partial \mathbf{x}}.$$

Notice that due to this relation V gives us a lot of information about the system. The gradient ∇V may also be defined relative to a non-standard inner product for $\mathbb{R}^n$ by considering the dual vector space for $\mathbb{R}^n$ (that is, the space of all linear mappings from $\mathbb{R}^n$ to $\mathbb{R}$). Many of the results for gradient systems defined with respect to the standard inner product for $\mathbb{R}^n$ continue to hold in this more general setting. This is due to the basic equality $\frac{\partial V}{\partial \mathbf{x}}\mathbf{y} = \langle \nabla V, \mathbf{y} \rangle$, where $\mathbf{y} \in \mathbb{R}^n$ and $\langle \cdot, \cdot \rangle$ is an arbitrary inner product. For more information about gradient systems see Hirsch and Smale (1974).

Proposition 3 *Let Γ be a partnership game and (A, A^t) be the corresponding payoff matrices. Then the following two statements are true:*

1. *(2) is a gradient system with $u(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot A\mathbf{y}$ as potential.*
2. *$(\mathbf{p}, \mathbf{q})$ is asymptotically stable for (2) and (3) if and only if $(\mathbf{p}, \mathbf{q})$ is evolutionarily stable.*

Proof. Let Γ be a partnership game and A be the corresponding payoff matrix. A proof of 1 can be found in Hofbauer and Sigmund (1998, Theorem 11.2.2). To prove 2, suppose that $(\mathbf{p}, \mathbf{q})$ is an equilibrium of (2) but not evolutionarily stable for a two population model. Thus, $(\mathbf{p}, \mathbf{q})$ is not a strict Nash equilibrium of Γ. We claim that $(\mathbf{p}, \mathbf{q})$ is not a strict local maximum of $\mathbf{x} \cdot A\mathbf{y}$ and, thus, not asymptotically stable for (2). This conclusion follows from the fact that $\mathbf{x} \cdot A\mathbf{y}$ is also a strict Liapunov function for (2).

To prove that $(\mathbf{p}, \mathbf{q})$ is not a local strict maximum of (2) note that, since $(\mathbf{p}, \mathbf{q})$ is not a strict Nash equilibrium, there is a $\mathbf{s}$ or there is a $\mathbf{r}$ such that $\mathbf{s}$ is an alternative best reply to $\mathbf{p}$ or $\mathbf{r}$ is an alternative best reply to $\mathbf{q}$:

$$\mathbf{s} \cdot A\mathbf{q} = \mathbf{p} \cdot A\mathbf{q} \quad \text{or} \quad \mathbf{p} \cdot A\mathbf{r} = \mathbf{p} \cdot A\mathbf{q}.$$

Suppose $\mathbf{s}$ is an alternative best reply to $\mathbf{p}$. Then every convex combination $\lambda\mathbf{s} + (1-\lambda)\mathbf{p}, 0 \leq \lambda \leq 1$ is also a best reply to $\mathbf{p}$. From this we can conclude that in every neighborhood of $(\mathbf{p}, \mathbf{q})$ there exist $\mathbf{x}, \mathbf{y}$, namely $\mathbf{x} = \lambda\mathbf{s} + (1-\lambda)\mathbf{p}$ and $\mathbf{y} = \mathbf{q}$, such that $(\mathbf{p}, \mathbf{q})$ is no strict maximum of $\mathbf{p} \cdot A\mathbf{q}$. A similar argument applies to the case where $\mathbf{r}$ is an alternative best reply to $\mathbf{q}$.

If $(\mathbf{p}, \mathbf{q})$ is evolutionarily stable, on the other hand, then it is a strict Nash equilibrium. Thus, there is no alternative best reply to $(\mathbf{p}, \mathbf{q})$ and it must therefore be a strict local maximum of $\mathbf{x} \cdot A\mathbf{y}$. This implies that $(\mathbf{p}, \mathbf{q})$ is asymptotically stable. ■

If Σ_n^d is a d-signaling game, then the only evolutionarily stable states are strategy combinations where the sender and the receiver deliberate in the

right way and the signaling parts of their strategies form a signaling system. In our first model there are two such strategy combinations, (s_{10}, r_{10}) and (s_{11}, r_7). Hence, these strategy combinations are the only asymptotically stable states for (2) and (3).

Lemma 1 *Let Σ_2^d be a signaling game with deliberation. If $(\mathbf{p}^*, \mathbf{q}^*)$ is an interior rest point for the replicator dynamics (2), then $(\mathbf{p}^*, \mathbf{q}^*)$ is linearly unstable.*

Proof. $(\mathbf{p}^*, \mathbf{q}^*)$ is linearly unstable if the Jacobian matrix evaluated at $(\mathbf{p}^*, \mathbf{q}^*)$ has at least one eigenvalue with positive real part. The tangent space at a point $(\mathbf{x}, \mathbf{y})$ in the interior of the state space consists of vectors (ξ, η) with $\xi, \eta \in \mathbb{R}_0^{16} = \{\zeta : \sum_{j=1}^{16} \zeta_j = 0\}$. Set $p_i = p_i^* + \xi_i$ and $q_i = q_i^* + \eta_i$ where (ξ, η) is a vector in the tangent space at $(\mathbf{p}^*, \mathbf{q}^*)$. Then

$$\begin{aligned}
\dot{\xi}_i &= \dot{x}_i = (p_i^* + \xi_i)(u(p_i, \mathbf{q}^* + \eta) - u(\mathbf{p}^* + \xi, \mathbf{q}^* + \eta)) \\
&= p_i^*(u(p_i, \eta) - u(\mathbf{p}^*, \eta)) + \xi_i(u(p_i, \eta) - u(\mathbf{p}^*, \eta)) - (p_i^* + \xi_i)u(\xi, \eta) \\
&= \sum_j L_{ij}^s \eta_j + \text{higher-order terms}
\end{aligned}$$

where $L_{ij}^s = p_i^*(a_{ij} - \mathbf{p}^* \cdot A\mathbf{e}_j)$ (a_{ij} being the ijth component of the sender's payoff matrix A) is the partial derivative of the ith equation of (2) relative to the jth variable. By a similar calculation we get

$$\dot{\eta}_i = \dot{q}_i \approx \sum_j L_{ij}^2 \xi_j$$

with $L_{ij}^2 = q_i^*(b_{ij} - \mathbf{q}^* \cdot Be_j)$, b_{ij} being the ijth component of the receiver's payoff matrix $B = A^t$. If we set

$$L = \begin{pmatrix} 0 & L^s \\ L^r & 0 \end{pmatrix}$$

then L is the Jacobian evaluated at $(\mathbf{p}^*, \mathbf{q}^*)$. L has at least one positive eigenvalue if L is not negative semi-definite on the tangent space at $(\mathbf{p}^*, \mathbf{q}^*)$, i.e. if

$$\langle (\xi, \eta), L(\xi, \eta) \rangle_{(\mathbf{p}^*, \mathbf{q}^*)} > 0$$

for some (ξ, η) in the tangent space at $(\mathbf{p}^*, \mathbf{q}^*)$ ($\langle \cdot, \cdot \rangle$ denotes the Shashshahani inner product; see Hofbauer and Sigmund 1998, 128). We claim that by setting $\mathbf{p} = \epsilon\mathbf{s} + (1 - \epsilon)\mathbf{p}^*$ and $\mathbf{q} = \delta\mathbf{r} + (1 - \delta)\mathbf{p}^*$, with $(\mathbf{s}, \mathbf{r})$ a strict Nash equilibrium of Σ_2^d, there exist such points $\xi = \mathbf{p} - \mathbf{p}^*$ and $\eta = \mathbf{q} - \mathbf{q}^*$ arbitrarily close to $(\mathbf{p}^*, \mathbf{q}^*)$.

Observe first that

$$\begin{aligned}
\langle (\xi, \eta), L(\xi, \eta) \rangle_{(\mathbf{p}^*, \mathbf{q}^*)} &= \sum_i \frac{1}{p_i^*} \xi_i \sum_j L_{ij}^1 \eta_j + \sum_k \frac{1}{q_k^*} \eta_k \sum_m L_{km}^2 \xi_m \\
&= 2(u(\mathbf{p}, \mathbf{q}) - u(\mathbf{p}^*, \mathbf{q}^*))
\end{aligned}$$

The last term is positive since $u(\mathbf{p},\mathbf{q}) = \epsilon\delta u(\mathbf{s},\mathbf{r}) + \epsilon(1-\delta)u(\mathbf{s},\mathbf{q}^*) + (1-\epsilon)\delta u(\mathbf{p}^*,\mathbf{r}) + (1-\epsilon)(1-\delta)u(\mathbf{p}^*,\mathbf{q}^*) > u(\mathbf{p}^*,\mathbf{q}^*)$. Thus L is not negative semi-definite. ■

Proof of Theorem 1. Let Σ_2^d be a d-signaling game with $\mathbb{P}(\sigma_1) = \mathbb{P}(\sigma_2)$. By Proposition 3, the replicator dynamics (2) for Σ_2^d is a gradient system and the only asymptotically stable fixed points are the strict Nash equilibria of Σ_2^d. This implies that all other rest points are either unstable or weakly stable (i.e. stable, but not asymptotically stable).

Let us consider the interior of the state space first. The existence of a potential function (the average payoff) excludes the existence of periodic orbits since the potential is strictly increasing along every non-stationary solution. This also implies that interior rest points cannot be weakly stable in that there are cycling solutions around an interior rest point (the eigenvalues of the Jacobian evaluated at rest points have zero imaginary part Proposition 3). Moreover, Lemma 1 implies that if $(\mathbf{p}^*,\mathbf{q}^*)$ is an equilibrium of (2), then there is no neighborhood U of $(\mathbf{p}^*,\mathbf{q}^*)$ such that every point in U is an equilibrium of (2).

These arguments show that interior rest points must be unstable. We claim that the set of unstable rest points together with the set of points converging to those unstable rest points has Lebesgue measure zero. In (Huttegger forthcoming) it is shown that for gradient systems the set of unstable fixed points has Lebesgue measure zero. If S is a connected, possibly singleton, set of unstable fixed points in the interior, then the center-stable manifold theorem (see Kelley 1967) implies that the set of points converging to points in S is contained in a manifold whose codimension is at least 1 if points in S have at least one positive eigenvalue. But this last fact follows from Lemma 1.

All equilibrium points on the boundary where the senders or receivers coordinate their deliberational states suboptimally with positive frequency will be unstable. To see this consider an arbitrary rest point $(\mathbf{x},\mathbf{y})$. Disregarding the deliberational states in the supports of $\mathbf{x}$ and $\mathbf{y}$ defines equivalence classes of strategies which are just characterized by a sender strategy or a receiver strategy, respectively. Define a copy $(\mathbf{x}',\mathbf{y}')$ of $(\mathbf{x},\mathbf{y})$ by requiring $\mathbf{x}'$ and $\mathbf{y}'$ to have the right deliberational states for both states of the world and by setting the frequency of each sender and receiver strategy in $\mathbf{x}'$ and $\mathbf{y}'$ equal to the frequency of the corresponding equivalence class based on strategies in the support of $(\mathbf{x},\mathbf{y})$. Perturbing $(\mathbf{x},\mathbf{y})$ in the direction of $(\mathbf{x}',\mathbf{y}')$ clearly leads away from equilibrium.

Thus the only rest points we still have to analyze are on the boundary where sender strategies and receiver strategies with suboptimal deliberational states are not present. Apart from the two signal system types, equilibria on this boundary are the same as in the two-population replicator dynamics for signaling games without deliberation. Let x_1 and x_2 denote the frequency of the two signaling systems s and s'. Let y_1 and y_2 denote

the frequency of the two signaling systems r and r', (s, r) and (s', r') being the strict Nash equilibria. The vertex $x_1 = 1$ and $y_2 = 1$ is linearly unstable as is the vertex $x_2 = 1$ and $y_1 = 1$. There is a manifold of equilibria M where $x_1 = x_2$ and $y_1 = y_2$. Rest points in M are also linearly unstable except when all one-to-one strategies have zero frequency. In this case, rest points are second order unstable: only many-to-one strategies are present, so introducing a signal system pair will increase the average payoff. There remain two types of sets of rest points. $x_1 = 0 = x_2$ defines a manifold of rest points which are linearly unstable if $y_1 \neq y_2$ and second order unstable otherwise (by the same argument involving average payoff as before). Similar arguments apply for the manifold where $y_1 = 0 = y_2$. ■

Proof of Theorem 3. In the second model there is not one pair of deliberational states which is optimal relative to a state of the world. There are two pairs for each state which are characterized by being asymmetric: if the sender deliberates, the receiver isn't supposed to deliberate, or the other way round. This means that the same arguments as in the proof of Theorem 1 go through except that convergence will depend on how many of one of the two configurations of deliberational states are intially present. ■

Acknowledgements

The paper profited most from numerous discussions with Brian Skyrms and from many helpful remarks by Bill Harms. Moreover, I am grateful to Peter Hurd, Natasha Komarova, Don Saari and Kevin Zollman for their assistance, to Josef Hofbauer for pointing out the relevance of the center-stable manifold theorem and to an anonymous referee for a number of useful suggestions.

Notes

[1] This question also puzzled 20th century positivists like Dubislav (1937).
[2] This is one way to translate the underlying extensive form signaling game into a game in strategic form.
[3] For rational choice interpretations see Lewis (1969) and Cubitt and Sugden (2003).
[4] For more on the semantic differences between human languages and animal signaling systems see Harms (2004a).
[5] From a representational point of view, signals, states and acts might be viewed as representations of each other if they are part of a signaling system. Notice that we do not need to introduce any kind of mental representations. Mental representations could mediate between states and signals or signals and acts. But our argumentations below do not rely on additional mediating representations. For more on the relationship between animal signals and mentalistic language see Radner (1999).

[6] For an analysis of signaling games which include the realistic feature of local interaction see Grim et al. (2001) and Zollman (2005).
[7] On this view, one may wonder about the status of a cry for help like 'Help!'. In many such situations the receiver is required to think about what to do and not to make a decision without deliberation. I don't think that this poses any serious problem to Lewis' proposal. Either it is clear what the appropriate response to a cry for help is. Or the act of starting to deliberate about further actions without thinking whether to start it or not is the required response.
[8] Discrete time dynamics are easier to program than dynamics in continuous time. We used a two population model in order to keep the number of types in the population reasonably low.
[9] Notice that (1) does not necessarily have to be a model for the evolutionary dynamics of two populations. It may also serve as a model for two agents that interact repeatedly and have strategies corresponding to the types for (1). The frequencies then represent the probability with which each agent chooses a certain strategy. Börgers and Sarin (1997) derive a continuous version of (1) as a model of individual learning from a variant of reinforcement learning that was studied in Bush and Mosteller (1955). Other reinforcement learning dynamics may also lead to a continuous version of (1) (Beggs 2002; Hopkins and Posch 2005).
[10] An asymmetric two-player game with payoff matrices (A, B) is a rescaled partnership game if there exist constants $\alpha, \beta > 0$ and c_j, d_i such that the game with payoff matrices (A', B') defined by $a'_{ij} = \alpha a_{ij} + c_j$ and $b'_{ji} = \beta b_{ji} + d_i$ is a partnership game. See Hofbauer and Sigmund (1998).

References

Beggs, A. W.: 2005, 'On the Convergence of Reinforcement Learning', *Journal of Economic Theory* **122**, 1–36.

Boesch, C.: 1991, 'Symbolic Communication in Wild Chimpanzees', *Human Evolution* **6**, 81–90.

Börgers, T. and R. Sarin: 1997, 'Learning Through Reinforcement and the Replicator Dynamics', *Journal of Economic Theory* **74**, 235–265.

Bush, R. and F. Mosteller: 1955, *Stochastic Models for Learning*, John Wiley & Sons, New York

Cheney, D. L. and R. M. Seyfarth: 1990, *How Monkeys See the World: Inside the Mind of Another Species*, Chicago University Press, Chicago

Crawford, V. P. and J. Sobel: 1982, 'Strategic Information Transmission', *Econometrica* **50**, 1431–1451.

Crespi, B. J.: 2001, 'The Evolution of Social Behavior in Microorganisms', *Trends in Ecology and Evolution* **16**, 178–183.

Cubitt, R. P. and R. Sugden: 2003, 'Common Knowledge, Salience and Convention: A Reconstruction of David Lewis' Game Theory', *Economics and Philosophy* **19**, 175–210.

Dubislav, W.: 1937. 'Zur Unbegründbarkeit der Forderungssätze', *Theoria* **3**, 330–342.

England, R. R., G. Hobbs, N. J. Bainton and D. Roberts: 1999, *Microbial Signaling and Communication*, Cambridge University Press, Cambridge.

Grim, P., T. Kokalis, A. Alai-Tafti, N. Kilb and Paul St. Denis: 2001, *Making Meaning Happen*, Technical Report #01-02, Department of Philosophy, Group for Logic and Formal Semantics, SUNY, Stony Brook, NY.

Harms, W. F.: 2000, 'Adaption and Moral Realism', *Biology and Philosophy* **15**, 699–712.

——— 2004a, *Information and Meaning in Evolutionary Processes*, Cambridge University Press, Cambridge.

——— 2004b, 'Primitive Content, Translation, and the Emergence of Meaning in Animal Communication', in D. K. Oller and U. Gabriel (eds.), *Evolution of Communication*, MIT Press, Cambridge, Mass., pp. 31–48.

Hauser, M. D.: 1997, *The Evolution of Communication*, MIT Press, Cambridge, Mass.

Hirsch, M. W. and S. Smale: 1974, *Differential Equations, Dynamical Systems, and Linear Algebra*, Academic Press, Orlando.

Hirsch, M. W., S. Smale and R. L. Devaney: 2004, *Differential Equations, Dynamical Systems and an Introduction to Chaos*, Academic Press, San Diego.

Hofbauer, J. and E. Hopkins: 2005, 'Learning in Perturbed Asymmetric Games', *Games and Economic Behavior* **52**, 133–152.

Hofbauer, J. and K. Sigmund: 1998, *Evolutionary Games and Population Dynamics*, Cambridge University Press, Cambridge.

Hopkins, E. and M. Posch: 2005, 'Attainability of Boundary Points under Reinforcement Learning', *Games and Economic Behavior* **53**, 110–125.

Huttegger, S. M.: forthcoming, 'Evolution and the Explanation of Meaning', *Philosophy of Science*.

Kelley, A.: 1967, 'The Stable, Center-Stable, Center, Center-Unstable, Unstable Manifolds', *Journal of Differential Equations* **3**, 546–570.

Komarova, N. and P. Niyogi: 2004, 'Optimizing the Mutual Intelligibility of Linguistic Agents in a Shared World', *Artificial Intelligence* **154**, 1–42.

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

Maynard Smith, J.: 1982, *Evolution and the Theory of Games*, Cambridge University Press, Cambridge.

Maynard Smith, J. and D. Harper: 2003, *Animal Signals*, Oxford University Press, Oxford.

Maynard Smith, J. and E. Szathmáry: 1995, *The Major Transitions in Evolution*, Oxford University Press, Oxford.

Millikan, R. G.: 1984, *Language, Thought and other Biological Categories*, MIT Press, Cambridge, Mass.

Nowak, M. A. and D. C. Krakauer: 1999, 'The Evolution of Language', *Proceedings of the National Academy of Sciences* **96**, 8028–8033.

Owings, D. H. and D. F. Henessy: 1984, 'The Importance of Variation in Sciurid Visual and Vocal Communication', in J. O.Murie and G. L. Michener (eds.), *The Biology of Ground Dwelling Squirrels*, University of Nebraska Press, Lincoln, pp. 169–200.

Radner, D.: 1999, 'Mind and Function in Animal Communication', *Erkenntnis* **51**, 129–144.

Skyrms, B.: 1996, *Evolution of the Social Contract*, Cambridge University Press, Cambridge.

——— 2000, 'Stability and Explanatory Significance of Some Simple Evolutionary Models', *Philosophy of Science* **67**, 94–113.

Snowdon, C. T.: 1990, 'Language Capacities of Nonhuman Animals', *Yearbook of Physical Anthropology* **33**, 215–243.

van Rooy, R.: 2004, 'Evolution of Conventional Meaning and Conversational Principles', *Synthese* **139**, 331–366.

Weibull, J.: 1995. *Evolutionary Game Theory*, MIT Press, Cambridge, Mass.

Zollman, K. J. S.: 2005, 'Talking to Neighbors: The Evolution of Regional Meaning', *Philosophy of Science* **72**, 69–85.