MODELING THE ORIGINS OF COMMUNICATION: ALIGNMENT OF INTERESTS AND SOCIAL CUES THEREOF

ROLAND MÜHLENBERND[*1] and ANDREAS BAUMANN[2]

[*]Corresponding Author: muehlenbernd@leibniz-zas.de
[1]Leibniz-Centre General Linguistics, Berlin, Germany
[2]University of Vienna, Vienna, Austria

Alignment of interests is a fundamental factor for the evolutionary stability of communication. We will apply tools from evolutionary game theory to analyze the relationship between the degree of alignment of interests among interlocutors and the evolutionary stability of their potential signaling strategies. We will study different scenarios and show that there is an evolutionary benefit for adopting a communication strategy which is conditioned on social cues: hints about the degree of alignment of interests in a given situation. We will show that this strategy i) is the only evolutionarily stable strategy in a scenario with social cues, ii) has very low underlying requirements (with respect to alignment of interests) to be evolutionarily stable, and iii) is less reliable for the receiver, since it determines the sender to be partially dishonest.

1. Introduction

A fundamental underlying factor for the stability of communication is *alignment of interests* (AOI). Searcy and Nowicki (2005) illustrate in a number of scenarios that the evolutionary stability of animal signaling depends on the AOI between sender and receiver. They distinguish a) scenarios where interests overlap, such as food begging signals or alarm calls, where stable signaling is suitable and common, from b) scenarios where interests diverge or oppose, such as mating signals or displays of aggression, where stable signaling is rather unexpected and can only be explained through additional conditions. Naturally, AOI does also play an important role in human language. For example, the Gricean *cooperation principle* tells us that interlocutors' interests are generally aligned in at least one aspect, namely to communicate appropriately and cooperatively (Grice, 1975).

Often, the purpose of a communicative signal depends on the extent to which interests are aligned. For example, birds such as the great titmouse use alarm calls in cases when there is actual danger (high AOI), but they also use them as 'false alarm' in situations where they want to move other birds away from some food source (low AOI) (cf. Matsuoka, 1980). That is, titmice are able to adapt their signaling strategies not only towards the actual information they want to transfer (the source of danger) but also towards properties of the receiver, i.e., social information. This form of deception is frequently found in human language as well. For example, the usage of polite speech can serve for sweetening the conversation (high AOI) or for persuading the hearer for doing something that is rather in the interest of the speaker (low AOI).

In this article, we will apply evolutionary game theory to show how the stability of communication strategies under population dynamics depends on the AOI in the underlying scenario. Moreover, we will demonstrate that the establishment of stable communication is more likely when senders adapt their signaling strategies towards *social cues*, i.e., hints about the AOI between interlocutors in a given situation. Finally, we will show that the benefit of the integration of social cues into communication comes with a hitch: it promotes dishonest signaling.

2. The mating game: communication with (partially) aligned interests

We study a fundamental question concerning the origin of communication: how can a signal become an entity of (meaningful) communication at all? We take a look at a simple form of communication, namely *signaling*, where a sender produces a signal that evokes a response from a receiver. One important aspect of signaling being sustainably successful is the alignment of interests between a sender and a receiver. When the sender has an advantage in evoking a particular response from a receiver by sending a signal, and the receiver, in turn, has an advantage in performing this response upon the signal, then both parties have an aligned interest in successful communication.

To emphasize the role of alignment of interests in communication, we want to draw the attention to a signaling scenario where the interests between sender and receiver are only partially aligned. Diverse game-theoretic versions of this scenario have been studied with tools from evolutionary game theory (cf. Crawford & Sobel, 1982; Grafen, 1990).We will present a simple model version of this scenario, which we call the *mating game*.

The scenario is as follows: we have a situation where a male and a female of a particular species get together to get involved in a potential mating act. The male can be a high quality (H) or low quality (L) type. The female can agree to mate (M) or not to mate ($\bar{M}$). For any type of male, high or low type, it is always evolutionarily beneficial to mate with the female. For the female, however, it is evolutionarily beneficial to mate with high type males and reject low type males. The simplest form of the underlying payoff structure of this scenario is given in Figure 1(a), where each strategy profile is attributed to two utility values, the first for the male (row player), the second for the female (column player). Utility values are either 1 (evolutionarily beneficial) or 0 (evolutionarily detrimental).

In this scenario, interests are only aligned for the case where the male is a high type. It can be shown that sustainable signaling cannot evolve. Let's assume that high type males start sending a signal to the females, upon which females mate, otherwise they don't mate. That works fine for the females as long as only high type males send that signal. But now low type males have an evolutionary benefit

	M	$\bar{M}$
H	$1, 1$	$0, 0$
L	$1, 0$	$0, 1$

(a) underlying payoff structure

	$\langle M, \bar{M}\rangle$	$\langle \bar{M}, \bar{M}\rangle$
$\langle s, s\rangle$	$1, 1/2$	$0, 1/2$
$\langle s, \bar{s}\rangle$	$1/2, 1$	$0, 1/2$
$\langle \bar{s}, s\rangle$	$1/2, 0$	$0, 1/2$
$\langle \bar{s}, \bar{s}\rangle$	$0, 1/2$	$0, 1/2$

(b) mating game

Figure 1. The mating game's (a) underlying payoff and (b) normal-form representation.

in sending the signal, and will therefore adopt it. But then the signal is meaningless for females, because it lost its function, namely to distinguish high type from low type males. Therefore, they can ignore it and signaling stops working.

This line of thought can be studied mathematically. Whether a signaling strategy is sustainable can be mathematically deduced by proofing if it is an evolutionarily stable strategy (Maynard Smith & Price, 1973). As a first step, we define a space of signaling strategies. We define a sender strategy as a tuple $\langle x_H, x_L\rangle$, whereby x_H represents the signaling behavior of high types, and x_L represents the signaling behavior of low types. In its simplest version, our scenario entails four different sender strategies: $\langle s, s\rangle$, $\langle s, \bar{s}\rangle$, $\langle \bar{s}, s\rangle$ and $\langle \bar{s}, \bar{s}\rangle$, whereby s stands for sending a signal, $\bar{s}$ stands for not sending a signal. Similarly, a receiver strategy is defined as a tuple $\langle y_s, y_{\bar{s}}\rangle$, whereby y_s represents the response behavior upon receiving a signal, and $y_{\bar{s}}$ the response behavior upon not receiving a signals. For simplicity, we assume that receivers never mate upon not receiving a signal. Then the receiver strategies are: $\langle M, \bar{M}\rangle$ and $\langle \bar{M}, \bar{M}\rangle$.

From the underlying payoff structure we can deduce a utility table over pairs of signaling strategies by computing *expected utilities* (see the Online Appendix Sec. B) which represent how beneficial (on average) a combination of a sender strategy and a receiver strategy (a *strategy profile*) is for the sender (first value) and the receiver (second value). The resulting utility table is given in Figure 1(b).

From these utilities we can identify strategy profiles that are evolutionarily stable (Maynard Smith & Price, 1973). Selten (1980) has shown for asymmetric games, such as the mating game, that a strategy profile is evolutionarily stable if and only if the row player's utility is unique maximum in its column, and the column player's utility is a unique maximum in its row. This means that any unilateral switch out of an evolutionarily stable strategy profile leads to a strictly lower payoff for the switching player. The readers can check for themselves that none of the strategy profiles of Figure 1(b) is evolutionarily stable. Particularly, the strategy profile that represents separating signaling ($\langle s, \bar{s}\rangle$, $\langle M, \bar{M}\rangle$) is not evolutionarily stable, since the sender has a benefit from switching to $\langle s, s\rangle$. And then, ($\langle s, s\rangle$, $\langle M, \bar{M}\rangle$) is not evolutionarily stable, since the receiver is not strictly worse off when switching to $\langle \bar{M}, \bar{M}\rangle$. A similar argument can be found for any other strategy profile of the mating game in Figure 1(b).

	M	$\bar{M}$
H	$1, 1$	$0, 0$
L	$0, 0$	$1, 1$

(a) underlying payoff structure

	$\langle M, \bar{M}\rangle$	$\langle \bar{M}, \bar{M}\rangle$
$\langle s, s\rangle$	$1/2, 1/2$	$1/2, 1/2$
$\langle s, \bar{s}\rangle$	$1, 1$	$1/2, 1/2$
$\langle \bar{s}, s\rangle$	$0, 0$	$1/2, 1/2$
$\langle \bar{s}, \bar{s}\rangle$	$1/2, 1/2$	$1/2, 1/2$

(b) lethal mating game

Figure 2. Lethal mating game's (a) underlying payoff and (b) normal-form representation.

Let us now consider a scenario where interests are completely aligned, as represented by the payoff structure in Figure 2(a). This could represent a mating scenario where the female is extremely selective and aggressive. For example, after agreeing upon mating, the female tests the strength of the male before the actual mating act happens. High-quality males pass the test and mate with the female, whereas low-quality males fail the test and get killed by the female before the mating can happen. Here, low types have an evolutionary benefit of not initiating a mating act with this type of female. Let us call this game the lethal mating game. Its normal-form game representation is given in Figure 2(b).

This game has exactly one evolutionarily stable strategy profile, namely $(\langle s, \bar{s}\rangle, \langle M, \bar{M}\rangle)$, since any unilateral switch of strategy leads to a strictly worse payoff for the switching player. In this strategy profile, high-quality males send a signal upon which females mate, and low-quality males don't send a signal and won't get killed by the female. The bottom line of this section is this: in contrary to the standard mating game, signaling in the lethal mating game is evolutionarily stable, particularly because the interests of sender and receiver are totally aligned.

3. Transfer to a scenario of cultural evolution

Note that these mating game scenarios study stability aspects of signaling strategies in the light of biological evolution, where utility represents fitness which drives the rate of biological reproduction. But when it comes to the evolution of human communications and language, we are frequently concerned with cultural evolution, where utility equals fitness which drives the rate of behavioral reproduction, for example in form of imitation or learning. Since the concept of evolutionary stability is assumed to be applicable for biological evolution as well as cultural evolution (O'Connor, 2020), we can be agnostic about the concrete underlying dynamics. Moreover, very often we can transfer games that describe scenarios that a relevant in the light of biological evolution to scenarios that are relevant in the light of cultural evolution. We will exemplify this hereafter.

Here is an idea how we can reinterpret the mating game as a communication scenario primarily exposed to dynamics of cultural evolution: Assume that we have a group of individuals which happen to use alarm calls for dangerous events, such as raptor attacks, etc. Let's say that the individuals can distinguish between events of high danger (H) and events of low danger (L). In our game, the row player is a watchpost, who is in charge of giving alarm calls. The column player is a gatherer, who is doing a utility-relevant activity, such as picking berries.

In a first scenario the gatherer is selfish: he is picking the berries for himself. He has an interest to be warned whenever there is a high danger event, so he can stop picking and look for a save place to move to (M). However, in a low danger event, it is more beneficial not to move ($\bar{M}$) and continue picking berries. The watchpost, however, has an interest in always interrupting the gatherer, since there will be more berries left for her when her watchpost shift is over. The underlying payoff structure of this scenario is exactly the one in Figure 1(a), and the resulting communication game is exactly the one in Figure 1(b). And the analysis leads to the same conclusion: the strategy profile of separating signaling ($\langle s, \bar{s}\rangle$, $\langle M, \bar{M}\rangle$) is not evolutionarily stable. And also the reason is the same: interests are not aligned. Picking a lot of berries is good for the gatherer, but bad for the watchpost.

Let us think about a similar scenario where the gatherer is not picking berries for himself, but for the whole group instead, including the watchpost. Then the watchpost would prefer only to interrupt the gatherer when the gatherer's life is at stake (in high danger situations), but otherwise prefers him to continue picking berries, since more work done by the gatherer means more food for the whole group, including the watchpost herself. This scenario is exactly the one presented in Figure 2(a). Accordingly, Figure 2(b) represents the resulting communication game. Here again, the conclusion is the same: the separating signaling strategy ($\langle s, \bar{s}\rangle$, $\langle M, \bar{M}\rangle$) is evolutionarily stable since interests are aligned. Picking a lot of berries is good for both the gatherer and for the watchpost.

4. Combined communication games and social cues

Note that both scenarios of the last section differ with respect to the social behavior of the gatherer. In the first case, he is selfish; in the second case, he is social. In this section we want to merge both scenarios to a combined one, where the gatherer is sometimes selfish and sometimes social. In more general terms, in such a *combined scenario* the situation of totally aligned interests (Figure 2(a)) occurs with a frequency p in $(0, 1)$, and the situation of partially aligned interests (Figure 1(a)) occurs otherwise, with frequency $1 - p$. When we combine the utility tables of both situations weighted by probability p, we obtain the underlying payoff structure as given in Figure 3(a). Its normal form game representation with signaling options is given in Figure 3(b), which we label *communication game 1*.

Here, it can be shown that the separating signaling strategy ($\langle s, \bar{s}\rangle$, $\langle M, \bar{M}\rangle$) is evolutionarily stable if and only if $p > 1/2$; for the proof see the Online Appendix Sec. A.1. In other words, when situations of totally aligned interests occur more frequently than situations of partially aligned interests, then the separating signaling strategy is evolutionarily stable. With respect to our concrete example, this is the case when gatherers behave more often social than selfish.

	M	$\bar{M}$
H	$1, 1$	$0, 0$
L	$1-p, 0$	$p, 1$

(a) underlying payoff structure

	$\langle M, \bar{M} \rangle$	$\langle \bar{M}, \bar{M} \rangle$
$\langle s, s \rangle$	$1 - p/2,\ 1/2$	$p/2,\ 1/2$
$\langle s, \bar{s} \rangle$	$1/2 + p/2,\ 1$	$p/2,\ 1/2$
$\langle \bar{s}, s \rangle$	$1/2 - p/2,\ 0$	$p/2,\ 1/2$
$\langle \bar{s}, \bar{s} \rangle$	$p/2,\ 1/2$	$p/2,\ 1/2$

(b) communication game 1

Figure 3. The (a) underlying payoff and (b) normal-form representation of communication game 1. The alignment of interests between row and column player depends on p.

	$\langle M, \bar{M} \rangle$	$\langle \bar{M}, \bar{M} \rangle$
$\langle s, s, s, s \rangle$	$1 - \frac{p}{2}, \frac{1}{2}$	$\frac{p}{2}, \frac{1}{2}$
$\langle s, s, \bar{s}, s \rangle$	$\boxed{1, \frac{1+p}{2}}$	$\frac{p}{2}, \frac{1}{2}$
$\langle s, s, \bar{s}, \bar{s} \rangle$	$\frac{1+p}{2}, 1$	$\frac{p}{2}, \frac{1}{2}$
$\langle \bar{s}, \bar{s}, \bar{s}, \bar{s} \rangle$	$\frac{p}{2}, \frac{1}{2}$	$\frac{p}{2}, \frac{1}{2}$

Figure 4. A subset of four selected strategies in communication game 2. The full game entails the same finding: there is only one evolutionarily stable strategy pair (and only for $p > 0$), namely $(\langle s, s, \bar{s}, s \rangle, \langle M, \bar{M} \rangle)$ (the boxed entry).

The definition of communication game 1 implies that individuals cannot distinguish situations with totally aligned interests from those with partially aligned interests. Let's assume that the sender is able to distinguish both scenarios, for example via *social cues* that give hints about the receiver's attitude. For example, when the sender can identify if the gatherer is behaving selfish or social, she can condition her strategy accordingly. We can model this by extending the sender's strategy space to a quadruple $\langle x_{Ht}, x_{Hp}, x_{Lt}, x_{Lp}, \rangle$, where x_{Ht} is the signaling behavior of a sender in a scenario with totally aligned interests in a high-danger event, x_{Hp} is the signaling behavior of a sender in a scenario with partially aligned interests in a high-danger event, and so on. In each case, the sender can either send a signal s or not $\bar{s}$, leading to $2^4 = 16$ sender strategies. We call the resulting game *communication game 2*. Figure 4 shows a subset of strategies in communication game 2. The full utility table is given in the Online Appendix Sec. A.2.

It can be shown that there is exactly one evolutionarily stable strategy pair, namely $(\langle s, s, \bar{s}, s \rangle, \langle M, \bar{M} \rangle)$, where the sender always sends the signal in high-danger events, but also in low-danger events in case that the receiver is selfish. Importantly, this strategy pair is evolutionarily stable for any p in $(0, 1)$; for the proof see the Online Appendix Sec. A.2. The bottom line of this section is this: there is a much higher AOI requirement for evolutionary stability *without a sender's access to social cues* (communication game 1; requirement: $p > \frac{1}{2}$) than *with a sender's access to social cues* (communication game 2; any p in $(0, 1)$). This is our main finding and we will come back to this point in the conclusion.

Note that the scenarios in this section assume that signaling comes without costs. In the Online Appendix Sec. C, we studied communication games 1 and 2 with signaling costs. The result shows that signaling costs strengthen the conditions for separating signaling strategies to be evolutionarily stable in both games. Since the separating signaling strategy in communication game 2 is not conditioned on social cues, we can conclude that signaling costs weaken the prerequisites for the evolutionary stability of signaling strategies that involve social cues.

5. Reliability and honesty

We want to study the communication strategies with respect to the concepts *reliability* and *honesty*. Following Searcy and Nowicki (2005), a signal is reliable if i) some characteristic of the signal (including its presence/absence) is consistently correlated with some attribute of the signaler or its environment, and ii) receivers benefit from having information about this attribute. Moreover, a signal does not have to be perfectly reliable, but it is enough when the signal is *honest on average*, such that the receiver on average is better off assessing the signal than ignoring it.

Following these lines of thought, we define a signal s to be honest, if and only if it is sent in situations such that the expected response of the receiver is evolutionarily beneficial to the receiver. Hereof, we define the reliability of a signal s as the ratio of the honest usage of s to the total usage of s. More precisely, when a sender uses a signal s with frequency f_T in total and uses it honestly with frequency f_H, then the reliability of the signal is given by the ratio $\frac{f_H}{f_T}$. Thus, reliability of a signal is a value between 0 and 1 (never/always honestly used).

Let's take a look at the reliability of signal s in sender strategy $\langle s, s, \bar{s}, s\rangle$ as part of the only stable signaling strategy pair $(\langle s, s, \bar{s}, s\rangle, \langle M, \bar{M}\rangle)$ in communication game 2. Here, totally aligned situations occur with frequency p and partially aligned situations with frequency $1 - p$. High danger and low danger events are implicitly assumed to occur both with frequency $\frac{1}{2}$. This gives us the following frequencies for the four situations that the sender can distinguish: $f(x_{Ht}) = \frac{p}{2}$, $f(x_{Lt}) = \frac{p}{2}$, $f(x_{Hp}) = \frac{1-p}{2}$ and $f(x_{Lp}) = \frac{1-p}{2}$. In the stable signaling strategy $\langle s, s, \bar{s}, s\rangle$ the sender sends the signal with total frequency $f_T = f(x_{Ht}) + f(x_{Hp}) + f(x_{Lp}) = \frac{2-p}{2}$. Furthermore, since the signal s is only honestly sent when there is a high danger event, we get frequency $f_H = f(x_{Ht}) + f(x_{Hp}) = \frac{1}{2}$. Now we can compute the reliability of signal s as follows: $\frac{f_H}{f_T} = \frac{\frac{1}{2}}{\frac{2-p}{2}} = \frac{1}{2-p}$.

From this analysis we can follow that the reliability of the signal s in strategy $\langle s, s, \bar{s}, s\rangle$ is greater than $\frac{1}{2}$ for any p in $(0, 1)$. However, this result also shows that strategy $\langle s, s, \bar{s}, s\rangle$ is not totally reliable (save for the borderline case with $p = 1$). On the other hand, it is easy to see that in the evolutionarily stable sender strategy $\langle s, \bar{s}\rangle$ of communication game 1, signal s has a reliability value of 1 (since s is only and always sent in H), therefore it is totally reliable.

6. Conclusion

We have argued that classical game-theoretical models of animal signaling can be transferred to scenarios that might have played an important role in the evolution of human language. The evolutionary analyses of the models in this paper have demonstrated a well-known general aspect of any form of communication: the evolutionary stability of meaningful signaling/communication depends to a great degree on the alignment of interests (AOI) between sender and receiver. The more particular results of our analyses are connected to the juxtaposition of the communication games 1 and 2. This analysis showed that access to social cues lowers the underlying AOI requirements for signaling strategies being evolutionarily stable.

We conclude that the ability to integrate social cues into communication is a conducive feature for stable communication and therefore a potentially important propulsive factor in the evolution of language. The latter analysis showed that the integration of social cues comes with a hitch, though: it promotes dishonest signaling. But then, dishonest signaling isn't necessarily something adverse in the presented scenario. It can also be seen as a form of punishment that the sender exerts on the receiver for selfish behavior. Here, dishonesty might have evolved for a good cause, namely to suppress the behavior of nonprosocial individuals in order to enhance sociality. This aspect ought to be pursued in subsequent research.

Online Appendix: The Online Appendix is part of the proceeding's supplementary material and accessible under https://doi.org/10.6084/m9.figshare.20004395

Funding: Roland Mühlenbernd was funded by the German Research Foundation (DFG) within the SFB 1412 under DFG agreement number 41659133.

References

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, *50*, 1431–1451.

Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*, *144*, 517–546.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Edited by P. Cole/J. Morgan, New York: Academic Press.

Matsuoka, S. (1980). Pseudo warning call in titmice. *Tori*, *29*, 87-90.

Maynard Smith, J., & Price, G. (1973). The logic of animal conflict. *Nature*(146), 15–18.

O'Connor, C. (2020). *Games in the philosophy of biology.* Cambridge University Press.

Searcy, W. A., & Nowicki, S. (2005). *The evolution of animal communication.* Princeton University Press.

Selten, R. (1980). A note on evolutionarily stable strategies in asymmetric animal conflicts. *Journal of Theoretical Biology*, *84*, 93–101.