SEPARATING DIRECTIVES AND ASSERTIONS USING SIMPLE SIGNALING GAMES*

Most contemporary accounts of meaning involve, in one way or another, the intentions of the speaker and/or listener. In contrast, David Lewis develops an account of meaning in *Convention* which does not require reference to intentional states.[1] He instead defines meaning as exclusively a feature of a social group—it arises when speakers and listeners behave in mutually compatible ways in an appropriately defined game. Because intentions can be left out of this account, meaning can be present in any group which can arrange itself appropriately, including many animal species. Since many examples of animal signals appear to miss a variety of features considered central to human languages, some regard this as a reductio of Lewis's account. Others have suggested that the meaning present in animal signals (and in Lewis's toy model) represents "primitive content" which can be further refined in various ways to arrive at meaning in human language.[2] Some of these refinement accounts attempt to make intentionality a derivative notion which emerges from language (or alongside it), rather than a requirement for it.[3]

These projects which refine Lewis's definition of meaning must address features of human language that cannot be captured in Lewis's simple game. Here I focus on one shortcoming of Lewis's account, its failure to distinguish between directives and assertions.[4] I present a new Lewis-style game which accounts for this distinction and thus demonstrates how, in this case, Lewis's strategy might be pushed forward to a more complete account of human language.

* The author would like to thank Bill Harms, Simon Huttegger, Mandy Simons, and Brian Skyrms for their helpful comments on earlier drafts. This paper grew out of an interesting conversation with Kevin Kelly and Peter Kriss.

[1] David Lewis, *Convention: A Philosophical Study* (Cambridge: Harvard, 1969).

[2] William F. Harms, *Information and Meaning in Evolutionary Processes* (New York: Cambridge, 2004), and "Primitive Content, Translation, and the Emergence of Meaning in Animal Communication," in D. Kimbrough Oller and Ulrike Griebel, eds., *Evolution of Communication Systems: A Comparative Approach* (Cambridge: MIT, 2004), pp. 31–48; Ruth Garrett Millikan, *Language: A Biological Model* (New York: Oxford, 2005).

[3] See Millikan, *op. cit.*

[4] John R. Searle, "A Taxonomy of Illocutionary Acts," in *Expression and Meaning: Studies in the Theory of Speech Acts* (New York: Cambridge, 1979), pp. 1–29.

0022-362X/11/0803/158–169

I. LEWIS SIGNALING GAMES

Central to Lewis's account of meaning is the definition of a sort of language game, which has come to be known as a *Lewis signaling game*. The story of Paul Revere serves as a motivating example. The sexton of the Old North Church and Revere face a coordination problem. The sexton can observe whence the British troops arrive but can do nothing about it. Revere can inform the city but cannot observe the troops. Seeing that together they can achieve an outcome they both desire, they agree on a convention: the sexton will hang one lantern if the British come by land and two if they come by sea. In this circumstance, few would disagree that *one lantern* means the British are coming by land and that *two lanterns* means they are coming by sea.

Generalizing this story, Lewis describes a simple game which he believes captures an important part of linguistic behavior. In this game there are two players, one named *the sender* and one *the receiver*. The sender is aware of some aspect of the world that the receiver is not, and she has at her disposal a set of signals which she can send to the receiver, who must then take some action. The sender and the receiver both benefit if the receiver takes an action which is appropriate to the state, and the appropriate act differs from state to state.

In the game's simplest form, there are two relevant states of the world, two signals, and two actions that the receiver can take. *Action one* (A1) is appropriate for *state one* (S1), and *action two* (A2) is appropriate for *state two* (S2). The signals are labeled Sig A and Sig B. We will also suppose for simplicity that both the sender and the receiver win when the appropriate action is taken and lose when an inappropriate one is taken.

This situation describes a simple game, which has many different Nash equilibria.[5] Of particular interest are the Nash equilibria dubbed by Lewis as *signaling systems*. In the simple two-state, two-signal, two-act game there are two such equilibria. One occurs when the sender sends Sig A in S1 and Sig B in S2 and the receiver performs the appropriate acts in response to the signals. The other signaling system permutes the signals for both the sender and receiver. These equilibria are notable because they are *Pareto superior*—both players fare better in these equilibria than in any others—and *strict*—each player would do strictly worse by unilaterally deviating.

Even if we suppose that individuals will always coordinate on Pareto superior or strict equilibria (a very strong assumption), there are two

[5] A set of strategies (contingency plans over all possible situations) constitutes a Nash equilibrium if and only if no player would do strictly better by changing her strategy while the other players keep their strategies fixed.

choices, and there remains a question of how individuals can come to coordinate on one signaling system rather than the other. Notice that if they disagree about which signaling system is in force they will completely miscoordinate and always lose. Lewis suggests that natural salience or prior agreement (as in the case with Revere) may contribute to coordination on a signaling system.

More recently, several scholars have focused on how individuals might come to play these strategies without the aid of prior agreement or salience.[6] In these models prior agreement is replaced by an evolutionary or learning process where individuals can come to coordinate on a signaling system with very weak assumptions about their underlying cognitive ability.

Lewis argues that in signaling system equilibria the signals have acquired meaning.

> I have now described the character of a case of signaling without mentioning the meaning of the signals; that two lanterns meant that the redcoats were coming by sea, or whatever. But nothing important seems to have been left unsaid, so what has been said must somehow imply that the signals have their meanings.[7]

Of course, intentions may often be involved in the meaning of signals even for Lewis. Intentions can shape the emergence of a signaling system and can be involved in its being sustained. However, one might come to these equilibria in many ways which do not involve intentional decisions at all—this is a central feature of the evolutionary accounts. These evolutionary accounts offer a sort of proof of possibility that much of human-like meaning can be had without an appeal to intentions. The evolutionary models, which describe how individuals come to play signaling systems, are akin to showing how meaning could emerge from a context where nothing is meaningful.[8]

Lewis's account opens up the possibility that nonhuman signaling systems might also be endowed with meaning. For instance, baby birds often signal their mothers or fathers of their hunger by begging. Extensive experimental work has been done, and it has been determined that in many species the chicks alter the amount or degree of begging dependent on their hunger, and parents determine how

[6] For a review, see Simon Huttegger and Kevin J. S. Zollman, "Signaling Games: Dynamics of Evolution and Learning," in Anton Benz, ed., *Language, Games, and Evolution* (New York: Springer, 2011).

[7] Lewis, *op. cit.*, pp. 124–25.

[8] Harms, *op. cit.*; Brian Skyrms, *Evolution of the Social Contract* (New York: Cambridge, 1996), and *Signals: Evolution, Learning and Information* (New York: Oxford, 2010).

much food to give a chick based on the amount of begging.[9] As a result, we can describe a game—similar to Lewis's—where the chicks and the parent are in equilibrium.[10] As a result, a theory of meaning which only requires that the players be in equilibrium would entail that the chick's begging means that the chick is hungry.

Many have found this conclusion unsettling. Human languages are of a different kind than animal languages, and meaning is uniquely a feature of our languages. As a result, scholars have attempted to point to the feature of human language which simultaneously differentiates it from animal language and which endows human language with meaning. Since many examples of human language use involve intentions, intentionality is often invoked to distinguish human from animal language. A naked appeal to intentionality is insufficient, however, since those who defend the meaningfulness of animal language are challenging exactly this claim—that meaning is essentially intentional. Instead, one must illustrate what, other than intentions, is left out of this nonintentional account of meaning. There are many potential avenues, each of which deserves its own discussion. We will here focus on the distinction between directives and assertions, which are often marked in human languages by the indicative or imperative mood.

II. ILLOCUTIONARY FORCE

Sender	Receiver
S1 → Sig A	Sig A → A1
S2 → Sig B	Sig B → A2

Table 1: A signaling system

Lewis and those after him have recognized the inadequacy of Lewis signaling games in explaining all the nuances of linguistic behavior. One can capture this shortcoming by considering how one might translate the signals of Table 1. Sig A might be translated as "we are now in S1," or alternatively as "you should do A1."[11]

These two English translations might lead one to object that meaning is not present in these signals because they do not capture a

[9] William A. Searcy and Stephen Nowicki, *The Evolution of Animal Communication* (Princeton: University Press, 2005).

[10] There is some disagreement over exactly what game is being played and exactly what the equilibrium is, but for our purposes such details can be ignored. For a discussion see *ibid.* and John Maynard Smith and David Harper, *Animal Signals* (New York: Oxford, 2003).

[11] See Harms, "Primitive Content"; Millikan, *op. cit.*

fundamental distinction. More modestly, one might concede that something like meaning is there, but understanding the emergence of this meaning is unlikely to help us understand how terms in our languages come to have the meanings they do.

The two English translations syntactically differ—one is an indicative and the other an imperative—but the concern can also be captured by considering the different illocutionary forces which this syntactic distinction usually marks. The sexton might want a single lantern to simply inform Revere about the British, or he might want a single lantern to cause Revere to ride to Lexington yelling "The British are coming by land!" The former has the force of an assertion, transmitting information, while the latter has the force of a directive, encouraging another to act in a certain way. Searle describes this difference as a difference in "direction of fit."[12] Assertions are meant to fit some preexisting fact in the world, while directives are meant to cause the world to fit the sentence. Searle characterizes this distinction in a thoroughgoing intentional way, but in keeping with the Lewisian strategy we might see if this can be done away with.

Rather than engage intentional meaning, Brian Skyrms suggests that a quantitative notion of information might capture all that is needed to explicate the meaning of the signals in a Lewis signaling game.[13] Consider the signaling-system equilibrium pictured in Table 1. Suppose that prior to hearing a signal or seeing an action the probability of S1 is 0.5. In the signaling system of Table 1, the conditional probability of S1 given that the sender sent Sig A is 1.0. In this sense, the signal carries perfect information about the state. But there is also perfect information about the act of the receiver. Conditioned on the sender sending Sig A, the probability that the receiver will take A1 is also 1.0. Here, a sort of symmetry parallels the two English translations of Sig A: it carries perfect information about both the signal and the act. The central argument of this paper is that, in order to account for the distinction between directives and assertions, we must find a way to break that symmetry, that is, to find a game where (in equilibrium) the signal carries information about the state and not the act or, alternatively, about the act and not the state. If we do this in the right way,

[12] Searle, *op. cit.*

[13] See Skyrms, *Signals*. This is not to say that a quantitative notion of information can always capture the representational content of signals in every context, just in the case of Lewis signaling games. Peter Godfrey-Smith, for instance, has suggested that a notion of information cannot provide a naturalistic foundation for indication generally. See Godfrey-Smith, "Indication and Adaptation," *Synthese*, XCII, 2 (August 1992): 283–312.

we might be able to find the nonintentional equivalent of Searle's "direction of fit."[14]

III. HUTTEGGER'S DISTINCTION

Following a suggestion by Lewis and Harms,[15] Simon Huttegger attempts to model the syntactic difference between indicatives and imperatives as a difference in deliberation.[16] He describes a signaling game where, in addition to the normal acts, both the sender and the receiver must choose whether or not to "deliberate."

Sender	Receiver
S1 → (Do not deliberate, Sig 1)	Sig 1 → (Deliberate, A1)
S2 → (Deliberate, Sig 2)	Sig 2 → (Do not deliberate, A2)

Table 2: Huttegger's signaling system

He considers a situation where one of the players must deliberate in one state and the other must deliberate in the other state. Table 2 shows an example of a signaling system in this game—in one state only the sender deliberates, and in the other only the receiver does. Huttegger suggests that in this equilibrium we should interpret Sig 1 as indicative (or an assertion) and Sig 2 as an imperative (or a directive). This corresponds to one feature of indicatives and imperatives. I inform you that a car is coming when I believe that you need to combine this information with some of your background knowledge or desires in order to determine the correct course of action. I yell "Stop!" when I believe I know the correct course of action, and I believe that any delay on your part would be harmful.

"Deliberation," however, is just a name attributed to some additional action added in Huttegger's model. The appropriate act in S1 is for the receiver to deliberate and then do A1. There is no combination of information or choice of additional action in either case.

[14] Not every way of breaking this symmetry would be sufficient to capture this distinction. Harms (in conversation) gave an insightful counterexample. Suppose that the sender is perfect in her implementation of a signaling system, but the receiver often makes mistakes. There will be an asymmetry in information because the signal carries perfect information about the state but less than perfect information about the act, but it would be odd to call the signal assertive as a result. I take it as necessary, if not sufficient, to break this symmetry in order to account for the difference between assertions and directives. In the case I present below, I break it in one extreme way—by giving either perfect or no information about the act or state. However, symmetry might also be broken in less extreme ways.

[15] Lewis, *op. cit.*; Harms, *op. cit.*

[16] Huttegger, "Evolutionary Explanations of Indicatives and Imperatives," *Erkenntnis*, LXVI, 3 (May 2007): 409–36.

Huttegger leaves this process at a more abstract level, supposing that a refined model could be constructed with similar results.

Although he describes his model as developing a distinction between indicatives and imperatives, Huttegger's model does capture one asymmetry between directives and assertions. Often we direct so that others need not think about what to do, and we assert so that others might decide the best action to take. However, one might object to Huttegger's model by raising the same problem as before. It seems as though signals in Huttegger's model retain the same two translations that were present in Lewis's original game. For instance, Sig A might be translated both as "deliberate and then do A1" and "we are in S1." Huttegger suggests that the latter is a more natural interpretation, but of course the former remains available.

This point can be underscored by appealing to the quantitative notion of information discussed above. As with the original Lewis signaling game, the signal gives perfect information both about the act and about the state. In the signaling system pictured in Table 2, the probability of A1 is 1.0 given the receipt of Sig A, even though this signal is the supposed indicative/assertive signal.

One might extend Huttegger's model in various ways to secure this distinction. For example, one might break A1 into two actions (say, A1.1 and A1.2). The deliberation by the receiver is necessary to determine which of these two actions is correct. Similarly, one might do the same for action A2 but require the sender to deliberate and then send one of two signals. Here, something like informational asymmetry might be constructed in Huttegger's models. As I will explain in the next section, while this is possible, I believe we can adequately model directives and assertions without any appeal to deliberation whatsoever.

IV. A NEW GAME

In order to illustrate the alternative game, I will construct a Lewis signaling game with one extra receiver. Although he has not considered a game like the one presented here, Skyrms has begun to study these types of modifications and the implications they have for the evolution of meaningful signaling systems.[17]

Consider the awkward situation of moving a piece of furniture, like a couch, that requires two people to carry. Often it is helpful to have a third person in position to observe the orientation of the object and direct the movement. That third person might help to level the

[17] Skyrms, *Signals*, and "Evolution of Signaling Systems with Multiple Senders and Receivers," *Philosophical Transactions of the Royal Society of London B*, CCCLXIV (2009): 771–79.

couch either by describing the position of the couch—"the back is too low"—or by directing the actions of the movers—"you on the back, lift up." In the former case, the director often can issue only one statement, leaving it to the movers to position themselves; however, in the latter he must issue two: a directive for each mover. The director would assert in the former case and would direct in the latter.

We will abstract away from cases like this in much the way Lewis did. In our game there will be one sender who is aware of the state of the world and two receivers that must take complementary actions. All three are rewarded only when both receivers take the appropriate actions. Suppose for simplicity that there are only two states, although this assumption is not critical to the discussion that follows. Suppose that the appropriate actions for each receiver are as listed in Table 3.

		States	
		S1	S2
Receivers	R1	A1	A2
	R2	A2	A1

Table 3: Appropriate actions for each state

The receivers must take complementary actions, and the state determines who should do what. In our furniture-moving example, we might imagine that the two states represent the tilt of the couch (either forward or backward). If the couch is tipped forward, the front mover must lift up while the back mover must lower, and vice versa when the couch is tipped backward.

If we allow the sender to differentiate the message that she sends to each receiver, we have a total of four signaling-system equilibria (there are many other equilibria as well). I will call two of these signaling-system equilibria *a-equilibria*. In the a-equilibria the sender sends the same message to both receivers using a different signal in different states. One such equilibrium is pictured in Table 4.

Sender	Receiver 1	Receiver 2
S1 → Sig A (to R1) S1 → Sig A (to R2)	Sig A → A1	Sig A → A2
S2 → Sig B (to R1) S2 → Sig B (to R2)	Sig B → A2	Sig B → A1

Table 4: An a-equilibrium

Consider the informational content of each signal. Like the simple Lewis signaling game discussed above, Sig A (to either receiver)

carries perfect information about the state. What about the act? If we restrict ourselves to considering only one particular receiver (say, Receiver 1), then here too the signal carries perfect information about the act. But if we average over both receivers we find that this information is lost. The probability that a randomly chosen receiver will take A1 given that she is sent Sig A is equal to the unconditional probability that the receiver will take A1. Here we have one example of the informational asymmetry we sought. The asymmetry suggests that we have an assertion because it carries perfect information about the state but no information about the act.

We can generate directives as well. Consider the signaling system pictured in Table 5, which I will call a *d-equilibrium*.

Sender	Receiver 1	Receiver 2
S1 → Sig A (to R1) S1 → Sig B (to R2)	Sig A → A1	Sig A → A1
S2 → Sig B (to R1) S2 → Sig A (to R2)	Sig B → A2	Sig B → A2

Table 5: A d-equilibrium

Here, in both states the sender sends both signals to a receiver but conditions to whom he sends each signal on the state. In this equilibrium it appears as though the signal means something about what the appropriate act is but says nothing about the state. Again, we can consider the informational content of each signal. Globally, there is now perfect information about the act, since the probability that a randomly chosen receiver takes A1 given he received Sig A is 1.0. There is local information about the states here, too. Given that Receiver 1 receives Sig A, the probability that we are in S1 is 1.0. However, if we average over both receivers, that information is lost. Knowing that a randomly chosen receiver has received Sig A gives us no information about the state, despite the fact that it gives us perfect information about the act. Again we have an asymmetry, this time one that suggests the signals are directives.

We have developed an informational asymmetry that corresponds to the distinction between assertions and directives. In a-equilibria there is perfect global information about the state but not the act—the signals are assertions. In d-equilibria there is perfect global information about the act but not the state—the signals are directives.

One might object to this model by noting that it appears to require three-person conversations in order to get off the ground. The distinction between directives and assertions is perfectly coherent in two-person conversations as well. How, then, is this model to account

for that? The three-way conversation described here is merely a heuristic; the game is equivalent to many two-person conversations. Suppose there is just one sender and one receiver. The sender observes the state of the world while the receiver remains ignorant, just as before. But now suppose that the receiver is in one of two "positions," namely, either Receiver 1 or Receiver 2, and suppose that both the sender and receiver can observe her position. The actions and payoffs are just as before, but now there is a single right action which depends on the state and the position of the receiver.

This two-person game is essentially equivalent to the game described above; it has all the same equilibria (as well as the same strategies and payoffs). Now imagine a population of players all observing the same linguistic conventions but having different relevant states and positions. Averaging over that population in an a-equilibrium will result in the signal carrying perfect information about the state and zero information about the action taken by the receiver, and averaging over a population in a d-equilibrium will reverse the role of the signal.

This two-player version suggests other possibilities as well. What if the receiver's position was not known by both players? If the receiver knows her own position but the sender does not, there cannot be any d-equilibria because the relevant strategies are not present.[18] The language here can only have assertions. Given that the sender does not know what the correct action is, how could he issue a directive? This is the situation faced by Cambell's monkeys, who give an alarm call when they spot a predator.[19] Since the right response for a listener depends on the situation that listener is in—and since the signaler is signaling to many different monkeys in different contexts of which the signaler is unaware—it seems proper to interpret the signal as an assertion. We might also imagine another situation where only the sender is aware of the receiver's position, and here only the d-equilibria would be possible. Again this seems plausible; if I have sufficient information to know the right action but you do not, I should issue a directive. Finally, this model might be expanded to consider situations where the position itself is a signal sent by another signaler, by nature, or by the perceptive apparatus of the receiver. Modifications of this sort would be needed to fully account for the evolution of such terms, but I will not consider them in detail here.

[18] I am indebted to Brian Skyrms for this suggestion.

[19] Klaus Zuberbühler, "Predator-Specific Alarm Calls in Campbell's Monkeys, *Cercopithecus campbelli*," *Behavioral Ecology and Sociobiology*, L, 5 (2001): 414–22.

While this model overcomes some of the challenges posed to Huttegger's model, I do not mean to say that this model is right and that Huttegger's is wrong. In fact, features of both models may be present in actual situations of indicative/assertion and imperative/directive use. The point here is not so much to illustrate the one true model, if there is such a thing, but instead to illustrate the strength and flexibility of this approach for developing models of many different linguistic phenomena.

V. THE INTENTIONALITY CHALLENGE

With respect to the directive/assertion distinction, has anything important been left unsaid? Does this game-theoretic account capture everything important about the behavior? No doubt many will regard it as inadequate. The directive/assertion distinction is usually characterized with ample references to the intentions of the speaker (and perhaps the hearer). My account does not require the speaker or the hearer to be the sort of thing that has intentions at all.[20]

One cannot now just require that any adequate account of this distinction make reference to intentions. The whole aim of the Lewisian project (or at least my characterization of it) is a minimalistic one—to see how much of linguistic behavior can be captured without appeal to intentions at all. If my account has failed to capture something central about the directive/assertion distinction, that gap must be explicated in nonintentional terms for the objection to not beg the question.

Following this line of argument, one might continue to push the objection from translation discussed above. In the a-equilibrium pictured above, one can interpret Sig A to mean "do A1 if you are Receiver 1, and do A2 if you are Receiver 2." Similarly, in the d-equilibrium one can interpret Sig A to mean "we are in a state where the best action for you is A1." We seem to have returned to the problem that was initially posed for Lewis and again for Huttegger.

It seems clear that in this case the plausibility of the two translations is stretched further than it was before, but not that those translations are somehow impossible. However, those who would point out that we have not eliminated the possibility of devising both indicative (or assertive) and imperative (or directive) translations of Sig A might have difficulty describing what would be necessary to demonstrate that such a real distinction exists. Even English indicatives and imperatives often convey some information about the state and some information about the act. Knowing that I have said "run" to someone

[20] As in Lewis's model, intentions *can* play a role in explaining how people come to play particular equilibria, but it *need not* play a role.

gives you some information about the world, namely, that the state was such that I wanted my conversant to run.

I do not believe that the possibility of these two translations demonstrates an inadequacy of the Lewis signaling games as a model for linguistic behavior. Instead, this possibility is an example of the more radical problem of translation suggested by Quine.[21] Merely because I can offer two different English sentences which both capture the meaning of the signals in the game is not sufficient to demonstrate that meaning is not present in these games, just as the presence of two possible interpretations of the term 'gavagai' does not demonstrate that the natives are not using meaningful language.

The onus instead should be on the objector to point out what exactly is missing from these games. The challenger must provide some test which one can apply to another human language in order to determine if that language has a directive/assertion distinction that can also be applied to artificial settings like signaling games. Merely asserting that something is not present in these cases leaves open the possibility that French (or Japanese, or whatever) may too be missing that feature, a conclusion that I believe no one would regard as appropriate.

KEVIN J. S. ZOLLMAN

Carnegie Mellon University

[21] W. V. O. Quine, *Word and Object* (Cambridge: MIT, 1960).