

The peacock, the sparrow, and the evolution of human language

Carl T. Bergstrom ¹ Rustom Antia ²
Szabolcs Számadó ³ Michael Lachmann ⁴

May 1, 2001

¹Department of Zoology, University of Washington, Seattle, WA, USA

²Department of Biology, Emory University, Atlanta, GA, USA

³Department of Plant Taxonomy and Ecology, Eötvös Lorand University, Budapest, Hungary

⁴Santa Fe Institute, Santa Fe, NM, USA

Abstract

How did human language arise, and what accounts for its present structure? Over the past decade, there has been great interest — and impressive progress — in using evolutionary theory to address to these and related questions. In particular, a number of studies have shown that several key features of language could plausibly arise and be maintained by natural selection when individuals have coincident interests. However, these models have largely ignored a vital strategic component of the social context in which language is employed: individuals in real societies, both past and present, do not have fully coincident interests. Simultaneously, theoretical models of animal communication have confronted these strategic issues head-on, but they have largely focused on costly (Zahavian) signals.

We approach this problem directly, asking the following question: *Can language evolve and be maintained under common biological scenarios of non-coincident interest?* Using a trio of examples — the peacock, the sparrow, and human language — we show that an explicit connection can be drawn between “costly signalling theory” as developed for the study of animal communication, and the “cheap” signals employed in human language. We argue that coincident interests are not a prerequisite for linguistic communication, and explore the structural features to be expected of languages employed in societies with non-coincident interests. We find that many of the results derived previously by assuming coincident interests will also be expected under less restrictive models of society.

1 Introduction

How did human language evolve? What accounts for its present structure? Recently, both within the field of linguistics and more broadly, there has been great interest and impressive progress in formulating these questions in terms of evolutionary games [1, 2, 3, 4, 5, 6, 7, 8, 9]. These studies have shown that several important properties of language can plausibly arise and be maintained by natural selection when individuals have coincident interests [10]. However, it is difficult to envision societies with this degree of internal harmony, at least outside of the Garden of Eden [11]. Individuals in real societies, both past and present, do not have fully coincident interests. We approach this problem directly, asking the following question: *Can language evolve and be maintained under common biological scenarios of non-coincident interest?*

In this paper, we show that an explicit connection can be drawn between “costly signalling theory” as developed for the study of animal communication [12, 13, 14, 15], and the “cheap” signals employed in linguistic communication. This approach allows us to demonstrate that coincident interests are not a prerequisite for language. We can also explore the structural features that are to be expected of the languages used in societies with non-coincident interests. Many of the results derived previously by assuming coincident interests also arise under less restrictive models of society.

We begin in Section 2 by briefly outlining recent results from signalling theory, which we will bring to bear on the problem of the evolution of human language. In Section 3, we illustrate these principles by describing two examples of animal communication — peacock tails and sparrow throat-bibs —

which highlight important aspects of the relationship between the meanings, costs, and verifiability of signals. Finally, in Section 4, we return to human linguistic communication. We explain how cost-free communication can be evolutionarily stable in a world of conflicting interests, and why the properties associated with linguistic communication would plausibly be favored under these circumstances.

2 Background and recent results from signalling theory

Signals and signalling games

Loosely speaking, a signal can be defined as a behavior or phenotype produced by one individual (the signaller) that serves to influence the behavior of a second individual (the signal receiver) by transmitting information¹. As discussed by Maynard Smith and Harper [18] — who consider in detail the

¹A more precise definition might allow the signal to change the “information state” of the receiver; whether this change in information state lead to changes in behavior can depend on the condition of the signal receiver and other aspects of the signal’s context [16, 17]. Because we focus on the strategic issues using a game-theoretic framework based in expected payoffs, the simpler definition, with behavior replacing information state, will be sufficient for the purposes of this paper. If the signal changes the information state in any way that is relevant to the strategic behavior of the signal receiver, the change in information state can be seen as changing the “conditional strategy” or “conditional behavior” of the signal receiver. This, in turn, changes the expected payoff to the receiver; so far as the strategic model is concerned, the change in information indeed causes the signaller to “behave” differently.

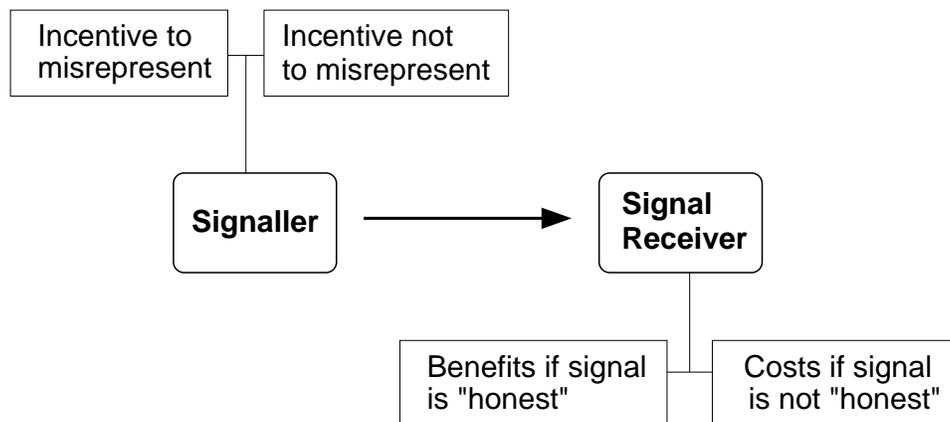


Figure 1: Schematic diagram of a signalling interaction.

terms and definitions associated with signalling — this definition highlights two important aspects of signalling. First, signals carry information from a signaller to a receiver. Second, signals influence receiver behavior. Thus communication can be viewed as a fundamentally game-theoretic process: in communication, two individuals choose actions (signal and response, respectively), the payoffs of which depend on the choices of both individuals.

Figure 1 provides a schematic diagram of communication viewed as a signalling game. The signaller must select a signal to inform the signal receiver about the state of the world. The signaller weighs the benefits of accurately representing the world against the benefits of misrepresenting the world, and based on this comparison chooses a signal. The receiver must select a response to the state of the world, and in doing so must choose whether or not to use the information provided by the signaller. The receiver thus weighs the benefits of heeding the signal should it be “reliable” against the costs of heeding it should it be misleading.

Communication will be stable, at least in the short-term, when the signaller and receiver pursue strategies (choice of signal given world state, and choice of response given signal, respectively) that together comprise a signalling equilibrium. Given a set of possible signals, a signalling equilibrium is simply a pair of signaller and receiver strategies such that neither party gains from unilateral change in strategy [19].

By the definition of a signal given above, we know that receiver behavior will be affected by the signaller's provision of information. At a signalling equilibrium, then, the receiver must on average gain from heeding the signal ([16, 18]), and in general it can be expected that the signal will be reliable on average (but see ref. [20]). This, in turn, tells us that the signaller must be in some sense reliably representing the state of the world. We can infer that on average, the incentive to the signaller to misrepresent the state of the world must be outweighed by the incentive not to do so.

But why would that happen? There are a number of possibilities. When interests of signaller and signal receiver coincide, the signaller has no incentive to misrepresent the state of the world to the receiver, and indeed may have an active incentive not to do so. Under coincident interests, it is easy to see why signallers would be inclined to accurately represent the world in their signals. But when interests conflict, the signaller will have an incentive to misrepresent the world in the choice of signals. Clearly, for the signals to nonetheless be reliable on average, this incentive must be balanced by a counterveiling incentive against misrepresentation. Much of so-called costly signalling theory focuses on understanding the source and form of this counterveiling incentive.

Incentives against misleading signals: Where do they come from?

From what sources do incentives against misleading signals arise? One possibility is that they can come from the actual costs of signal production. This is, in essence, the critical idea behind Zahavi's original formulation of the handicap principle [12, 21]. As Zahavi explains, when signal production is inherently costly — and differentially costly according to the meaning and/or validity of the message — signallers can be induced to reliably report aspects of their condition or of the environment. Indeed, numerous models have established that this idea is mathematically tractable [22, 13, 15, 14, 23].

However, the countervailing costs necessary to ensure reliable signalling need not arise from the actual production of the signal. Other models have shown that reliability can also be induced by the costs of misleading the receiver when interests are totally [24] or even partially [19] shared.

Yet another source of cost can be social in origin; a number of models have shown that the costs to misleading receivers can be prohibitive if signallers risk being “found out” or otherwise penalized for doing so [25, 26, 27, 28, 29]. In cases such as these, the incentives against misleading signals are imposed on the signallers by the receivers; these incentives can be viewed as costs incurred in the course of the interaction between signallers and receivers, as a consequence of the receivers' responses [18, 30, 17].

Incentives against misleading signals: What form do they need to take?

Clearly, in order to prevent deceptive signalling, the “threat” of cost must be present. But this does not mean that signal cost — in any of the forms described above — will necessarily be paid at equilibrium, even when signalling occurs between individuals with conflicting interests. Contrary to a principle claim of costly signalling theory [12, 21, 13, 15, 31, 32, 33], wastefulness is *not* a prerequisite for signal reliability. Recent theoretical developments for discrete [34] and continuous [35] signalling games have revealed that reliable signalling can occur even when the signals employed at equilibrium are virtually cost-free, so long as out-of-equilibrium signals are costly.

To understand these results, it helps to think in terms of a “signal cost function”: a mapping from signal and world state to the fitness “cost” paid by the signaller. While strictly speaking, a signalling game merely defines a mapping from the trio of signal, world-state, receiver strategy to the signaller payoff, we can often intuitively select an obvious separation of payoff into cost components and benefit component. For a peacock, we might assume that the effects of tail production on viability are the costs, whereas the mating success is the benefit. With sparrows, we might decide that fights incurred are the costs of a signal, and the adversaries scared away are the benefits. And so on². Looking at signalling games in this way, it is not the magnitude

²When costs are social, we somehow have to treat cost-inducing behaviors of the receivers as fixed evolutionarily just as ecological cost of producing a tail is assumed to be fixed for a peacock. That is, we have to assume that whatever the receiver does to induce cost is already optimal and not subject to adjustment within the scope of the signal game

of the cost function but rather its slope which enforces signal reliability [35].

We can illustrate these points by examining a toy model (Figure 2) in which signallers communicate their quality to signal receivers; all else held equal, signallers benefit from being seen as higher quality whereas receivers benefit from accurately assessing signallers. The “signalling technology” employed has the following cost structure: it is cost-free to signal up to true quality, but lethal to signal above it. At equilibrium, in this signalling game, all signallers will signal their true quality, with zero cost.

In practice, it may be hard to find a cost function like this, especially one that is not subject to evolutionary breakdown as signallers evolve ways to send equivalent signals more cheaply. Signal cost functions may assume a form more like that in Figure 3b, with costs gradually increasing as signallers overstate their quality. In such situations, signallers will have to pay non-zero costs.

Basically, what is happening here is that as cost functions become “smoother,” signal costs at equilibrium increase. Indeed, the curvature of the cost function imposes a lower bound on signal cost at equilibrium, with greater curvatures allowing less expensive equilibrium signals (details provided in ref [35]). Hence the ability of signalling systems to “find” cost functions with high curvature will be very important in determining signal cost at equilibrium. Different sources of signal cost (physiological, social, etc.) may vary in the degree to which the relation between signal and cost is flexible and, as a consequence, they may also vary in the level of equilibrium cost. We explore this possibility (among other things) in the sections that follow.

under consideration.

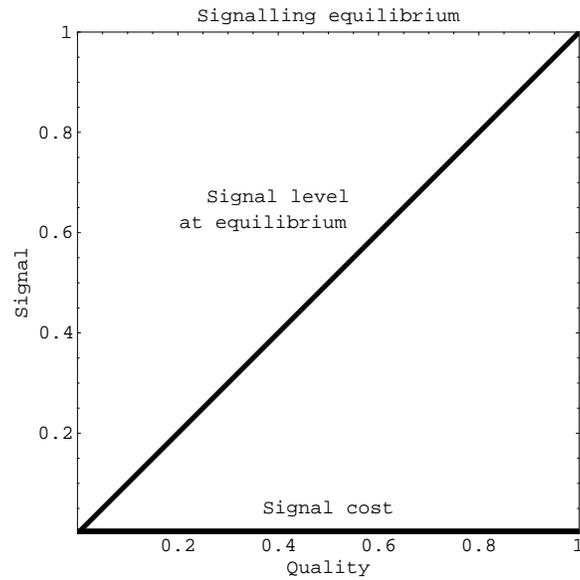
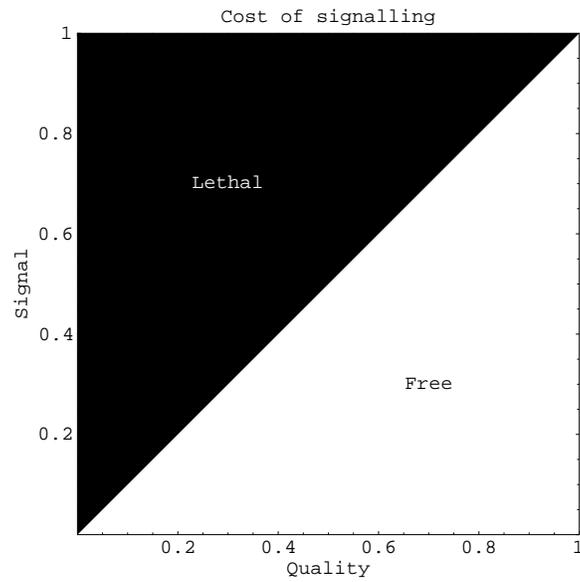


Figure 2: A toy model: signallers can send any signal up to their true quality for free; signalling in excess of their true quality is lethal. (A) the cost function, (B) the equilibrium signal level and signal cost.

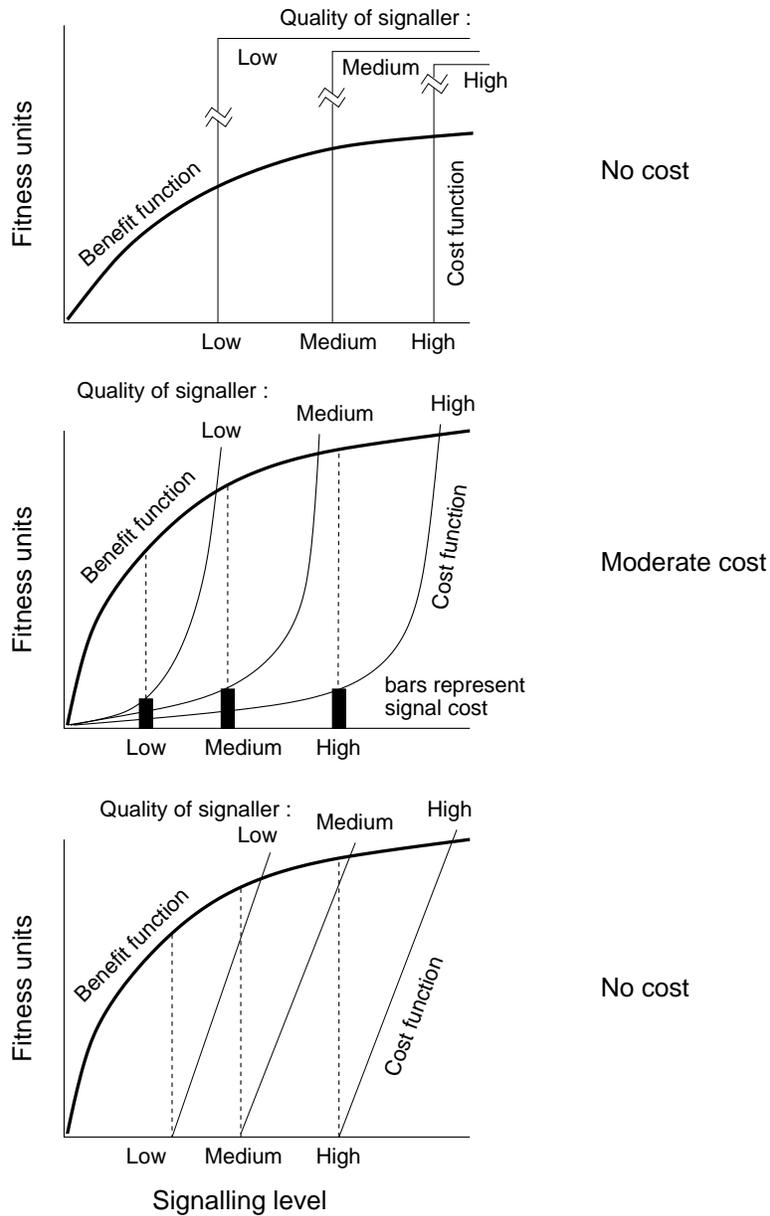


Figure 3: A) Signal cost is given by a discontinuous step function as in the toy model of Figure 2. Equilibrium signal costs are zero. B) Signal cost is a smooth function increasing with signal magnitude. Equilibrium signal costs (shaded bars) are non-zero. C) Signal cost is a continuous but kinked function. Equilibrium signal cost is zero, provided that the slope to the right of each “kink” exceeds the slope of the benefit function at that point.

3 The peacock and the sparrow

In this section, we introduce a pair of examples based on peacock and sparrow signalling behavior. With these examples we do not attempt to provide realistic models of peacock and sparrow signals, nor to reconstruct a hypothetical chain of steps by which language evolved; instead, we focus on these idealized systems because of their utility as models. As models, they can offer valuable insights into the relations among cost, conflict, and communication.

The peacock

Ever since Zahavi introduced the handicap principle [12], the elaborate tail of the peacock has been invoked as one of the most striking examples of costly signalling. Basically, the story is usually told as follows: the male peafowl expends a great deal of energy and effort to produce, maintain, and live with an elaborate tail; female peafowl preferentially mate with males who have spectacular tails, because a large tail is a reliable indicator of high genetic quality. While Zahavi's original formulation left some ambiguity in the exact mechanism by which this flamboyant appendage serves to facilitate female choice, later discussion settled upon the "strategic choice" mechanism: Peacocks weigh the marginal viability costs of increasing tail size against the marginal gains in mating success from doing so in order to select an optimal tail length; higher quality peacocks end up choosing to produce longer tails and thus quality is honestly conveyed [13, 33].

Peacocks pay signal costs which are physiological consequences of the extreme effort involved in signal production (and, in the peacock's case, from

the subsequent hassle of living with the prodigious plumes). When costs derive from such a source, signal cost becomes intertwined with the broader physiology of the organisms, and there will almost invariably be constraints on shape of the “signal cost function” described in the previous section. Costs are likely to relate to quality and signal level in complex ways which will be more likely to take forms qualitatively similar to the curves with gradually increasing slopes shown in Figure 3b. As a consequence, when costs result primarily from the physiology of signal production and maintenance, there will almost inevitably be costs associated with signalling at equilibrium.

The sparrow

If the peacocks’ tails provide the paradigmatic example of a costly signalling, then the dominance badges of passerines — for example, the black “bibs” of male house sparrows (*Passer domesticus*) or Harris sparrows (*Zonotrichia querula*) — arguably serve as the classic (if somewhat contentious) example of “conventional” signalling [36, 37, 27]. In these signalling systems, relatively minor and inconspicuous variations in plumage, such as variable throat bibs or forehead patches, serve as reliable indicators of aggressiveness, fighting ability, or other correlates of resource holding potential. Birds with “dominant” badges are more likely to win agonistic interactions, while “cheats” — birds which exaggerate their own resource-holding potential (RHP) in their choice of badge — are attacked and punished by conspecifics [36, 38, 39, 25, 40]. The actual production costs of the signal are very low; the cost on the producer’s side is only that needed to convey the message — the efficacy cost [27] — and nothing more. The remaining costs, those necessary

to keep signalling honest, arise not from signal production but instead as a consequence of the receivers' actions³.

Comparing the two examples

But why do peacocks use costly signals, while sparrows use free ones? The messages being sent appear to be closely related; both the peacock and the sparrow are communicating some aspect of quality, and in both cases the interests of signaller and receiver are not entirely in accord. The answer to this question lies not in the message itself, but rather in the ability of the receivers to verify the integrity of the message.

In our peacock example, signal receivers have no way of verifying the signals. The quality advertised by the signal — some form of genetic quality — only manifests itself stochastically and in the far future, in terms of viability and fecundity of the offspring; this simply cannot be verified by the signal receiver within the time-frame of her decision. And so the mechanism of signal reliability must involve production costs for signals (or at least for misleading signals). Indeed, when interests conflict and the trait being signalled cannot be verified by the signal receiver, there is often no other way to influence

³This distinction is not new; Maynard Smith and Harper [18] contrast peacock-like signal costs that are independent of the interaction with the signal receiver with sparrow-like signal costs that are incurred in the course of the interaction. Guilford and Dawkins [30] contrast signals stabilized by costs which are paid independently of the target receiver's response with signals stabilized by costs which arise from the target receiver's response. Hasson [17] separates the payoff of signalling into basal and social components of fitness, the peacock's tail imposing costs of the former sort and the sparrow's patch imposing costs of the latter.

the receiver *except* through a signal which has physiological costs which are associated with the quality of the signaller. Costs to deception cannot be imposed through the social interaction, because the signal receiver has no way of ever learning whether the signal was an accurate reflection of the signaller's quality, and thus the receiver has no way of "punishing" deception. And without costs, as we have already seen, reliable communication will not be stable. Truly costly signals such as the peacock's tail may not be efficient as cost-free signals, but peafowl have no other choice.

In the sparrow example, by contrast, the signal serves primarily as a shortcut to facilitate assessment of resource holding potential. Receivers can verify the signal with relative ease, should they choose to do so — all they have to do is approach more closely or, at worst, start a fight. Because the signals are so easily verified (and because receivers severely punish signallers with misleadingly-large badges), production costs are not needed to ensure reliable signalling.

The comparison between this pair of examples leads us to the following prediction: *Conventional signals will be used when communicating about (1) coincident interests or (2) verifiable aspects of conflicting interests. "Costly" signals will be used otherwise.* In Section 4, we will extend this prediction to human language, to explain why much of human communication uses conventional signals, but certain messages are transmitted via more costly means of expression.

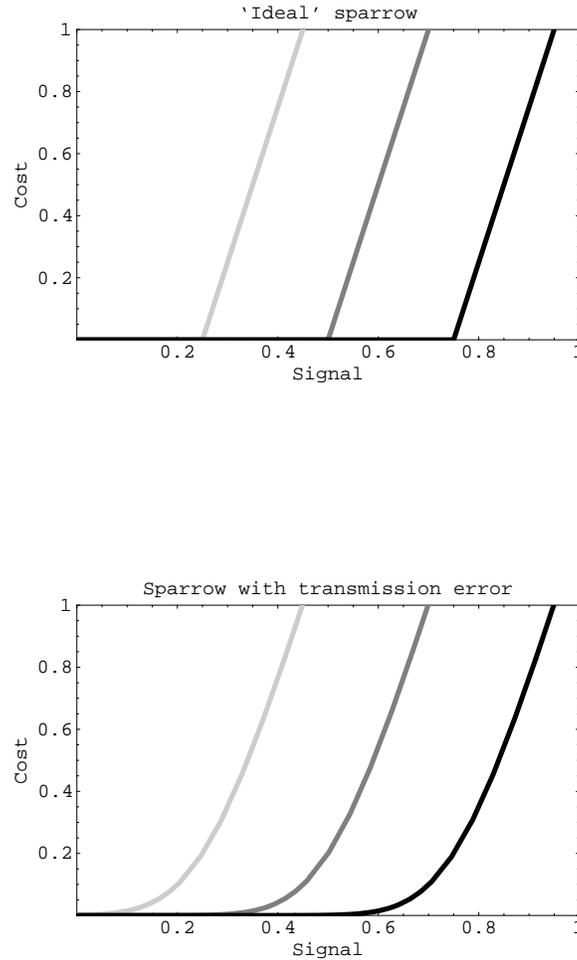


Figure 4: Imprecision decreases curvature of signal cost function and thus increases signalling costs in conventional signalling systems based on punishment of liars. A) Kinked cost functions for low quality (light gray), medium quality (gray), and high quality (dark gray) signallers. Here, signal costs are zero unless signallers use signal value larger than their quality, and thus signals can be free at equilibrium. B) If receivers are subject to normally-distributed error ($\mu = 0$, $\sigma^2 = 0.1$ in this figure), signal cost functions are smoothed and signal costs are non-zero.

Precision and cost in conventional signalling

As we explained in Section 2, to avoid high signal costs at equilibrium, signalling systems must employ cost functions that rise sharply around the signalling equilibrium. When cost functions take this shape, a new problem arises for socially-imposed costs: imprecision can be extremely costly when putatively deceptive signals are punished. When the costs are imposed socially, the signaller doesn't pay the costs associated with the signal level that he chose: rather, he pays the costs associated with the signal level that the receiver *thinks* that he chose. A misunderstanding or imprecision can hence be very costly, if the signal cost function rises sharply near the equilibrium signal level. In the sparrow case, this suggests that misunderstandings and imprecisions would lead to a considerable number of costly fights. If signallers can reduce the error (either by reducing errors in generation of the proper signal or by reducing errors in perception) they can lower the costs at equilibrium. Therefore, when costs are imposed socially, there will be selection for precise and accurate signals, in order to reduce costly errors.

Figure 4 illustrates this principle. Receivers punish signallers who signal higher than their true quality, generating the cost functions shown in the top graph. If there is no noise (error), signallers will signal their true qualities and there will be no signal cost at equilibrium (compare to figure 3c). However, perceptual error will introduce costs into the system. When there is normally distributed perceptual error, receivers sometimes will mistakenly punish signallers who have actually signalled their true quality. Under these circumstances, the cost functions take on the shape shown in the lower figure: costs to rise and signalling is costly at equilibrium.

Evolutionary stability

We have already seen that socially-imposed signal costs may allow cheaper signals because of their greater flexibility in the form of the cost function; socially-imposed signal costs may also be more resistant to breakdown over evolutionary time. A detailed treatment of evolutionary stability in the face of new signalling innovations is beyond the scope of the present paper and will be presented in a subsequent report; nonetheless we will briefly sketch this argument here.

When signal costs result from the physiology of signal production, natural selection actively works to destabilize a signalling system. Signallers will be selected to find cheaper ways of sending equivalent signals; over time, they are likely to acquire evolutionary innovations which reduce signalling costs. While the evolutionary dynamics which follow such an innovation can be extremely complicated, it is plausible that a series of innovations could ultimately destabilize a signalling system [41, 42]. In summary, when signal costs arise during production we might expect a dynamic tension between stabilizing forces of signal cost and the destabilizing forces of evolutionary innovation.

When costs arise through receiver response, the situation is different: natural selection will not necessarily favor innovations that reduce signal costs, because signal costs are no longer under the evolutionary control of the signallers who pay them. To put this another way, the only individuals able to reduce signal costs (the receivers) are unlikely to have any evolutionary incentive to do so. As a consequence, the destabilizing pressure of selection for reduced signal costs will not be a threat to signalling systems in which

cost is imposed by the signal receiver.

4 The evolution of human language

As mentioned in introduction, a number of authors have explored the evolution of human language by modeling linguistic communication as an evolutionary game played among individuals with coincident interests (a *team game*, in the parlance of game theory). But is human communication properly modeled as a team game? And is this assumption of common interest necessary in order to explain the evolutionary stability of linguistic communication? In this section, we apply the insights obtained from the peacock and sparrow to understand how the features of human language can be evolutionarily stable despite conflicting interests among the communicating parties.

Arbitrary signal-meaning association

In human language, the relationship between signal and meaning is largely arbitrary [43]. In models of costly signalling this is not the case; in these models, non-coincident interests require that signal meaning be related to the signal's production costs [44]. Our first challenge, in reconciling costly signalling theory with evolutionary models of human language, is to explain how arbitrary meanings can be assigned to signals in a stable way, despite conflict of interest and without signal cost. This problem has been noted before. Nowak and Krakauer [6] employ a simple model of cost-free signalling, and find that under non-coincident interests, the signal-object associations fail to form. In response, they conjecture that “cooperation may represent

an important prerequisite for the evolution of language.” Grassly et al. [45] observe that signal costs are low in human communication, and rule out Zahavian arguments on these grounds: “One property of arbitrary sign systems is that costs of production are very low, and hence conventional handicap style arguments do not seem pertinent.”

But do conflicting interests really force us to rule out arbitrary sign systems? When taken alongside the recent results from signalling theory, the sparrow analogy suggests that they do not. The argument comes in three steps: (1) reliable signalling can be maintained by the Zahavian mechanism without signal cost at equilibrium so long as the cost function takes the appropriate form. (2) Social enforcement of signal costs may be particularly likely to give rise to this form, where the cost of the signal is exquisitely sensitive to the quality of the signaller around the point of “true quality.” (3) When costs arise through social enforcement, the signal-meaning association can be arbitrary, so long as it is “properly understood” by the signal receiver (or whatever other party imposes the social cost.)

Conversely, human language provides a very nice example of how cost functions can assume forms that allow very cheap signalling, such as those discussed in ref. [35] and depicted in figure 3c. Namely, language allows cost functions that are flat up to a threshold, and then increase sharply. In such signals, social enforcement need impose no cost when signals are “honest” — but when signals are interpreted as dishonest, substantial social penalties may be imposed.

Moreover, the cost functions for linguistic signals with social costs will often rise sharply only on one side of equilibrium, while remaining essentially

flat on the other. That is, a signal misleading in one direction may not be penalized, whereas a signal misleading in the other may receive sharp sanction. We can explain this by noting that human sentences are typically uttered within a mutually understood context. Receivers can generally distinguish between misleading signals that are, and are not, harmful to themselves⁴. If a signal is misleading but not harmful, it probably will not merit punishment. Suppose, for example, that I plan to buy a used car. The car dealer tells me that one particular model gets 20 miles per gallon, and based in part on this information, I purchase the car. If I later discover that the car gets only 10 miles per gallon, I am likely to be angry and to take some sort of punitive action, be it a lawsuit, a fist-fight, or a verbal smearing of the dealer's reputation. If instead I were to find that the car gets 30 miles to the gallon, I would be quite unlikely to be angry or to punish the dealer in any way, even though, strictly speaking, the dealer's initial claim is every bit as false. What matters is more a matter of context than one of literal truth.

In this subsection, we have merely sketched an argument for how arbitrary signal-meaning associations can be stable despite conflicting interests. A full "demonstration of possibility" would include (1) a formal model by which signal and meaning take on arbitrary associations and are maintained stably by social enforcement and (2) a formal explanation of why the particular acts of social enforcement needed to stabilize the signalling system are themselves evolutionarily stable. Though we will not get into detail here, our preliminary investigations suggest that it will be possible to provide solutions to both

⁴They may also be able to make inferences about whether the signaller stands to gain from the deception, but we will not go into this here.

problems.

Selection for precision and accuracy

Under coincident interests, the advantages to sharing more information with reduced chance of error serve as a plausible selective force favoring higher levels of linguistic structure. Will these results carry over to scenarios of non-coincident interest as well?

It is not immediately apparent that they must. While assembly of phonemes into words and words in sentences will indeed allow increases in the volume, precision, and accuracy of communication, such gains could easily come at too high a cost. Indeed in certain signalling games, reductions in precision — of a particular sort, namely, the switch from separating to pooling equilibria — can drastically reduce signal costs, leaving all parties better-off despite the reduction in the precision of the information conveyed [44]. Moreover, reductions in accuracy can also be favored in certain signalling games. For example, in the Sir Philip Sidney game, where costs arise as a consequence of signal production, transmission error can reduce signal costs and can even benefit both parties [44]. Why, given these results, should we expect the evolution of increasing communication efficiency when interests are non-coincident?

First of all, the same advantages to improved communication that are described by Nowak and colleagues will be present to the degree that communication is beneficial to both parties, even in strategic signalling models without completely coincident interests. Second, the reductions in cost associated with a shift from separating to pooling equilibria will necessarily

apply only when signal costs arise as consequences of signal production (and cost functions are thus somewhat constrained). Once costs are imposed by signal receivers (or the constraints on cost functions are otherwise relaxed) signalling need not be costly and thus reduced precision due to pooling will not be required in order to enjoy low signal costs. Third, we note that as illustrated in Section 3, misunderstandings increase the total signal costs in systems where costs are imposed by the signal receivers on signallers. By reducing the chance of misunderstandings, signal costs can be reduced as well. Language can do this nicely; messages can be encoded carefully and redundantly, to minimize the chance of a misunderstanding.

So we can say that the features of language allow — and even favor — two distinct types of precision. First, the relaxation of constraints upon the signal cost function allows sharing of more information (a shift from pooling to separating equilibria) without increased costs. Second, the increased need to prevent misunderstandings when costs are socially imposed will select for the more careful and redundant signals that become possible through linguistic communication.

Language requires dissociation of meaning and cost

In human language, meaningful sentences are constructed combinatorically out of smaller discrete units that have their own referents. In terms of the volume and accuracy of information transmission, this striking property of language offers a number of advantages [6, 8] and thus will be favored when the interests among communicating parties coincide.

When interests do not coincide, however, it is difficult to see how this type

of combinatorical communication can be stabilized by a Zahavian mechanism. In Zahavian systems, signalling is kept honest by an intrinsic relationship between the meaning of a complete signal (hereafter, a *sentence*) and the production cost for this sentence. As we see from the peacock example, such a relationship can fairly easily be obtained for non-combinatorial forms of communication. This is much harder to do when a large number of sentences are built up by using a much smaller set of words in alternative combinations.

Consider an absurdly simple example, a Wittgensteinian “language game” in which an individual solicits food for his family from a somewhat sympathetic kinsman, using a six-word language. His request is made as follows. He first names the type of food item that he needs — “goat,” “rabbit,” or “apple” — and then specifies how many of that item he needs, “one,” “two,” or “three”. With just six signals, this language could conceivably be used to convey nine different requests. But in this system, how can costs be assigned to *individual words* to ensure reliable communication? For example, how much more should it cost to use the word “two” than to use the word “one”? If communication is to be reliable, the cost of using the word “two” must be sufficient to deter an individual who needs one goat but not two from saying “goat, two” instead of “goat, one”. But if this is the case, the word “two” will be too expensive to use in reference to the lesser food items, rabbits and apples. Thus when stabilizing costs are assigned to individual words, several of the nine possible “food,number” sentences (e.g. “apple,two”) will be so expensive that they are not worth sending. There is no way to associate words and costs such as to facilitate full yet reliable use of this “language.”

More complex languages, particularly those that employ negation, con-

junctions, and other complex grammatical constructs, will face even greater difficulties of this sort. In order to employ a Zahavian mechanism to stabilize reliable signalling, then, signal cost will somehow have to be associated not with the component phonemes or words but instead with each sentence as a whole. For the reasons described above, this is very difficult when signal costs arise from the physiology of signal production. However, an association between whole-sentence meaning and signal cost more easily obtains when costs are imposed by the individuals receiving and interpreting the signals, e.g., for conventional signalling systems as in the sparrow example of the previous section.

Therefore, we argue that the transition from Zahavian to conventional signals is a prerequisite for the evolution of combinatorical sentence meaning. In order to reap the benefits of building sentences from words, a signalling system must separate the meaning of a signal from the mechanism that keeps its use honest. How? The results discussed in Section 2, taken together with the idea that social control and high signal accuracy facilitate the proper cost function forms to allow cheap signalling, suggest an answer: when the signal receivers impose the stabilizing costs, these costs can be associated with *sentences* rather than phonemes.

Costly forms of human communication

Not all human communication is linguistic, despite the many advantages of this mode. As Thorstein Veblen noted over a century ago, humans make liberal use of costly signals in their myriad displays of conspicuous consumption [46]. Why resort to this inefficient form of communication?

In Section 3, we predicted that conventional signals will be used when talking about coincident interests or verifiable qualities, and costly signals will be used otherwise. Because of the social context in which human linguistic communication often occurs (structured populations with repeated interactions), much of human communication falls into the former domain. However, not all human interactions will meet these criteria; under circumstances that do not, we might expect to see the use of costly signals. In particular, we might expect humans to use costly signals (1) in one-time interactions, (2) when communicating about otherwise unverifiable properties, or (3) when the potential gains of deception outstrip the deterrent ability of community- or individually-imposed sanction or punishment.

5 Discussion

Chomsky [47] and Bickerton [48] have argued that human language could not have evolved gradually by natural selection from systems of animal communication [2]. In order to make a convincing counter-argument, one needs to demonstrate that (1) arbitrary meanings can be assigned to signals in an evolutionarily stable manner and (2) the various higher-level structures of language can emerge by evolution. Over the past decade, a number of authors have put together a relatively plausible response. In particular, Nowak and colleagues have presented a set of mechanisms by which language could have done precisely this.

However, these arguments have typically viewed human communication as an evolutionary game with coincident interests among the communicating

parties. In this paper, we have strengthened previous results considerably; by appealing to signalling theory, we have been able to show that the results derived under the assumptions of coincident interest actually hold much less restrictive conditions, including environments of conflicting interest as would have inevitably been present during language evolution.

Though we have presented arguments that human language can be evolutionarily stable in scenarios of conflicting interest, we have not shown that language must have arisen under such scenarios. Indeed, we see at least two hypotheses for how language could have come to be employed for such games. On one hand, language could have directly evolved under scenarios of non-coincident interest. Alternatively, language could have first evolved for use in “games” or interactions with coincident interests, and only later have been coopted for use in less cooperative situations. Further work on the actual evolution of language (rather than on the evolutionary stability of language) will be required to assess the relative feasibilities of these alternatives.

One reason to suspect that language may have initially evolved for use in coincident interest situations is that individual reputations — in which individuals become known as liars or as good sources of information — may provide an important “cost” to deception in human communication. While reputations can emerge even without communal “discussion” of an individual’s reliability, language certainly facilitates the spread of reputations by allowing individuals to share information about others’ reliability. Once language is established, reputations can be a powerful incentive against unreliable signalling — but until language has been founded, reputation effects could be hard to maintain.

Using signalling theory, we derived results similar to those of Nowak and colleagues, for scenarios of conflicting interest. If this were all that signalling theory could do for the study of language evolution, then its contribution would be brief. Having shown via signalling theory that language evolution worked must as previously supposed but under more general circumstances of non-coincident interest, evolutionary biologists could safely leave signalling theory behind and retreat to the earlier coincident-interest approach to language evolution. But signalling theory has a utility beyond what we have expressed explicitly here. First, signalling theory can inform us about the *dynamics* of signal evolution; while we have shown that similar linguistic properties will be evolutionarily stable whether interests are coincident or conflicting, we know very little about what sorts of evolutionary dynamics are to be expected at the origin of linguistic communication in scenarios featuring conflicting interests. Second, signalling theory allows us to predict when signals will be costly and when they will not, and explains the form of signals, both costly and free. Third, signalling theory predicts that when social enforcement is in place, signals still may not be cost-free. Thus signalling theory helps us to predict how much social cost will be incurred at equilibrium, given some details of how the social enforcement works (i.e., the form of the signal cost function). For all of these reasons, we expect that signalling theory will provide a useful foundation for further exploration of human linguistic communication as an evolutionary game.

Acknowledgments

The authors thank E. A. Smith for his insightful suggestions.

References

- [1] J. R. Hurford. Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua*, 1989.
- [2] S. Pinker and P. Bloom. Natural language and natural selection. *Behavioral and Brain Sciences*, 13:707–784, 1990.
- [3] M. Oliphant. The dilemma of Saussurean communication. 37:31–38, 1996.
- [4] J. Maynard Smith and E. Szathmáry. *The Major Transitions in Evolution*. Oxford University Press, 1997.
- [5] A. Rubinstein. *Economics and Language*. Cambridge University Press, 2000.
- [6] M. A. Nowak and D. C. Krakauer. The evolution of language. *Proceedings of the National Academy of Sciences USA*, 96:8028–8033, 1999.
- [7] M. A. Nowak, J. B. Plotkin, and D. C. Krakauer. The evolutionary language game. *Journal of Theoretical Biology*, 200:147–162, 1999.
- [8] M. A. Nowak, J. B. Plotkin, and V. A. A. Jansen. The evolution of syntactic communication. *Nature*, 404:495–498, 2000.
- [9] M. A. Nowak, N. L. Komarova, and P. Niyogi. Evolution of universal grammar. *Science*, 291:114–118, 2001.
- [10] S. Pinker. Survival of the clearest. *Nature*, 404:441–442, 2000.

- [11] C. Knight. Ritual/speech coevolution: a solution to the problem of deception. page 68. 1998.
- [12] A. Zahavi. Mate selection - a selection for a handicap. *Journal of Theoretical Biology*, 53:205–214, 1975.
- [13] A. Grafen. Biological signals as handicaps. *Journal of Theoretical Biology*, 144:517–546, 1990.
- [14] J. Maynard Smith. Honest signalling: The Philip Sidney game. *Animal Behaviour*, 42:1034–1035, 1991.
- [15] H. C. J. Godfray. Signalling of need by offspring to their parents. 352:328–330, 1991.
- [16] O. Hasson. Cheating signals. *Journal of Theoretical Biology*, 167:223–238, 1994.
- [17] O. Hasson. Towards a general theory of biological signaling. *Journal of Theoretical Biology*, 185(2):139–156, 1997.
- [18] J. Maynard Smith and D. G. C. Harper. Animal signals: Models and terminology. *Journal of Theoretical Biology*, 177:305–311, 1995.
- [19] C. T. Bergstrom and M. Lachmann. Signalling among relatives. III. Talk is cheap. *Proceedings of the National Academy of Sciences USA*, 95:5100–5105, 1998.
- [20] Sz. Számadó. Cheating as a mixed strategy in a simple model of aggressive communication. *Animal Behaviour*, 59:221.

- [21] A. Zahavi. The cost of honesty (Further remarks on the handicap principle). *Journal of Theoretical Biology*, 67:603–605, 1977.
- [22] N. Nur and O. Hasson. Phenotypic plasticity and the handicap principle. *Journal of Theoretical Biology*, 110:275–297, 1984.
- [23] R. A. Johnstone and A. Grafen. The continuous Sir Philip Sidney game: A simple model of biological signalling. *Journal of Theoretical Biology*, 156:215–234, 1992.
- [24] J. Maynard Smith. Must reliable signals always be costly? *Animal Behaviour*, 47:1115–1120, 1994.
- [25] S. Rohwer. The evolution of reliable and unreliable badges of fighting ability. *American Zoologist*, 22:531–546, 1982.
- [26] M. Enquist. Communication during aggressive interactions with particular reference to variation in choice of behaviour. *Animal Behaviour*, 33:1152–1161, 1985.
- [27] J. Maynard Smith and D. G. C. Harper. The evolution of aggression: can selection generate variability? *Philosophical Transactions of the Royal Society of London, Series B*, 319:557–570, 1988.
- [28] H. Viljugrein. The cost of dishonesty. *Proceedings of the Royal Society of London Series B*, 264(1383):815–821, 1997.
- [29] J. B. Silk, E. Kaldor, and R. Boyd. Cheap talk when interests conflict. *Animal Behaviour*, 59:423–432, 2000.

- [30] T. Guilford and M. S. Dawkins. What are conventional signals? *Animal Behaviour*, 49:1689–1695, 1995.
- [31] A. Zahavi. On the definition of sexual selection, fisher’s model, and the evolution of waste and of signals in general. *Animal Behaviour*, 42:501–503, 1991.
- [32] A. Zahavi. A reply to John. *Animal Behaviour*, 54:228–229, 1997.
- [33] T. Getty. Handicap signalling: when fecundity and viability do not add up. *Animal Behaviour*, 56:127–130, 1998.
- [34] P. L. Hurd. Communication in discrete action-response games. *Journal of Theoretical Biology*, 174(2):217–222, 1995.
- [35] M. Lachmann, C. T. Bergstrom, and S. Szamado. The death of costly signalling? *Santa Fe Institute Working Paper Series*, (00-12-072), 2000. <http://www.santafe.edu/sfi/publications/working-papers.html>.
- [36] S. Rohwer. The social significance of avian winter plumage variability. *Evolution*, 29:593–610, 1975.
- [37] D. P. Whitfield. Plumage variability, status signalling and individual recognition in avian flocks. *Trends in Ecology and Evolution*, 2:13–18, 1987.
- [38] S. Rohwer. Status signaling in harris sparrows: Some experiments in deception. *Behavior*, 61:107–129, 1975.
- [39] S. Rohwer and F. C. Rohwer. Status signals in Harris’ sparrows: experimental deceptions achieved. *Animal Behaviour*, 26:1012–1022, 1978.

- [40] A. P. Moller. Social control of deception among status signalling house sparrows *Passer domesticus*. *Behavioral Ecology and Sociobiology*, 20:307–311, 1987.
- [41] G. E. Hill. Trait elaboration via adaptive mate choice: sexual conflict in the evolution of signals of male quality. *Ethology Ecology and Evolution*, 6:351–370, 1994.
- [42] H. C. J. Godfray. Evolutionary theory of parent-offspring conflict. *376:133–138*, 1995.
- [43] S. Pinker. *The Language Instinct*. Harper, 1995.
- [44] M. Lachmann and C. T. Bergstrom. Signalling among relatives. II. Beyond the Tower of Babel. *Theoretical Population Biology*, 54:146–160, 1998.
- [45] N. C. Grassly, A. von Haeseler, and D. C. Krakauer. Error, population structure, and the origin of diverse sign systems. *Journal of Theoretical Biology*, 206:369–378, 2000.
- [46] T. Veblen. *The Theory of the Leisure Class*. Dover Publications, Mineola, N.Y., 1889.
- [47] N. Chomsky. *Rules and representations*. Columbia University Press, New York, 1980.
- [48] D. Bickerton. *Language and species*. University of Chicago Press, Chicago, 1990.