Communication and Representation Understood as Sender-Receiver Coordination

Ronald J. Planer and Peter Godfrey-Smith

To Appear in *Mind and Language*

1. Introduction

In 1969 the philosopher David Lewis published *Convention*. Lewis wanted to defend the idea of *conventions of meaning*, a notion which had been attacked by Quine in some of the most influential work in mid-20th century philosophy. Lewis was especially interested in linguistic conventions, but he set out from simple cases such as Paul Revere's lantern code – *one if by land, two if by sea*. He did this with the aid of a simple model which described the interaction between pairs of behaviors on "each side" of a sign—behaviors of *making* signs and behaviors of *interpreting* them. He showed how some of these combinations of behaviors could be stable, giving rise to useable conventions, when the agents making and using signs share common interests, make rational choices, and are characterized by specific asymmetries in what they can perceive and do.

Lewis's model has always been seen as important in discussions of convention itself, and in related parts of philosophy of language, but for many years it had little impact elsewhere. In particular, it was not seen as the basis for a *naturalistic* semantic theory. It is not surprising that Lewis's model seemed unimportant for naturalistic projects, as Lewis took intentions and beliefs for granted in specifying his model. Further, Lewis seemed to build a great deal *in* to his model in order to get quite limited results *out* of it. Lewis's agents were assumed to be rational, and to have many beliefs about each other, while the signaling that arises between them was very simple.[1]

Brian Skyrms, in his 1996 book *Evolution of the Social Contract*, generalized the Lewis model. Skyrms showed that Lewisian signaling can evolve by natural selection as well as be rationally chosen. It can evolve in organisms with little or no psychology. The *conventionality* that Lewis took as his topic has analogues in systems of other kinds, where different selection processes shape the behaviors of sign production and use.

This project has continued.[2] Bridges have been established between the Lewis model and work in economics, biology, and linguistics. Connections between treatments of communication in all these fields have become clear in retrospect. The Lewis model can be seen as a "minimal model" of communication in general.

---

[1] See Rescorla (2017) for an overview.

[2] See, e.g., Skyrms (2004; 2010), Huttegger (2007), Zollman (2011), Barrett (2009), Smead (2014), O'Connor (2014), Harms (2004), and Godfrey-Smith and Martínez (2013).

In this article, we explore how the Lewis model and its variants relate to different domains, and see what picture emerges of the underlying general features of sign use. We will begin with a presentation that covers the Lewis model's formal core but is minimal with respect to technicality. This presentation includes a treatment of syntax, often seen as a problem for views of this kind but in fact something that can be handled in natural and illuminating ways. Then we will discuss four further topics: (i) cooperation and common interest; (ii) signaling within organisms; (iii) meaning; and (iv) human communication and language. We argue that the Lewis model not only has great utility as a modeling framework; it fosters a novel perspective on the nature of communication itself, and the nature of semantic and representational properties.

2. Sender-Receiver Models

Lewis imagined two agents, who we will call the "sender" and "receiver" (Lewis said "communicator" and "audience"). The sender can perceive the state of the world, which is determined exogenously, but cannot act except to produce signs of some kind. The receiver can only perceive these signs, but can act in a way that has consequences for both agents.

The sender applies some *sender's rule*, a mapping from states to signs; the receiver applies a *receiver's rule*, mapping signs to acts. Composed, the result is a mapping from states to acts. Those pairings of acts and states determine payoffs for both agents. So the model assumes a causal flow from state to sign to act, with behavioral choices at two points that implement mappings.

The payoffs to each agent depend on the pairing of receiver actions with states of the world, and can be represented in a matrix (Table 1). Here each cell contains two numbers which represent the sender's and receiver's payoffs, respectively, for a particular combination of receiver's action and world state.

		States		
		S1	**S2**	**S3**
	A1	5,5	0,0	0,0
Acts	**A2**	1,1	2,2	1,1
	A3	0,0	0,0	5,5

Table 1: A payoff matrix for a Lewis signaling model.

Lewis assumed *common interest*, *common knowledge*, and *rational choice* on both sides. The Sexton of the Old North Church in Boston and Paul Revere both want to defeat the British army, and agree on the actions best suited to different states of the world. There are various ways they can achieve the coordination they aim at, including using the famous lantern code: *one of by land, two if by sea*. The Sexton is to hang one lantern in the church tower if the British are invading by land, two lanterns if they are coming by sea. That is a one-to-one mapping from states to signs implemented by the sender, the Sexton. The receiver, Revere, maps signs to acts in way that ensures appropriate preparation of defenses, given the state of the world. This combination of rules comprises a *Nash equilibrium*: neither side can improve their payoff by unilaterally changing their behavior.

The sender's rule need not be one-to-one. Suppose the sender always sends the same sign, regardless of what they observe. Then the signs carry no information, in Claude Shannon's (1949) sense, about the state of the world; they do not change any of the probabilities of states of the world. In this case, the best response for the receiver is to always produce their best available cover-all behavior. This combination of rules is also a Nash equilibrium.

Other options are possible. The sender may refuse to distinguish some states of the world, but send a unique sign in others. That is a case of "partial pooling" of states. The sender may also use a mixed strategy, in which they respond to a particular state of the world by sending one sign with probability p and others with other probabilities, summing to one. The receiver, also, may "pool" some or all of the signs when working out how to act, and may probabilistically "mix" their behaviors in response to some signs.

In *Evolution of the Social Contract*, Skyrms dropped Lewis's assumptions of common knowledge and rational choice. He showed that evolution by natural selection can shape behaviors in a way that yields Lewisian signaling. Among other things, this required that combinations of sender and receiver rules be assessed in terms of their effects, not in terms of what an agent represents as preferable. Skyrms's emphasis in developing an evolutionary interpretation of Lewis's model has been on *robustness*, showing that Lewisian signaling can arise in many variants of an evolutionary model.[3] As William Harms (2010) pointed out, naturalistic versions of the Lewis model bear a close relationship to the approach to communication and meaning developed in the 1980s by Ruth Millikan (1984).

Skyrms then noted that simple models of change by natural selection are not only models of biological evolution. Any process in which successful behavioral dispositions are retained and unsuccessful ones dropped, by "trial and error" in a broad sense, has some fit to the dynamics of evolution. This might be reinforcement learning (trial and error by individuals). It might be differential imitation, rather than differential reproduction. If successful behaviors at one time-step are imitated at the next step more than unsuccessful behaviors—*imitation of the successful*—this also allows adaptive behaviors to proliferate in the population.

Hence, there is a family of selection processes that can stabilize Lewisian signaling. The four we've mentioned can be roughly ordered by their increasing cognitive demands: *biological evolution*, *reinforcement learning*, *imitation of the successful*, and *rational choice*. These selection processes work on different time-scales and produce somewhat different outcomes, but all can stabilize Lewisian signaling in some cases.

[3] See Skyrms (2010) for an overview.

In all these cases there is coadaptation of sender and receiver behaviors. The rules on each side of a sign can each be partly explained in terms of the nature of the rule seen on the other side. The sender produces signs in a certain way because of the receiver's rule downstream, while the receiver interprets signs in a certain way because of the sender's rule behind the signs encountered. In the case of rational choice, it is *envisaged* sender or receiver behavior that shapes the behavior of the other side. In biological evolution and reinforcement learning, the behaviors must be actually produced and feedback involves actual effects. The selection processes differ, but there is broad similarity in the existence of a mutual shaping of behaviors on each side. The orientation the model lends to questions about communication is: *Why send messages at all? Why attend to what is sent? Why give information to another, on the senders side, and on the receiver side, why guide your actions with what another is saying?*

As we proceed, we will contrast this view of communication – sometimes explicitly, sometimes more implicitly – with other historically influential frameworks (e.g., Gricean communication, Relevance Theory, Interpretavism). There is also an interesting connection between the Lewis model and information theory, as developed by Claude Shannon (1948). Shannon's model featured a *source* (a variable state of the world), a *transmitter*, who sends a *message* along a *channel*, a *receiver*, and a *destination* where the information is put to use in some sense. Shannon gave a quantitative treatment of uncertainty at the source, and the role of the sign in reducing uncertainty about the source. The theoretically relevant properties of a sign are the ways it reduces that uncertainty. In retrospect, we can see Lewis's theory as fitting hand in hand with

Shannon's: Shannon took for granted occupants of the sender and receiver roles, and gave an account of the channel conditions that would make communication between them possible; Lewis took for granted channel conditions that would allow sender and receiver to communicate, and gave an account of why agents would ever want to occupy those roles.

The Lewis model is a model of the use of signs to enable coordination of receiver acts with states of the world. This can be called *act-state coordination*. There is another kind of coordination that communication can help to achieve, coordination between two agents' acts: *act-act coordination*. It might be that the sender first decides how to act and then sends a sign, with the receiver conditioning their own act on what they see. Alternatively, it might be that both sides send and receive.

The game traditionally called "Battle of the Sexes" is one where payoffs are determined by combinations of acts, for example, and models of this kind often feature pre-play signals sent by one or both agents. Two pairs of acts achieve coordination between the partners (both attend the ballet, both attend the football match). Each agent prefers one of these combinations to the other, but prefers both of those to a failure to coordinate. The *Stag Hunt* is another example (Skyrms 2004). Here there is a cooperative option (a pair of agents hunt stag together). The best response to a cooperative act is to cooperate, while the best response to non-cooperation is non-cooperation. In all these cases, payoffs result from pairings of acts, and if signal use can be maintained, its role is to coordinate act with act, not an act with an exogenously given state.

Many empirical cases will contain a mixture of both kinds of coordination. The general outlook we get from the Lewis model still applies, though the signs now have semantic involvement with different kinds of things. Rather than representation of an exogenous state, a sign's role is more directly tied to social coordination—expressing intentions and directing action. We will use the term "sender-receiver model" to cover both the Lewis model and its extensions, as well as models of act-act coordination. These models share a common structure in featuring senders and receivers engaged in some form of interdependent sign production and interpretation. This structure is like a hub from which many more detailed case-specific models can be derived. We will use the term "sender-receiver framework" for the orientation to communication and representation that is engendered by these models. This orientation, while shaped by formal models, is not itself essentially formal.

The sender-receiver framework is sometimes thought too simple to illuminate anything but the most primitive sorts of signs; in particular, syntax has been seen as a problem.[4] We disagree: the framework is in fact very useful for understanding the origins of syntax and related phenomena. Syntax, and related features of sign systems, are to be understood in terms of their role in sender-receiver interaction. The question is not, or not simply, "What kind of structure is present?" It is what kinds of structure are put to use in communication, and how. This constraint falls naturally out of the sender-receiver

[4] See Rescorla (2017) for a discussion.

framework. Here we present a taxonomy of signs that makes more distinctions than is usual. These distinctions are set up to illuminate possible evolutionary relationships and transitions. This framework complements a growing family of formal and informal models of the emergence of structured signaling.[5]

We regard syntax as the bringing together of two simpler features, *combinatorial structure* and *encoding*. All macroscopic entities have structure in some sense. What matters is whether that structure makes a difference to behaviors of sign production and interpretation. We define combinatorial signs as follows. First, there is *sharing of parts* among signs within a system. Signs that do not satisfy this condition are *atomic*. Second, the *sequence properties of the signs matter* in the sense that, by intervening on them, one can change how the signs function. By "sequence properties," we mean the different ways a sign's parts are ordered with respect to one another in space or time. Imagine, for example, a system in which each sign consists in a sequence of 'a's and 'b's but the receiver only attends to the *ratio* of 'a's to 'b's in a given sign. These signs satisfy our first condition, but not our second. Two signs that differ with respect to their sequence properties but have the same ratio of 'a's to 'b's are not distinguished by the receiver. We call signs that satisfy only our first condition *composite signs*. Composite signs have parts, but the order of these parts is not causally relevant. So there is a three-way distinction – *atomic, composite, combinatorial* – bearing on this first aspect of syntax.

[5] See, e.g., Nowak and Krakauer (1999); Nowak et al. (2000), Kirby (2000), Barrett (2009), Tria et al. (2012), Scott-Phillips and Blythe (2013), Franke (2014).

To grasp the second feature, which we call "encoding," it is best to start from a related feature that is weaker. We call this weaker feature "organization." In an organized system, there is some relation among signs that plays a communicative role. For example, imagine an alarm call system where, in a very rough way, louder calls are produced in response to closer and more immediate threats, and receivers attend to call intensity in their responses. There is a relation between sounds, *louder_than*, that maps to a relation between states of the environment, *closer_than*. It is possible for receivers to be oblivious to this relation between calls; senders might produce louder calls the more agitated they are, but receivers might attend only to the presence of a call. Then the *louder_than* relation between signs would be epiphenomenal; present, but unused. This would be a case of what we call a *nominal system* – one where there is no communicative role for a relation between signs. On the other hand, if receivers *do* attend to this relation—something that might evolve quite readily—then we have an organized sign system. If this relation-between-relations (between *louder_than* and *closer_than*) is rough and unsystematic, then the system is not yet a system of *encoded* signs. But there is a natural pathway to encoding; in time, refinement of the system might produce a situation where the distance of a threat maps the decibel level of a call by means of a particular rule (e.g., N meters $\rightarrow$ (110 – N) decibels.)

When combinatorial signs have their meaning specified by an encoding principle that gives a semantic role to the sequence properties of signs, we have *syntax*. A familiar example of syntax in this sense is binary encoding. Here each sign is composed out of a sequence of ‘1’s and ‘0’s, where the right-most bit stands for the number of factors of 2^0, the second right-most bit stands for the number of factors of 2^1, and so on, with concatenation standing for addition. These signs share parts *and* their sequence properties matter. Moreover, they are governed by an encoding principle that gives a role to the sequence properties. Taking a set of binary encoded signs and randomizing the mapping from signs to numbers would result in the loss of syntax, on our view. They would still be combinatorial, but not encoding, and hence lacking in syntax. If one were to further intervene so that sequences having the same number of ‘1’s and ‘0’s were made into functional equivalents (resulting in a loss in expressive power), the signs would then be composite but not combinatorial.

So, in identifying syntax we make use of a pair of three-way distinctions: *nominal-organized-encoding*, and *atomic-composite-combinatorial*. The result is a 3x3 classification which is represented in Table 2 below. We include possible examples of each category, some of which are merely hypothetical. Many empirical cases will only approximate these categories.

Syntax is then the product of two kinds of transitions, one that introduces relations among signs in successively stronger forms, and one that introduces part-whole structure into the signs, again in successively stronger forms. These sequences correspond to plausible evolutionary pathways, though we do not contend that more elaborately structured systems must *always* pass through less elaborately structured ones first.

It might be questioned whether all 9 cells of this table represent genuine possibilities. Ruth Millikan (1984) has argued that the usual descriptions of simple, apparently "atomic" signs like alarm calls overlook a hidden kind of structure. An alarm call is typically said to mean *danger* (or something similar), but in fact its content is something like *danger here now*, where the time and place of the call maps to the time and place at which danger is said to be present. (In Millikan's terminology, time and place are "aspects" of the sign.) We agree that simple signs are very often (though not always) "articulated" in the ways Millikan describes; the time of call production contributes to the call's content, to what it says, and similarly for place. This puts pressure on the nominal category in our classification, though not on the atomic category, as time of production is not a physical part of the call.

We think this sort of quasi-organization is something of a special case, and does not prevent us from marking out nominal signs as a useable category. In these alarm call cases, the receiver need not employ anything like a nontrivial rule or device for the interpretation of these aspects of the call. The time at which the call is produced does indeed contribute to its content, but interpretation of this aspect of the sign comes "for free" as long as the receiver is disposed to treat the call as a sign of danger. Similarly, a mating call means something like *mate with me*, but the "with me" aspect requires no specific interpretative processing. It would be possible for a mating call to mean *mate with the individual two places to my right* or for an alarm call to mean *danger this time next week*, but these would be significantly different systems; some nontrivial interpretation of the spatial and temporal aspects of the calls would be needed. We thus treat simple alarm call systems as cases where the signs are nominal and atomic (see Table 2).

	Nominal →	Organized →	Encoded
Atomic ↓	No sharing of parts across signs; no utilization of relations among signs. Ex.: Many animal alarm calls.	No sharing of parts across signs; some utilization of relations among signs. Ex.: An alarm call system where call intensity relates to distance of predator in a rough way.	No sharing of parts across signs; sign form governed by an encoding principle. Ex.: An alarm call system where call intensity precisely maps distance of predator.
Composite ↓	Sharing of parts across signs; sequence properties do not matter; no utilization of relations among signs. Ex.: A bird song system where receivers attend only to the presence or absence of species-typical syllables.	Sharing of parts across signs; sequence properties do not matter; some utilization for relations among signs. Ex.: Early stages of honeybee waggle dance; a bird song system where receivers attend only to overall song complexity	Sharing of parts across signs; sequence properties do not matter; governed by an encoding principle. Ex.: Actual honeybee waggle dance.
Combinatorial	Sharing of parts across signs; sequence properties matter; no utilization of relations among signs. Ex.: (A few) great-ape gesture sequences (gesture sequences that function differently than their constituent gestures taken individually, and where effect depends on order of the gestures).	Sharing of parts across signs; sequence properties matter; some utilization for relations among signs. Ex.: Human language words (nouns share more parts with other nouns than verbs, and vice-versa); DNA triplet code (optimized for error minimization).	Sharing of parts; sequence properties matter; governed by an encoding principle (that gives a semantic role to sign parts). Ex.: Human language sentences; genes.

Table 2. A taxonomy, distinguishing different kinds of structure seen in sign systems.

Another perceived limitation of sender-receiver models arises in connection with talk of a “sender's rule” and a “receiver's rule.” This language might be taken to suggest that there has to be a kind of fixity in the use of a sign. This is just the simplest case. Lewis considered cases where there is “discretion” on one side or the other; the sender can make available information despite not knowing what the receiver will do, if the sender is confident that whatever the receiver might do, given all their available evidence, will be acceptable from the point of view of the sender. Lewis said that the distinction between indicative (descriptive) and imperative (command) contents depends on which agent's policies include more discretion. When receiver has substantial discretion and the sender does not, this yields an indicative content. When discretion is greater on the other side, the content is imperative. This proposal has been refined in later discussions, largely in the spirit of Lewis’s suggestion. [6]

We have been referring to the Lewis “model” and similar “models.” By this, we mean that these structures are idealizations. In application to just about any empirical case, they will simplify in many ways. Applying a sender-receiver model to human language with its inherent complexity and ontogenetic links will involve greater

[6] See Huttegger (2007) and Zollman (2011) for attempts to improve on Lewis's treatment.

idealization than applications to non-human communication. Evolutionary models that yield equilibria and other results embody their own idealizations.

A model-based orientation to communication naturally leads to a view on which the status of an interaction as communicative is a matter of degree. There are *clear* or *paradigm* cases of sender-receiver systems and *marginal* ones, cases that only partially fit the sender-receiver structure. A case might be marginal because there are not well-demarcated and distinct entities playing the essential roles, or because only some of what is going on in a complex situation is shaped by what the models cover. This perspective generates a cluster of research questions: what processes, in evolution, culture, or learning, take us *from* marginal *to* clearer status, and which take us in the opposite direction? When will the sender and receiver roles become clearer and more distinct, as opposed to entangled and non-differentiated?

A diagnostic point follows as well. We as humans are not just accustomed to participating in sender-receiver systems, but also to talking *about* them. We have habits of response to sign-like objects that are shaped by experience with paradigm public cases—words, diagrams, maps. Psychologists often claim that we are (over) enthusiastic agency-detectors.[7] We think that we are also enthusiastic sign-detectors, and the sender-receiver structure is one we tend to see very readily. The sender-receiver structure, then, has a dual role. It describes a natural configuration, one that occurs on many scales. And

[7] See, e.g., Barrett (2004); Gray and Wegner (2010).

it describes a target of our *interpretive habits*. A full treatment of communication and meaning will include both of these: it will describe the actual role played by signs in mediating between sender and receiver, and also how we respond to sign systems when we encounter them as onlookers and theorists. Given these habits, we will often be tempted to describe marginal cases in ways that really apply only to paradigm cases. We have habits of response that are shaped by experience with paradigm public cases and these habits carry over when we encounter partial cases in areas like biology. Sometimes we are aware of the extent to which these descriptions are not literally true, but sometimes these habits produce genuine confusion.[8]

3. Common Interest

Communication is often seen as a fundamentally cooperative affair, a view closely associated with Paul Grice. According to Grice, communication is governed by a *cooperative principle*: each participant makes “a conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange” (Grice, 1975). Similar views have been expressed in many other literatures.

There is a question regarding how “cooperation” is to be taken in this context. Some theorists—in particular, Gloria Origgi and Dan Sperber (2000)—hold that only “cooperation” in a very minimal sense is required: the sender must be motivated to be

[8] Arnon Levy (2011) treats the concept of information in genetics and cell biology in terms of a fictional application of a sender-receiver framework.

understood, and the receiver to *understand*, where understanding does not imply *accepting-as-true* what is said. Others make stronger claims, where there can be no doubt that sender and receiver interaction is held to require a broader form of cooperation. It is this stronger view that is most relevant here. We see these theorists as guided by the following thought: if a sender provides information about the world to another agent whose interests are diametrically opposed, then any receiver's actions guided by that information are likely to be unwelcome, and the sender is better off providing no information at all. As in the "Miranda rights" formulation used by US police forces, you may want to remain silent when “anything you say can be used against you.”

Lewis dealt only with situations in which, as he put it, common interest “predominates.” He did not assume that sender and receiver receive the *same* payoff from a given pairing of act with state, but assumed that the differences were small, and in general, the sender and receiver want the same acts performed in a given state. Clearly, this will often not be the case. What becomes of communication with lower degrees of common interest?

Situations of this sort have been looked at in biology and economics using models somewhat different from the Lewis model. A famous example in economics is Crawford and Sobel (1982). Here the sender has private information about the state (the sender’s quality as an employee) and both the state of the employee and the actions available to her prospective employer (a judgement of sender quality) vary in one dimension. The sender sends a message that can be interpreted as representing their quality, and the receiver acts on that basis. As interests diverge, fewer distinct signals can be maintained, until informative signaling collapses altogether. Similar models have been used in the biological literature to understand the origins and stability of mating displays advertising quality.[9] The message of this work has been that when interests are significantly divergent—and, in particular, when the sender benefits from advertising that it is of higher quality than it is—signs will become meaningless unless special factors are present that enforce honesty (factors which cause dishonest senders to pay higher costs or benefit less).

Some have suggested that when common interest cannot be taken for granted, we should abandon talk of "information transmission" between senders and receivers altogether, and conceive of signalers as attempting to *manipulate* or *influence* receivers. This approach to animal signaling was pioneered by Dawkins and Krebs (1979) and has been defended more recently by Owren, Rendall, and Ryan (2010). Interaction will often still occur in such cases; when interests diverge, senders will make use of signs which, in virtue of their intrinsic properties, are difficult to ignore or affectively salient, thereby securing effects on receivers.

Let us look more closely at these issues. First, it is necessary to represent the idea of common interest in a more exact way. Suppose a situation features *n* possible states of the world and *m* possible receiver actions. In each state, the sender and receiver both have

[9] See, e.g., Maynard-Smith and Harper (2003) and Searcy and Nowicki (2005).

a preference ordering over the *m* actions. Godfrey-Smith and Martínez (2013) define several measures of common interest based on the degree of divergence between those preference orderings. In their simplest measure (*C*), *complete common interest* obtains when sender and receiver have the same preference ordering over acts in every state, while *complete conflict of interest* obtains when sender and receiver preference orderings are reversed in every state. Between these extremes are cases of *partial common interest*, measured by the number of pairwise disagreements over the relative value of actions in a state. They then ask: to what extent does the degree of common interest in a signaling game predict the presence and informativeness of communication in that game? And what is the minimum level of common interest that permits informative communication to be sustained at all?

A computer search of hundreds of thousands of 3-state games found that the level of common interest in a game was highly predictive of whether communication was sustainable, and predictive also of the amount of information conveyed in signals. The search assayed whether a given game has at least one Nash equilibrium in which informative signals are sent and used. Surprisingly, the search also found a small number of cases where communication could exist despite the sender and receiver disagreeing *entirely* on their preferences about actions in *every* state of the world. This might be interpreted as showing this reversal of preferences is not a strong enough criterion for "complete conflict of interest," but the finding is surprising however it is described. A model using dynamic methods, rather than a search for equilibria, yielded similar results (Martínez and Godfrey-Smith 2013).[10]

One stronger conception of "complete conflict of interest" draws on the idea of a *constant-sum game* – a game where the payoffs to sender and receiver on any occasion always sum to a constant value, and so a gain for one agent is always at the expense of the other. The maintenance of information-using equilibria does appear to be impossible in constant-sum games, though we are unaware of a proof of this result. Elliott Wagner (2010) found that the replicator dynamics could produce chaotic dynamic regimes in constant-sum games, with continually shifting mappings of states to signs and signs to acts.

These results are based on idealized models, but they have consequences for claims about the links between communication and common interest. They show, first, that there is *some truth* in the informal claims routinely made in this area. Common

---

[10] They found that, for intermediate and high levels of C, the evolutionary dynamics found Nash equilibria. In contrast, for very low values of C, if a static Nash equilibrium exists for the game, the replicator dynamics will not reach the state that corresponds to an equilibrium, but will orbit closely around it. There are also C=0 games with no Nash equilibria in which the evolutionary process leads the sender and receiver populations to maintain a state in which informative signals are sent and used, but no equilibrium is reached.

interest helps maintain communication and predicts the amount of information made available by a sender. But there are surprises; cases where the best act for one side is always the worst act for the other can support informative signaling at equilibrium. Models also show the possibility of chaotic and cyclical outcomes in which signs retain some information about states, but the mappings between states and signs and between signs and acts are always in flux. This bears upon the assumption, mentioned above, that in the absence of common interest or differential costs, signs will evolve towards meaninglessness.

4. Within the Organism

The schema seen in the Lewis model can be understood as existing *between* agents or *within* them. The causal flow from senses to effectors in an organism is a simple example of the latter. Millikan's analysis of semantic properties, with its emphasis on (cooperating) producers and consumers, has this feature too. As she noted, the set-up recognized by her theory can exist whether the producer and consumer are inside one head or are different agents. This suggests that semantic properties may be given the same treatment whether they are properties of internal or external representations. Neither sort of representation is primary in principle, and the same family of patterns can be used to understand both. However, showing that internal representations really do fit a sender-receiver structure is not straightforward. A sender-receiver system is not just any series of causal factors operating in a row. A receiver or *reader* (as we will now often say) is not just everything downstream of a putative representation, nor is a sender or *writer* simply everything upstream.

As we said above, there are both clear and partial cases of sender-receiver systems. The clearest are between-agent cases. Signs exist “between” two organisms or similar agents, each of which has its own agenda. The sign is produced by one agent and interpreted or used by another. Materially, a good vehicle for communication in the paradigmatic, highly cooperative, between-organism cases is something that is cheap and easily controlled, and these signs have a "passive" character in relation to the more active sender and receivers who use them. In the case of "signals" made and used within the boundaries of an organism, new possibilities arise. The "channel" between sender and receiver is often made up of living material. Hormones are quite a lot like external signals in the way they are produced and received, but patterns of neural activity are different.

Neurons might start out as simple intermediaries at early evolutionary stages, but once they are present, there is much more they can do. Early discussions of the evolution of neurons often emphasized conduction as their essential role – neurons enable activity in one part of an organism to be rapidly made relevant to the activities of another distant part (Parker 1919). But it became clear that conduction is only a small part of what neurons do, and probably did. Neurons do not just propagate activity from one place to another; they *create* new patterns and activity. They do this both in ways that admit information-processing description (inference, filtering) and ways that may not. This takes the system away from a clear sender-receiver pattern, in ways that make adaptive sense. Nevertheless, there remains a strong temptation to describe the system in ways that really apply only to clearer, paradigmatic cases. This, we believe, sheds light on certain foundational disputes in cognitive science regarding mental representation and cognitive architecture. Anti-representationalists emphasize disanalogies between supposed internal representations and paradigm cases of intermediaries in send-receive systems, and view talk of "representations" as leading to illusions of explanation. Representationalists, on the other hand, insist that the representational framework remains the best one available, and can be used in a way mindful of disanalogies between internal representations and familiar external cases.

An organism can be divided into *stages* as well as spatial parts, and earlier stages often have access to information that will be useful in guiding the actions of later stages. This situation also fits a sender-receiver structure: one can see *memory* as communication between stages. An earlier stage can leave marks that are interpreted and used by later stages. Memory in some respects lies *between* interpersonal and intrapersonal signaling, because a person at any time has a complete psychological profile, a set of beliefs, desires, etc., of their own. The spatial parts of an organism do not. Moreover, the preferences of the different stages of a person can conflict in ways for which there are no clear analogues in the between-parts case, a fact motivating the use of "commitment devices" (Ainslie 2001).

The fact that stages function in some ways as whole agents makes it possible for some forms of memory to have an overtly communicative character, utilizing external media such as notes, computer files, and *ad hoc* arrangements of objects intended to appear salient and suggestive at later times. A controversial topic in cognitive science is whether on-board, psychological memory makes use of write-read cycles that have significant similarity to external mnemonic technologies, or whether neural mechanisms, such as long-term potentiation at synapses, work in ways that lack clear "write" and "read" processes, and hence depart further from a sender-receiver configuration (Gallistel and King 2009, Donohoe 2010, Godfrey-Smith 2014a).

Earlier we set out a taxonomy of signs. This taxonomy can be applied within the organism as well, and in both dimensions. At one extreme are representations that are atomic and nominal; at the other, mental representations that are combinatorial and encoding. The latter sort have syntax in a strong sense and can be appropriately conceived of making up a “language of thought.”[11] Between these extremes lies a variety of other forms mental representations which have only recently begun to feature in cognitive science (e.g., Danks 2014). As in the between-agent case, this gives rise to a family of research questions: Which sorts of factors can drive, and which can constrain, evolutionary transitions from one type of representational system to another? What are

[11] An internal sign system can count as a language of thought for us but only remotely resemble an actual natural language (it might be devoid of logical connectives, tense, etc.). This is consistent with Jerry Fodor’s original conception of a language of thought (Fodor, 1975). However, Fodor often had a richer conception in mind, as seen in his views about language learning and comprehension. Fodor held that such phenomena can be explained only if we possess a language of thought that is equally expressively powerful.

the most plausible evolutionary pathways via which more elaborate representational systems in the brain emerge?

Sign systems in the brain are generally used not just for the transmission of information, but also for computation. As the set of states one wishes to represent increases in size, there are well-known economic benefits to the use of complex (composite, combinatorial) signs as opposed to atomic ones. However, when computation is part of the picture, the organizational properties of a sign system become critical as well. To see this, consider an animal that is able to compute, for any two food items it has cached, which of those items was cached more recently. If an organized sign system is used, this computation might be trivial, and hence impose minimal physical demands on the animal's brain. Suppose, for example, that the animal uses the connection strength among one pair of neurons (an "A pair") to represent the elapsed time since item A was cached, the connection strength between another pair of neurons (a "B pair") to represent the elapsed time since item B was cached, and so on, and that the relation among signs *stronger_connection_than* maps to the relation among items *cached_more_recently_than*. Then, for any two items, all the animal has to do to determine which item was cached more recently is compare the connection strengths among the relevant pairs of neurons (e.g., to determine whether A was cached more recently than B, it need only compare the strength among the A pair and the strength among the B pair). Matters are different if a nominal sign system is used, however. To compute this same relation, the animal would have to know in advance which of the two signs being compared—which of the two connection strengths—represented the longer duration, and know this for every possible pair of signs it might token. This is because there is no way to extract this information from one's signs with a nominal sign system.

Between-agent sign use functions primarily in transmission. This is not so in the within-the-organism case. Internal signs are routinely combined and processed so as to yield new information that is useful for the control of behavior. In general, they are *for* carrying out computations of various sorts at least as much as they are *for* transmitting information. We think this facilitates the evolution of highly organized sign systems, and see a co-evolutionary feedback loop here. As an organism's internal signs become more organized, computation becomes more efficient, and as computation becomes more efficient, it "pays" to represent more information, more states of the world, because one can get more out of them. This point applies equally to communication between parts and between stages. It is sometimes taken to be a puzzle why syntactically-structured signs, or even proto-syntactically structured signs for that matter, figure so rarely in non-human communication systems.[12] We think part of the answer lies in the function of these systems. If we broaden our focus to include the within-organism context, such sign systems might not be rare at all.

5. Meaning

The models treat sender-receiver coadaptation, constructed broadly, as the core property of communication; the way signs are produced is sensitive to and shaped by the way

[12] See, for example, Scott-Phillips and Blythe (2013) and Cloud (2014).

signs are interpreted, and vice-versa. Against this background, we now look at a long-standing divide among theories of meaning ("meta-semantic" theories). We see these theories as tending to fall into two traditions or styles of explanation, namely, *expressive* and *interpretative traditions*.

> *Expressive tradition*: The basis of meaning is what senders are trying to get across to a receiver, or trying to get them to do.

> *Interpretative tradition*: The basis of meaning is what receivers make of a sign, of how they interpret it, or what they get from it.

A vivid example of this divide in the philosophy of language can be seen in the work of Paul Grice and Donald Davidson. Grice gave a theory of how speakers use language to express ideas and to achieve things, where an important theoretical category is *what is said* using some bit of language. This idea was developed differently by neo-Griceans and Relevance Theorists, though in both cases, the focus remained on the speaker. The opposite tendency is illustrated by Davidson; he gave a view of language based on the properties of *good interpretation*, of how an interpreter might *make sense* of a stream of speech.[13]

---

[13] Here we refer to Davidson's classic work in the 1960s and 1970s (collected in Davidson 1984). As a referee pointed out, some of Davidson's later work (Davidson 1992) defends an approach in some ways akin to the view defended here.

We think that each of these traditions, with their tendency to privilege one side of communication, gives rise to a skewed treatment of meaning. The Wittgenstein of the *Philosophical Investigations* (1953), in contrast, did emphasize the two-sided nature of the *use* of signs, but his treatment was anti-theoretical and lacking in positive detail. Sender-receiver models make this notion of use central. But what, in more positive terms, can the models teach us in this area? Let us start out by thinking about the attribution of meaning to signs in simple Lewis-style systems. What properties do the signs themselves possess?

One way to approach this question is to invoke the concept of information in Shannon's sense. Useful signs are ones that have reliable associations with behaviorally relevant states of the world, and hence *reduce uncertainty* about those states. In other words, they *carry information* about those states. In philosophy, Fred Dretske and Skyrms have been prominent defenders of this sort of explanation.[14] In a familiar Revere-style case, a message carries the information that the British are coming by land because it raises the probability of that event to 1. Refinements may be added to make sense of the idea of error or false content. On this sort of view, semantic content is a richer relative of information content.

[14] See, especially, Dretske (1981) and Skyrms (2010).

This is not the only way to recognize a relation of involvement between sign and world that amounts to a kind of meaning. Rather than looking "backward" to the typical cause of a sign, we can look "forward" to the conditions under which actions prompted by the sign will be successful, or will help stabilize the sender-receiver system. This is the approach taken in the teleosemantic theories of Millikan (1984, 2004), David Papineau (1984), and others. Success is connected to content via the idea of a *biological function*, which is in turn understood in terms of an evolutionary history of selection. A *truth condition* of a sign is a historically normal success condition for the use of the sign, where this is the condition under which actions based on the sign were successful and led to the maintenance of that pattern of sign use.

How might we decide between this approach and the information-based approach? In the simplest cases, the two approaches gain equal traction. The two kinds of content *coincide*. Let us assume that Revere and the Sexton are at their familiar signaling equilibrium. It seems natural to say that one lantern *means* that the British are coming by land.[15] This is the information content of the sign: one lantern raises the probability that the British are coming by land to 1. But it is *also* the success condition of actions based on the sign (i.e., preparing for a land invasion), given the receiver's rule that has been stabilized. In simple cases, often (though not always) the selection process operating will bring the two kinds of "content" into alignment with one another.

[15] Here we set aside any imperative content the signal might have.

Once we move away from the simplest cases, however, the two kinds of content come apart. Perhaps the sender cannot discriminate all of the states that matter, and hence cannot send distinctive signs in every state that calls for a different act. For example, suppose that the sender (a sensory system) cannot discriminate nutritious flies from harmless flying specks. The situation can be seen as one where the sender "pools" these states, treating them as equivalent, and the receiver produces an act that works best as a cover-all across. Then, as discussed in the naturalistic literature on content in the 1980s, the informational content will be a disjunction of all the states treated as equivalent by the sender, while the success condition of actions based on the sign will be logically stronger, as success is associated only with a subset of the states pooled by the sender (those containing a fly). The informational content might be *fly or speck* while the success condition or teleological content will be *fly*. In other situations, the information-based and success-based interpretations come apart in other ways. For example, when there is only partial common interest between the two agents, the success condition for a sign, given the receiver's rule, is different for the sender than it is for the receiver, though the informational content will be unaffected.[16]

One response to this situation is to try to find a combination of informational properties, action-related properties, and perhaps other factors, that delivers an intuitively correct meaning ascription across all of the cases. Another response is to say these cases show that there are various different kinds of *involvement* that a sign can have with a state

[16] See Shea, Godfrey-Smith, and Cao (2017) for a discussion of this issue.

of the world, where these forms of involvement may be routed through features of the sender, the receiver, or both. In simple cases, the same state of the world can be associated with a sign by various different routes. In other cases, this convergence or concurrence will not occur. This is not a problem that needs to be solved by choosing one kind of involvement between sign and world and saying it is *the* basis of content. Instead, various kinds of sign-world relationship can be recognized.

Here we have chosen some simple naturalistic ways in which signs can be associated with world states, but the conclusion to be drawn applies more generally. We see sender-receiver models as supporting a rethinking of the project of philosophical semantics. There is a partial deflation of that project, and in part a move towards pluralism or contextualism. For any sign, there is a totality of facts about how and why the sign is used by sender and receiver in the way that it is, and how and why these usage patterns have come to exist. This includes facts about the internal structure and/or organizational properties of signs, and also associations between signs, acts, and states of the world. When some sign-world relation is especially conspicuous or salient in a given context, we will often be inclined to treat it as determinative of *the meaning* of the sign. This is understandable. However, that same fact might be less important in another context, and a fact of the same type might be less important in the consideration of another case. As Lewis noted in his original 1969 discussion, once we have described all the facts about the sender's and receiver's policies, what they achieve, and why they exist, we have explained all there is explain. All of the facts there might be about "meaning" are determined by those facts about the sender's and receiver's policies, what they achieve, and why they exist, but it is not necessary to localize the basis of meaning further than this. A variety of features of signs are then semantically significant, and appropriately picked out in particular contexts of interpretation. But it is an error to try to work out which of these features of a sign is *the* meaning – the content, what it says.

The role of the sender-receiver structure as a magnet for interpretative habits is also relevant at this point. As noted earlier, a complete treatment of communication and meaning should also cover the ways in which we are disposed to talk about signs from the "outside," as both commonsense onlookers and as theorists. Upon observing a sender-receiver interaction, we are often led to form judgements about the meaning of the signs involved. This may have downstream effects on the way signs are produced and interpreted *in* communication. But we also talk with one another about the meaning of signs in a general sense; e.g., "schnee" means *snow*. This kind of meta-semantic discourse can create the impression that there is a definite *thing* that each sign means, but that conclusion does not follow.

6. Human Communication and Language

We now turn to human communication and language. Kim Sterelny (2017) has distinguished two challenges to incrementalist theories of language evolution: syntactic and semantic. The syntactic challenge is to show how human language syntax might have emerged gradually via a series of small, incremental steps. Chomsky and colleagues[17]

[17] See, especially, Berwick and Chomsky (2015).

have been the most ardent supporters of anti-incrementalism about syntax. We believe our account of syntax can be used to mount a response to this challenge, but that is a task for another day. Here we use the sender-receiver framework to offer a response to the semantic challenge. The essence of this challenge is to show how semantic phenomena distinctive of human linguistic communication might have evolved gradually out of those associated with simpler systems of animal communication.

The semantic challenge has been forcefully raised by Scott-Phillips (2014). Following Sperber and Deidre Wilson (1986), Scott-Phillips claims that there are just two possible forms of communication. One is *coded communication*. Coded communication is said to be made possible by *associations* between states and signs, on the one hand, and signs and acts, on the other. It is the home of "natural meaning," according to Scott-Phillips. The other is *ostensive-inferential communication* ("inferential communication" for short). Inferential communication consists in the expression and recognition of communicative intentions. A communicative intention is an intention that one's audience recognize that one has some *informative intention*, an intention to change the mental states of one's audience in some way. The story here is broadly Gricean, including the idea that "speaker meaning" is determined by the content of one's informative intention.[18] This kind of communication is said to be made possible by meta-psychology, the ability to attribute mental states to others.

[18] In the context of Relevance Theory (Sperber and Wilson, 1986), these two intentions are recast in more technical terms.

Scott-Phillips and others (Origgi and Sperber 2000; Tomasello 2010, 2014) argue that language evolved out of ancestral hominins' capacity for inferential communication. This is a natural view to hold if one thinks, as these theorists do, that only humans satisfy the meta-psychological demands for such communication. Those demands include the ability to process high orders of intentionality, a robust understanding of intentional agency, and tracking of "common ground."[19] While there are different versions of this view, the core idea is that advanced meta-psychological abilities evolved in response to the increasingly intense social demands of hominin life over the last 2-0.5 million years. These abilities made inferential communication possible, with language following in its wake as a tool for increasing the efficiency and precision of such communication. Only hominins evolved language, then, because only hominins evolved the capacity for inferential communication.

On this view, linguistic communication is treated as special case of inferential communication: it is inferential communication that makes use of a rich, conventional code (a natural language). The main alternative to this picture is one on which there was a gradual transition from coded communication to linguistic communication. Put

[19] Richard Moore (2014) argues that the meta-psychological requirements of inferential or "Gricean" communication have been overstated. If one accepts his deflationary account, this changes the cogency of the above move, as great apes are then plausibly seen as capable of such communication.

differently, one might hold that linguistic communication evolved via a series of small steps where, to use Scott-Phillips' language, "[t]he starting point [was] a coded communication system, with a fixed and limited set of signals, and the end point [was] a system of ostensive-inferential communication, made possible by a large set of conventional codes" (2014, p. 45). Scott-Phillips takes issue with this view for several reasons; here we focus on one. This is that the view suffers from a kind of conceptual confusion, or at least a lack of clarity. Scott-Phillips says, "it's not even clear what continuity *could even mean here …*" (*ibid.*, p. 46) [italics ours].

While we agree there are differences between paradigm cases of coded and inferential communication, we doubt that a sharp distinction can be drawn between the two categories. Meta-psychology can take on a striking variety of forms, and consequently, inform communication in a variety of ways. One could recover a strict dichotomy here by simply defining "coded communication" as any type of communication that falls short of inferential communication. That would turn Scott-Phillips' claim into a truism, but more importantly, it would conceal a lot of structure within the coded category. The psychological preconditions for inferential communication have partial forms, and there is a natural evolutionary trajectory from coded to inferential communication. As increasingly sophisticated meta-psychology appears and begins to play a role in how senders and receivers use signs, there is a shading of coded communication into the inferential.

As we have emphasized, sender-receiver models do not require sophisticated cognition but nor do they preclude it. When the sender and receiver are intentional systems, they can have a richer or poorer understanding of each other *as* intentional systems. Experimental research over the last two decades suggests that human "theory of mind" is a cognitive mosaic, a mix of different tracking and inferential abilities with separate evolutionary histories (and in some cases, cultural evolutionary histories). It is now widely accepted, for example, that chimpanzees can attribute goal states and states of visual perception to others.[20] In children, a suite of "low-level" or "implicit" mindreading abilities come on-line early, allowing them to make certain behavioral predictions in a limited range of circumstances. This is followed by the development of "explicit" abilities to recognize goals, perceptual states, thoughts, and emotions in the second and third year. Only later do children develop explicit belief tracking, higher-order intentionality, and adult-like understanding of functional relations among mental states and among mental states and behavior.[21] So, the menu of mental states and relations that an agent tracks can be richer or poorer. Moreover, these states can be tracked more or less *robustly*; the agent can make use of a larger or smaller set of cues in recognizing their presence (Sterelny 2003). Agents may also exhibit greater or lesser *response breadth* in the face of such information; they may be able to *do* more or less, behaviorally speaking, with that information (*ibid.*). When there is some understanding

---

[20] For evidence of goal attribution, see Uller and Nichols 2000; Tomasello et al. 2003. For evidence of sensitivity to visual states, see Kaminski et al. 2008.

[21] See Apperly (2010) for an overview.

present that the other agent is an intentional system, this might feed into communicative behavior. The questions the models lead us to ask is whether behaviors of sign production or interpretation are influenced by such understanding, and if so, how.

Meta-psychology might play more of a role on one side of a sender-receiver interaction than the other. In principle, individuals might recruit their understanding of intentional agency only when deciding how to react to a sign. Some cases of vocal communication in monkeys might fit this pattern. Robert Seyfarth and Dorthy Cheney (e.g., 2014) have described a wide range of cases where baboons draw sophisticated inferences on the basis of the calls they hear. In work with Anne Engh, they have shown that subordinate females interpret threat-grunts as expressing aggressive intentions towards them if the they have recently fought with the caller, but not if the two have recently groomed (Engh et al. 2006). The converse is possible too. An example might be the deceptive use of an alarm call in a setting where receivers are disposed, at least initially, to respond reflexively to the call. The fork-tailed drongo is an African bird that mimics a large range of different species' alarm calls to steal food from other birds and some mammals (e.g., meerkats). Field studies have shown that drongos flexibly change call types during repeated food-theft attempts, and are more likely to change call type following food-theft failure (leading to increased success) (Flower et al. 2014). The authors note that the drongo's deceptive behavior satisfies the conditions of so-called "tactical deception" (Whiten and Byrne 1988), which is standardly assumed to require some theory of mind abilities. Depending on the details of the case, and especially on how the interests of sender and receiver line up, an asymmetry of either sort could easily drive an increase in meta-cognitive sophistication on the other side; we might, for example, see a trend towards greater "epistemic vigilance" (Sperber et al. 2010) or greater cognitive collaboration. And as communication comes under the control of more powerful meta-psychology, we can expect to see new semantic phenomena and/or intensification of existing ones: increased rates of signal creation as well as novel and more flexible usage of already established signals, for example (Planer 2017).

Once sender and receiver possess advanced meta-psychological abilities, and an unbounded capacity for recruiting these abilities in communication, we see the phenomena distinctive of full-blown inferential communication—improvisation, indirectness, irony, etc. In selecting a sign, a sender can take into account the receiver's thinking about the sender's thinking about the receiver's own thinking (4^{th}-order intentionality). And, correspondingly, the receiver's interpretation can be sensitive to all of this (5^{th}-order intentionality). For example: Hunter A wants hunter B to know that he has already searched some territory, but there is no established convention for doing so; A cuts an "X" on a tree at the foot of the path leading that way; in so doing, A expects B to recognize A's intention to inform him of something, and moreover, expects that B's recognition of this intention, together with general background knowledge and information in common ground, will induce the desired change to B's mental states. B grasps all of this, and the impromptu communication is a success. High-level, cooperative mindreading of this sort is transformative with respect to signal creation, as it greatly expands the range of things that can be communicated and, relatedly, the range of behaviors that can be put to communicative use. Add language, and just about anything a sender and receiver can *think*, they can *communicate*.[22]

In sum: Where Scott-Phillips sees two fundamentally distinct forms of communication in nature separated by a clear boundary, we see a spectrum of sender-receiver systems ranging from ones where meta-psychology plays no role to those where communication is shaped by various forms of strategic and interpretative sophistication. Linguistic communication probably evolved incrementally out of less inferential forms of communication.

7. Conclusion

---

[22] A referee objected that, whereas Relevance Theory *explains* the spontaneous creation of signs in the absence of a code, here we merely *acknowledge the existence* of that phenomenon. We disagree: the sender-receiver framework does offer an explanation, just a more abstract one. Such cases are explained by sender and receiver appropriately adjusting sign production and interpretation behavior in light of the envisaged behavior on the other side. However, the main point of this section is to show how the sender-receiver framework can be used to undermine Scott-Phillips' anti-incrementalist challenge about the evolution of inferential communication, and one can agree with us about that even if one thinks we owe a fuller treatment of sign creation in the absence of a code.

The desire for an overarching theory of signs has been a longstanding one, especially clear in the semiological tradition associated with Ferdinand de Saussure. A central error of that tradition was to theorize about structure in a way largely unconstrained by sender-receiver interaction and hence the selection processes that shape sign use. This error permeated not just early forms of semiological analysis but also the more sophisticated forms of structuralist semiotics that came later. Post-structuralist semiotics followed, which more or less abandoned the aim of theoretical understanding in the usual sense. We suspect that this failure of structuralist semiotics to give a more direct role to sender and receiver use was pivotal to its demise, though either way, that failure remains.

This history has led some to think that there is no general theory of signs to be had (e.g., Sperber and Wilson [1986]). We are more optimistic. There has been a tendency on the part of general theories in this area to turn every *prima facie* case of sign use into a *paradigm* one. This is most evident in structuralist semiotics' search for the "grammatical principles" underlying every sign system, thereby revealing their likeness to human language. A general theory of signs must instead recognize that there are partial and marginal cases, and provide us with an understanding of *why* those cases count as partial or marginal, rather than attempting to shoehorn them into a paradigm form.

The sender-receiver framework unifies the many literatures on how and why communication arises and stabilizes. Here we have presented the framework in an ambitious form, treating it as identifying *the* basis for semantic phenomena. The usefulness of the framework does not require that it be construed in this way, however. It might be used alongside other theoretical ideas – perhaps alongside an explanation of some basic semantic relationships in different terms (a possibility pointed out by a referee). The general outlook – the focus on interlocking behaviors of sign production and sign interpretation – can be fruitfully employed whether one holds that communication and representation inhere in sender-receiver coordination, or construes the framework differently.

References

Ainslie, George. *Breakdown of will*. Cambridge University Press, 2001.

Apperly, Ian. *Mindreaders: the cognitive basis of" theory of mind"*. Psychology Press, 2010.

Barrett, Jeffrey A. "The evolution of coding in signaling games." *Theory and Decision* 67, no. 2 (2009): 223-237.

Barrett, Justin. *Why Would Anyone Believe in God? (Cognitive Science of Religion)*. Altamira: Walnut Creek, 2004.

Berwick, Robert C., and Noam Chomsky. *Why only us: Language and evolution*. MIT press, 2015.

Cloud, Daniel. *The domestication of language: Cultural evolution and the uniqueness of the human animal*. Columbia University Press, 2014.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 1431-1451.

Dawkins, Richard, and John Richard Krebs. "Arms races between and within species." *Proc. R. Soc. Lond. B* 205, no. 1161 (1979): 489-511.

Danks, David. *Unifying the mind: Cognitive representations as graphical models*. Mit Press, 2014.

Davidson, Donald. *Inquiries into truth and interpretation: Philosophical essays*. Vol. 2. Oxford University Press, 1984.

Davidson, Donald. "The second person." *Midwest Studies in Philosophy* 17.1 (1992): 255-267.

Donohoe, John. "Man as Machine: A Review of Memory and the Computational Brain: Why Cognitive Science will Transform Neuroscience, by CR Gallistel and AP King." *Behavior and Philosophy*, no. 38 (2010): 83-101.

Dretske, Fred. "Knowledge and the Flow of Information." MIT Press, 1981.

Engh, Anne L., Rebekah R. Hoffmeier, Dorothy L. Cheney, and Robert M. Seyfarth. "Who, me? Can baboons infer the target of vocalizations?." *Animal Behaviour* 71, no. 2 (2006): 381-387.

Flower, Tom P., Matthew Gribble, and Amanda R. Ridley. "Deception by flexible alarm mimicry in an African bird." *Science* 344, no. 6183 (2014): 513-516.

Fodor, Jerry A. *The language of thought*. Vol. 5. Harvard University Press, 1975.

Franke, Michael. "Creative compositionality from reinforcement learning in signaling games." *Evolution of Language: Proceedings of the 10th International Conference (EVOLANG10)*. 2014.

Gallistel, Charles R., and Adam Philip King. *Memory and the computational brain: Why cognitive science will transform neuroscience*. Vol. 6. John Wiley & Sons, 2009.

Godfrey-Smith, Peter. "Sender-receiver systems within and between organisms." *Philosophy of Science* 81, no. 5 (2014a): 866-878.

Godfrey-Smith, Peter. "Signs and symbolic behavior." *Biological Theory* 9, no. 1 (2014b): 78-88.

Godfrey-Smith, Peter. "Senders, receivers, and symbolic artifacts." *Biological Theory* 12, no. 4 (2017): 275-286.

Godfrey-Smith, Peter, and Manolo Martínez. "Communication and common interest." *PLoS computational biology* 9, no. 11 (2013): e1003282.

Gray, Kurt, and Daniel M. Wegner. "Blaming God for our pain: Human suffering and the divine mind." *Personality and Social Psychology Review* 14, no. 1 (2010): 7-16.

Grice, H. Paul. "Logic and conversation." In Syntax and Semantics, Vol. 3, Speech Acts, ed. by Peter Cole and Jerry L. Morgan. New York: Academic Press 1975, 41–58; here 45–47.

Harms, William F. "Primitive content, translation, and the emergence of meaning in animal communication." *Evolution of communication systems: A comparative approach* (2004): 31-48.

Harms, W. F. (2010). Determining truth conditions in signaling games. *Philosophical studies*, *147*(1), 23.

Huttegger, Simon M. "Evolution and the explanation of meaning." *Philosophy of Science* 74, no. 1 (2007): 1-27.

Kaminski, Juliane, Josep Call, and Michael Tomasello. "Chimpanzees know what others know, but not what they believe." *Cognition* 109, no. 2 (2008): 224-234.

Kirby, Simon. *Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners*. na, 2000.

Levy, Arnon. "Information in Biology: A Fictionalist Account 1." *Noûs* 45, no. 4 (2011): 640-657.

Lewis, David. *Convention: A philosophical study*. Harvard University Press 1969.

Millikan, Ruth Garrett. *Language, thought, and other biological categories: New foundations for realism*. MIT press, 1984.

Millikan, Ruth Garrett, and Ruth G. Millikan. *Varieties of meaning: the 2002 Jean Nicod lectures*. MIT press, 2004.

Moore, Richard. "Ontogenetic constraints on Grice's theory of communication." *Pragmatic development in first language* (2014): 87-104.

Nowak, Martin A., David C. Krakauer, and Andreas Dress. "An error limit for the evolution of language." *Proceedings of the Royal Society of London. Series B: Biological Sciences*266.1433 (1999): 2131-2136.

Nowak, Martin A., Joshua B. Plotkin, and Vincent AA Jansen. "The evolution of syntactic communication." *Nature* 404.6777 (2000): 495.

O'Connor, Cailin. "The evolution of vagueness." *Erkenntnis* 79, no. 4 (2014): 707-727.

Origgi, Gloria, and Dan Sperber. "Evolution, communication and the proper function of language." *Evolution and the human mind: Language, modularity and social cognition*(2000): 140-169.

Owren, Michael J., Drew Rendall, and Michael J. Ryan. "Redefining animal signaling: influence versus information in communication." *Biology & Philosophy* 25, no. 5 (2010): 755-780.

Parker, George Howard. *The elementary nervous system*. Vol. 1920. JB Lippincott, 1919.

Papineau, David. "Representation and explanation." *Philosophy of Science* 51, no. 4 (1984): 550-572.

Planer, Ronald J. "Protolanguage might have evolved before ostensive communication." *Biological Theory* 12, no. 2 (2017): 72-84.

Rescorla, Michael. "Convention." The Stanford Encylopedia of Philosophy (Summer 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2017/entries/convention/>.

Scott-Phillips, Thom. *Speaking Our Minds: Why human communication is different, and how language evolved to make it special*. Macmillan International Higher Education, 2014.

Scott-Phillips, Thomas C., and Richard A. Blythe. "Why is combinatorial communication rare in the natural world, and why is language an exception to this trend?." *Journal of the Royal Society Interface* 10, no. 88 (2013): 20130520.

Searcy, William A., and Stephen Nowicki. *The evolution of animal communication: reliability and deception in signaling systems*. Princeton University Press, 2005.

Seyfarth, Robert M., and Dorothy L. Cheney. "The evolution of language from social cognition." *Current opinion in neurobiology* 28 (2014): 5-9.

Shannon, Claude (1948). "A mathematical theory of communication." *The Bell System Mathematical Journal* 27: 379-423, 623–656.

Shea, Nicholas, Peter Godfrey-Smith, and Rosa Cao. "Content in simple signalling systems." *The British Journal for the Philosophy of Science* (2017).

Skyrms, Brian. *Evolution of the social contract*. Cambridge University Press, 1996.

Skyrms, Brian. *The stag hunt and the evolution of social structure*. Cambridge University Press, 2004.

Skyrms, Brian. *Signals: Evolution, learning, and information*. Oxford University Press, 2010.

Smead, Rory. "Deception and the evolution of plasticity." *Philosophy of Science* 81, no. 5 (2014): 852-865.

Smith, John Maynard, and David Harper. *Animal signals*. Oxford University Press, 2003.

Sperber, Dan, and Deirdre Wilson. "Relevance: Cognition and communication." Blackwell, (1986).

Sperber, Dan, Fabrice Clément, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson. "Epistemic vigilance." *Mind & Language* 25, no. 4 (2010): 359-393.

Sterelny, Kim. *Thought in a hostile world: The evolution of human cognition*. Wiley-Blackwell (2003)

Sterelny, Kim. "From code to speaker meaning." *Biology & Philosophy* 32, no. 6 (2017): 819-838.

Tomasello, Michael. *Origins of human communication*. MIT press, 2010.

Tomasello, Michael. *A natural history of human thinking*. Harvard University Press, 2014.

Tomasello, Michael, Josep Call, and Brian Hare. "Chimpanzees understand psychological states–the question is which ones and to what extent." *Trends in cognitive sciences* 7, no. 4 (2003): 153-156.

Tria, Francesca, Bruno Galantucci, and Vittorio Loreto. "Naming a structured world: a cultural route to duality of patterning." *PLOS one* 7.6 (2012): e37744.

Uller, Claudia, and Shaun Nichols. "Goal attribution in chimpanzees." *Cognition* 76, no. 2 (2000): B27-B34.

Wagner, Elliott O. "Deterministic chaos and the evolution of meaning." *The British Journal for the Philosophy of Science*63, no. 3 (2011): 547-575.

Whiten, Andrew, and Richard W. Byrne. "Tactical deception in primates." *Behavioral and brain sciences* 11, no. 2 (1988): 233-244.

Wittgenstein, Ludwig. "Philosophical investigations." Basic Blackwell, 1953.

Zollman, Kevin JS. "Separating directives and assertions using simple signaling games." *The Journal of philosophy* 108, no. 3 (2011): 158-169.