What Do Generalizations of the Lewis Signaling Model Tell Us About Information and Meaning?

Peter Godfrey-Smith
City University of New York

Given at MIT and Columbia, November 2013.

1. Introduction

This paper is about the model of communication that David Lewis presented in his first book, *Convention*, in 1969. The book was based on his dissertation at Harvard, advised by Quine. Lewis wanted to defend the idea of conventions of meaning, aiming at the case of language, but starting from simple cases such as Paul Revere's lantern code – *one if by land, two if by sea*. The model addresses how some pairs of behavioral rules can be stable – behaviors of *making* signs and behaviors associated with their *interpretation* – when there are some particular asymmetries between agents in what they can see and do.

The model has always been seen as important in discussions of convention itself, in related parts of philosophy of language, but not so much elsewhere. It was not seen as a model that can tell us about the nature of meaning in broad sense, or the basis for a naturalistic, ground-up semantic theory. I think this was at least partly because Lewis seemed to build so much *in*, to get not that much *out* – Lewis's agents were rational, and had many beliefs about each other and the world, while the signaling that arose between them seemed quite simple.

Brian Skyrms, in his 1996 book *Evolution of the Social Contract*, began generalizing the Lewis model, in a way that also made it more naturalistic. First he 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 had his eye on has analogues in systems of other kinds, where other selection processes shape behaviors of sign production and use.

Skyrms's two later books (*The Stag Hunt* and *Signals*) continued this project, and others have too.[1] There are now bridges between the Lewis-Skyrms models and work in economics and biology. Links between treatments of communication in all these fields have become clear in retrospect. (I think there is relevant work in

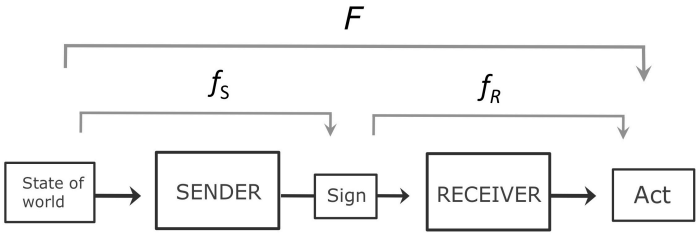
linguistics, but I don't know that literature well.) In the last 15 years or so, the Lewis model has come to look very different in its importance. I think of it as *the* minimal model for thinking about a wide range of semantic phenomena – not all of them, but many of them, roughly those having to do with communication and representation.[2]

The aim of this talk is to take stock of this work and ask: what are its messages? I'll discuss applications of the model in different areas, asking how it unifies these areas and gives us a story about basic, underlying general features of sign use. Is there an orientation to signs and semantic phenomena that comes from the model's core features, where we get a handle on these core features by looking at variations on the model that have, and might be, explored?

I'll first outline the Lewis model, then discuss three topics: cooperation, mental representation, and semantic content itself.

2. The Lewis model and its generalizations

Lewis imagined two agents, who I'll call *sender* and *receiver* (though Lewis said *communicator* and *audience*). The sender can see the state of the world, but cannot act except to produce signs of some kind. The receiver can only see the signs, but can act in a way that has consequences for both sides. The sender applies some *sender's rule*, a mapping from states to signs; the receiver applies a *receiver's rule*, mapping signs to acts. The set-up can be pictured like this:



f_S : sender's rule, maps states of the world to signs.
f_R : receiver's rule, maps signs to acts.
F : the resulting mapping from states to acts.

Figure 1: Lewisian Sender-Receiver System. Black arrows represent causal relations, grey arrows represent mappings.

Lewis assumed *common interest*, *common knowledge*, and *rational choice* by both sides. The Sexton and Revere both want to defeat the British army, so they want the same acts done in each state of the world. There are various ways they can achieve the coordination they aim at, including using the *one if by land, two if by sea* rule. That mapping from states to signs was implemented by the sender, the Sexton, and the receiver, Revere, mapped signs to acts in way that solved their problem. The result is a Nash equilibrium – neither side can improve their payoff by unilaterally changing their behavior.[3]

In *Evolution of the Social Contract*, Skyrms dropped the assumptions of common knowledge and rational choice. The rules of making and using signs, indeed, are simple in Lewis's central cases: if you see this state, produce this sign. But how do the agents reach a good pairing of behaviors, without rational choice and common knowledge? Skyrms showed that evolution can do this – a mechanical selection process, not requiring rationality on either side. If different rules are tried out by different individuals, individuals pass on their behavioral tendencies when they reproduce, and successful individuals reproduce more than others, adaptive behaviors can proliferate in a population.[4]

Skyrms then noted that the simple models of change by natural selection are not only models of biological evolution. Any process in which successful behaviors are retained and unsuccessful ones dropped, by "trial and error" in a broad sense, has some fit to the dynamics of evolution. This process might be reinforcement learning (trial and error by individuals), or imitation of successful behaviors (social learning that is sensitive to success). There is a family of selection processes that can stabilize Lewisian signaling. The four I've mentioned can be roughly ordered by their increasing cognitive demands: biological evolution, reinforcement learning, imitation of the successful, rational choice. These selection processes work on different scales, can operate at once, and produce somewhat different sorts of outcomes. But the first point is that a range of success-based or payoff-based rules of change have broadly similar effects in this context.

This generality with respect to selection process is one of several dimensions of generality the models have. Two others I'd emphasize are the following: Lewisian signaling can occur *between* organisms or *within* them, and signaling of this kind can bridge both *space* and *time*. The temporal dimension gives the models a connection to *memory*, a topic I'll discuss later.

The orientation the model gives us is to the interactive shaping of the behaviors on each side of a sign. Lewis's attitude (a little augmented perhaps) was like this: if we know what the sending and receiving rules are, and why they are that way, we know all there is to know about the interpretation of the signs. I interpret this in a non-trivial way, to mean: do not get hung up about what content is *in* the signs. They have the role they have, mediating between sender and receiver behaviors. All the facts about what the sender and receiver behaviors achieve, and why they do so, have their own sort of relevance to questions about the "meaning" of the signs. ("Significance" would probably be a better word to use than "content.")

Signs themselves do have notable properties – I'll discuss some in a moment. In a Lewisian system, these are due directly to the sender, but indirectly to the receiver, insofar as the sender's behavior is shaped by the prevailing patterns of interpretation waiting downstream. The way those interpretations matter varies according to the selection process operating – it might involve actual feedback from past instances, or envisaged consequences in a smart sender.

One important set of properties signs have are those described in information theory, developed especially by Claude Shannon. Here is Shannon's famous diagram of a "general communication system."

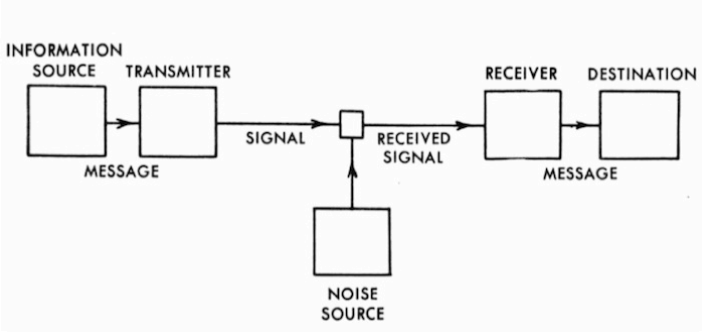
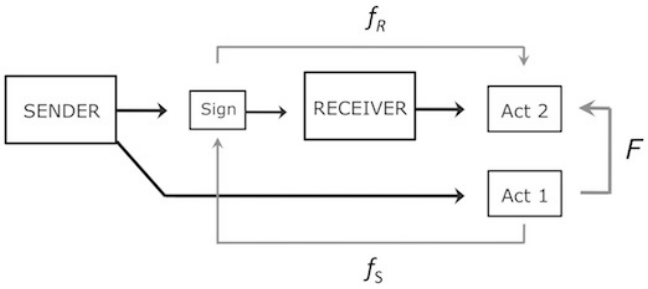


Figure 2: Shannon's 1948 diagram of a "general communication system"

In retrospect, and without any apparent influence, we can see Shannon's and Lewis's theories as fitting together in the following way: Shannon *took for granted* the sender and receiver roles, and gave a theory of channels that could make communication possible between them; Lewis *took for granted* the existence of a suitable channel, and gave an account of why anyone would want to occupy the sender and receiver roles.

I'll only briefly mention another kind of modification of the original model. The Lewis model is a model of the use of signs to enable one kind of coordination, the coordination of receiver actions with states of the world. There is also a different kind of coordination, of sender's act to receiver's act, where the difference between an "act" and a "state" is that a state is exogenous, not chosen by one of the agents whose choices are being modeled. The use of signs to achieve *act-to-act*, as opposed to act-to-state, coordination might be drawn like this:



f_S : sender's rule, maps acts to signs.

f_R : receiver's rule, maps signs to other acts.

F : the resulting mapping from acts to acts.

Figure 3: Sender-receiver system with act-to-act coordination

This is like the Lewis model in the way it focuses on the stabilization of sender and receiver behaviors, unlike it in the kind of coordination achieved. (In game theory, the "Battle of the Sexes" has this second structure, when it includes signaling.) The border between the two might be unclear, and many empirical cases will have a mixture of both. I'll use the phrase "sender-receiver system" to cover cases like this as well.

I'll cover a couple of other points more quickly. First, syntax might be seen as problem for this approach, but it's not. Rules on both sides can be systematic – can include substitutions and transformations, and so on. Skyrms's and Jeffrey Barrett's extensions of the model include cases with simple syntax. Second, talk of "rules" here might suggest that there has to be a kind of fixity in use of a sign. This is just the simplest case. Lewis himself 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 sure that whatever the receiver might do, given all their available evidence, will be something that is OK for the sender. Lewis

said that the distinction between indicative (descriptive) and imperative (command) contents depends on which agent's policies include more discretion.

I've been referring to this work as a "model" or a family of models, and earlier as a "minimal" model. What is meant? I don't want to wade into general questions about modeling, but I take it that in application to just about any empirical case, a model like those discussed so far is idealized; it simplifies in various ways. For example, in the case of human sign systems like language, what is "selected" – what is shaped by evolution and other selection processes – is a mass of dispositions, made possible by underlying mechanisms that also have further effects, including giving rise to many additional behaviors beyond those covered by the model. The core phenomena of language use include the fact that great torrents of what we say is aimed at making subtle changes to psychological states. Changing these states is generally the only way to eventually affect others' actions, but once we get going on this, we end up doing a lot, it seems, just to change the psychological states.

Given all this, what *is* the role of the Lewis model? In what way is it progress? Some particular fruits of this approach will, I hope, become clear below, but I can say something more general now. What the model does is induce us to approach signs and symbols by orienting us to the pairs of behaviors on each side of a sign, and to how they fit together. "Behaviors" is perhaps too narrow a term here; it would be better to talk of the mechanisms of production of signs and the mechanisms of their reception. The orientation is to the pair of questions: why send (in the way you do); why pay attention to what is sent, and why interpret it in the way you do? From this viewpoint, other approaches have often been one-sided, focusing on either the expressive or the interpretive side of an essentially two-sided phenomenon. The simple, Paul Revere-like cases make this clear; it would be a mistake to give a theory of those cases that gave primacy to one side or the other. Compare Grice and Davidson's large-scale theories about language. 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* by using a bit of language (we could ask *what the Sexton said* with his one lantern). Donald Davidson gave a view of language based on the properties of good *interpretation*, how an interpreter makes sense of the stream of incoming sound. Those are one-sided views in the sense I have in mind. I say this, and partisans of both sides object: Grice(/Davidson) do include both sides. Yes, in a sense, but in Grice, the receiver appears only as seen through the eyes of the speaker. (After all, the speaker is the one producing the language.) In Davidson, the

speaker appears as seen through the eyes of the interpreter. (Maybe that is not true of Davidson's later work, from the 1990s – see "The Second Person"). A more important objection to my point here is that it is easy to *say* "take a two-sided approach," but this admonition is empty if there is not much you can do with it. It is never a good criticism in science to simply say "you left this out." You always leave things out, and the choice is always about what. This is one reason why I see the Lewis model as so important, as it gives a nontrivial model of the interaction between the two sides, treating the co-evolution or co-selection of sender and receiver behaviors as the way to think about signs and their meaning.

I'll make one more point linked to this model-based orientation. Empirically, there are *clearer* or *paradigm* cases of sender-receiver systems, and *marginal* ones, cases that fit the model to some extent but only partially. This partial fit might be because there are not separate entities playing the essential roles, or because some of what is going on is shaped by what the model covers, but other factors are operating too. A diagnostic point follows: there will often be temptation to describe partial cases of a sender-receiver set-up using language that only really applies to clear, paradigm cases, in a way leading to confusion. We as people are not just very used to working with sender-receiver systems, but also used to talking *about* them, from the outside. We have habits of response to sign-like objects that are shaped by experience with paradigm public cases – words, maps – and these habits carry over when we encounter new partial cases in areas like biology. There is also a broadly evolutionary point here: if sender-receiver systems appear in clearer and more marginal forms, we can ask: which factors in an evolving system will push it towards one or the other?

3. Common Interest

Now I'll discuss some specific topics. The first is common interest and cooperation.

Communication is often seen as a fundamentally cooperative affair, an interaction between agents whose interests are at least fairly well aligned. Forms of this view are seen across different literatures. Paul Grice proposed a *cooperative principle* governing conversation: make your contribution at each stage the one required by the "accepted purpose or direction" of the conversation.[5] Ruth Millikan has a sender-receiver-like view in its general shape, and signs or "intentional icons" for her exist between two "cooperating devices." In Michael Tomasello's book *Origins*

of Human Communication, the origins of that communication lie in special kinds of fine-grained cooperation. (You might also see a link, within interpretation-based views, to "principles of charity.")

How does this look within the Lewis model? Lewis himself only dealt with situations in which, as he put it, common interest "predominates."[6] Sender and receiver want the same acts performed in each state. Lewis's later work on language itself (eg., "Languages and Language") also assumed a cooperative framework. What happens as we move away from this situation? Skyrms looked at a few cases in his 2010 book *Signals* (discussed in my review of his book in *Mind*). As interests diverge, signaling is initially, at least, still possible, but it changes in character. An old model from economics, Crawford and Sobel 1982, using a different set-up, had the result that as interests diverge, fewer distinct signals can be used, until signaling collapses altogether.

Let's look more closely. Suppose a situation has *n* equally probable states and *n* possible receiver actions. Each act-state combination is associated with a payoff for both sender and receiver. Those can be represented in a matrix like this:

		States		
		S1	**S2**	**S3**
	A1	5,5	2,2	5,5
Acts	**A2**	6,6	4,5	3,3
	A3	2,1	6,6	0,0

Within each state, both sender and receiver have *preference orderings* over acts. They might agree entirely in their orderings, or reverse them entirely. Between those extremes are many intermediates, and these are only partially ordered. The two agents might agree always on what is worst but not on what is best, might agree entirely in some states but not others, and so on. Manolo Martínez and I, in a 2013 paper, accept this but give some coarse-grained overall measures of common interest. We said there is *complete common interest* when sender and receiver have the same ordering in each state (as in the example above); we said there is *complete conflict of interest* when the orderings are reversed in every state; and between those, *partial common interest* is measured by counting *discordant pairs* in their orderings – pairs of acts for which their preferences, in a state, are switched.[7] We average the number of discordant pairs over states and scale the results to lie

between zero and one, to yield a measure called *C*. (We also discussed a finer-grained measure, *C**, but for now I'll just discuss *C*.)

In a Lewis model, the only reason for the receiver to attend to signs in guiding their behavior is that signs may have some predictive relation to the state of the world – might carry some information about it, in Shannon's sense. But the only reason, it seems, for the sender to put this information in the signs is that the receiver will pair acts with states in a way the sender wants, or at least does not mind. Given this, what is the *minimum* level of *C* for sender and receiver to be in an equilibrium when signs carry some information about the state and the receiver guides their actions with them – formally, this means there is *mutual information* between states and acts. Once we know the minimum, what other relations are there between *C* and the chance of getting signaling to be stable?

Martínez and I addressed this question with a computer search of hundreds of thousands of games, in which the payoffs defining each game are set independently and at random, and the three states treated as equally probable. The computer looked for equilibria in which informative signals are produced and used. To our surprise, the minimum level of *C* that allows signaling is zero. In a very small fraction of *C*=0 games (of the order of a tenth of a percent), sender and receiver can signal at equilibrium when they have completely reversed preferences over actions in every state. Here is an example:

	S_1	S_2	S_3
A_1	31,7	0,95	57,26
A_2	5,71	99,1	15,62
A_3	17,66	62,23	28,48

Sender: $S_1 \rightarrow m_1; S_2 \rightarrow [(29/47)m_1, (18/47)m_2]; S_3 \rightarrow m_2$

Receiver: $m_1 \rightarrow A_3; m_2 \rightarrow [(37/99)A_1, (62/99)A_2]; m_3 \rightarrow [(37/99)A_1, (62/99)A_2]$

These games uncovered by the computer search are pretty opaque. But Martínez was able to use them to construct some "artificial" cases. Here is one:[8]

	S_1	S_2	S_3
A_1	1,8	8,1	0,6
A_2	3,7	6,3	1,5
A_3	8,1	1,8	5,3

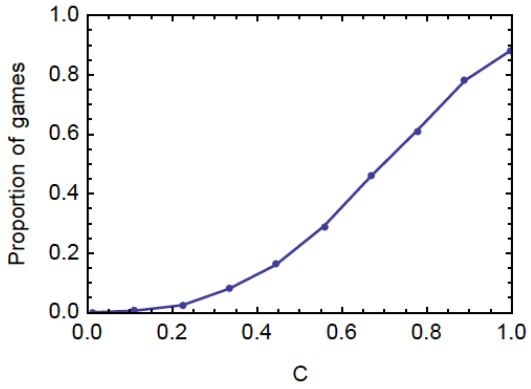
Sender : $S_1 \rightarrow m_3; S_2 \rightarrow [(3/7)m_2, (4/7)m_3]; S_3 \rightarrow m_2$

Receiver : $m_1 \rightarrow A_2; m_2 \rightarrow [(5/7)A_1, (2/7)A_3]; m_3 \rightarrow A_2$

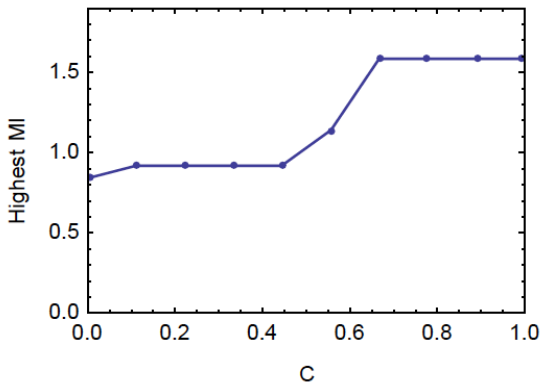
What is going on? There are various generalizations that can be made about these cases, but the deepest analysis I can give derives from some emails about the paper by Elliott Wagner and Carl Bergstrom.[9] Even when *C*=0 and all preferences are reversed, a sender can sometimes put together a pair (or larger number) of *uncertainty bundles* over which there is a kind of common interest. Suppose there are two such bundles (as in the case above). A bundle is created by the sender by associating a probabilistic mix of states with a signal. What is needed for informative signaling is for the sender to create two such bundles where there is a pair of acts (or pair of mixes of acts) such that both parties would rather have each act (or mix) associated with a bundle than have the receiver act the same way all the time. This is sometimes possible with *C*=0. When senders can do this, interests of senders and receivers come far enough into alignment for information – over which uncertainty bundle the world lies in – to be exchanged.[101112]

These cases are rare, and some of the equilibria are very fragile (others less fragile). One way to respond is to say that reversed preference orderings over acts in every state turns out not to be strong enough as a notion of *complete* conflict of interests. The concept of a *zero-sum game* is a stronger concept, and we don't claim that signaling can be stable in a zero-sum game. Applying this concept, though, requires making much stronger assumptions – our measures don't require that payoffs be commensurable across agents.

Setting aside extreme cases, here's a more general question. What proportion of games of a given level of *C* have *at least one* equilibrium with informative signaling?



And what is the *maximum* level of mutual information between states and acts seen in any equilibrium in any game with a given level of *C*?



It's interesting to see quite high values of mutual information at very low values of *C*, including zero.[13]

What does this show? It shows there is *some truth* in the informal claims about common interest and communication – common interest helps communication. But there are surprises. Levels of common interest so low that we initially defined them as complete *absence* of common interest – cases when the best act for one side is always the worst act for the other – can support signaling. And they can support quite a lot of information exchange.

4. Mental Representation

Next I'll look at the content of thought – not in relation to language meaning, but in its own right. A first response you might have here is that the Lewis model is not relevant, because that model is about communication, and mental content is not communicative, but a different sort of thing. Perhaps, but let's take a closer look.

In the mental representation debates, there is a distinction sometimes made, often hesitantly, which Tim Crane put like this: it is one thing to ask whether our minds *represent* the world, something else to ask whether our minds *contain representations* of the world. What's the difference? Perhaps a system like a brain can, in virtue of many facts about all of it, have a kind of involvement with external things – things it responds to and acts on, and so on – without there being discrete representation-like objects in the brain. OK; what then *would* be required for little representation-like objects to be there? That there be internal structures that are *used* in a particular way – consulted or read. The notion of a "cognitive map" in neuroscience is a good example.[14] Any such representations must have been made, produced. So the representations then mediate between two things, a producer and user. We seem to be back with the Lewis model.

We might be, but we might not. Not just any making and reacting-to fits the model. What *are* these producers and users? In Millikan's theory, the "producers and consumers" of signs can be two organisms, or parts of one organism. In both cases we can have "cooperating devices," of the right kind. Millikan's schema is clear, but it has never been clear what these entities would have to be like in order to exist within a brain. Sometimes the "consumer" looks like it is everything downstream of a putative representation and the producer is everything upstream.

I see this situation like this. As I said above, there are clear and partial cases of sender-receiver systems. In the paradigm between-agent cases, signs exist "between" two organisms, they have a role in the lives of each, and each has their own agenda. A good sign in the paradigm, highly cooperative, between-organism cases is something that is cheap, stable, and easily controlled. (It may also lend itself to the compact organization of information.) In the within-organism case, the Lewis model is there as a kind of skeleton, in the axis from senses to effectors. The senses can see but not act, the effectors can only see signs but can act in a way that affects both sides. But the "channel" between them is made up of living material, of cells with all their active capacities. Because of common interest and this special raw material, it is natural for the intermediate structures to take on further roles – not just transmitting but modifying and "processing."[15] Neurons, for example, might start out as a mere intermediaries, but once they are present there is much more they can do. This takes the system away, to some extent, from the sender-receiver pattern, in ways that make adaptive sense. To the extent that this is true, and the separation between signs and sign users is lost, "signaling" in the brain will be a

marginal case of an sender-receiver structure, but one that our habits of interpretation will tend to make us describe in richer ways, which are somewhat misleading ways – ways that anti-representationalist cognitive science reacts to and criticizes, without yet being able to put much in their place.

That first run-through was organized around the division between senses and effectors in an organism. This is a roughly spatial separation of roles. But there is another axis to consider: time. Think of an organism as divided into stages as well as spatial parts. It is often the case that earlier stages have access to information that will be useful in guiding the actions of later stages. This also fits the Lewis model: one can see memory as communication between stages, perhaps between earlier and later selves. Memory in some respects lies *between* interpersonal and intrapersonal signaling, because a person at a time has a complete psychological profile. Memory seems to obviously have a partly communicative character when it is routed through external media (notes and so on), and the same, I think, applies to on-board, psychological memory too. This varies across mechanisms; not all memory mechanisms call on an agent's rich, time-specific psychological properties. Some of the processes are genuinely subpersonal. But some of it draws on the agent's whole psychology.

If memory is communication between stages, can there be divergent interests across stages? This I discuss elsewhere in detail, but here is the situation as I see it. The basic architecture of memory was shaped by evolution. Stages of an organism cannot diverge in their evolutionary interests – not feasibly, anyway. But it's entirely possible for stages to differ in their preferences, and sometimes to do so in a way foreseen by the earlier stage. In principle, all the detail of the extensions of the Lewis model that concern divergent interests, discussed in the previous section, become relevant. We are familiar with *commitment devices* as means by which earlier stages constrain later ones. Can earlier stages also do this by controlling the flow of information to later stages? This seems possible in principle, hard in practice, and harder to the extent that information flow stays within the skull, and does not rely on external media. It is also easier to simply obstruct the information flow (trying to forget) than to actively deceive your later self. Attempts to deceive your later self will tend to be undermined by retention of information into the future about your own deceptive plan. Here a special role is played by the fact that later stages of an organism are material continuations of earlier ones, something that is not true in other realizations of a Lewis-like structure.

5. Information and content

Finally, I'll look at the "content" of signs, especially in relation to attempts to give a naturalistic semantic theory.

I said earlier that Lewis in his original discussion steered us in a particular way. Expressing the message in a strong form: if we know what the sending and receiving rules are, and why they are that way, we know all there is to know about the interpretation of the signs. We shouldn't get hung up about what content is *in* the signs. They have the role they have, mediating between sender and receiver behaviors. All the facts about what sender and receiver behaviors achieve, and why they do so, have their own relevance to questions about the "meaning" of the signs. *Meaning-talk* is talk of many sorts of involvement that signs have with things in the world, and with psychological states of agents, via the pair of sign-related behaviors on each side. In different contexts, different relations of this kind will be worth talking about. *Talking about* the meaning of signs is another kind of talk, which the model can be applied to, inasmuch as it can be applied to other complicated things that people say.

That is the orientation I take from the model. The last part of what I said might go beyond or away from Lewis's own attitude, at least in the case of language. This view is somewhat deflationary of attempts to give definite theories of meaning, of the *content* of linguistic and other signs. What I'll do here is pursue one path in this area, which relates to attempts to give naturalistic accounts of content, as seen in a lot of work in the 1980s.

Here's one general view about content, tied perhaps to a larger philosophical outlook. Meaning has to do with informativeness, evidence, with showing how things are. This is the approach that Dretske, drawing on Shannon, explored, and Skyrms favors this way of handling content in the sender-receiver models. In a Sexton-and-Revere case, a message carries the information that the British are coming by land because it raises the probability of that state to one. Perhaps content in a richer, semantic sense is a near-relative of that simple notion of *informational content*.

Another approach ties meaning not to *evidence* but to *success* – the success conditions for behavior based on a sign. A cartoon version might say this: a sign means that *p* iff when acted on in a normal or rational way, the result is success if and only if *p*. There are steps towards this idea in Bain and Peirce, but a clearer

version is in Frank Ramsey, in "Facts and Propositions."[16][17] Ramsey's version is behaviorist, treating beliefs as sets of actions.

> Thus any set of actions for whose utility *p* is a necessary and sufficient condition might be called a belief that *p*, and so would be true if *p*, i.e. if they are useful.

Since then, two literatures have pursued this approach. One is *teleosemantics* (Millikan 1984, 2004, Papineau 1984). Success is relevant to content via the idea of a biological function, which is in turn understood in terms of evolutionary history – a history of selection. A *truth* condition of a sign is a historically normal *success* condition for use of the sign. A smaller literature has tried to directly follow up Ramsey (Whyte 1990, Mellor 1998, Blackburn 2005). Jamie Whyte's version: "A belief's truth condition is that which guarantees the fulfillment of any desire by the action which that belief and desire would combine to cause." (1990). This asserts a non-historical link between truth and success. There are many objections that might be raised to this proposal, but let's work within it for a moment. I'll call the success-based notion of content the *adaptive* content of a sign, as contrasted with its *informational* content.

How do they look from point of view of the sender-receiver models? Did Lewis, for example, have one or the other of these in mind, in his original discussion? We can't tell, not just because he didn't say, but because in the simplest cases the two kinds of content are brought together. Remember the Sexton and Revere, and assume they are at their signaling equilibrium. One lantern means the British are coming by land. This is the case if we have informational content in mind: that signal raises the probability that they are coming by land to 1; it indicates that state. This is *also* the success condition of actions based on the sign, given the receiver's rule. The production condition is the success condition and everything goes as well as it can.

In the simple cases, this is no accident: sometimes (not always) the selection process brings the two kinds of "content" into alignment. If the sign is produced in circumstances outside its success condition, one side or the other can adjust, raise everyone's payoffs, and reach equilibrium. That is the point of Lewisian signaling, to have production conditions and success conditions aligned, and with a lot of common interest and some other assumptions, this can often be done. Selection brings the backwards-looking direction of involvement and the forward-looking direction of involvement together on the same target.

Once we leave the simple cases, the two come apart. Suppose the sender can't discriminate all the states that matter, and hence can't send distinctive signs in every state that demands a different action. The sender will "pool" some states and the receiver will produce an act that works best as a cover-all across those states (see the Appendix, first case). Then, as discussed in many cases in the 1980s, the informational content will be disjunctive or weak, and the success condition will be logically stronger (*moving black dot* versus *fly*). There are also cases that go the other way (see again the Appendix). Suppose the sender can discriminate all the states that matter and the receiver can produce optimal acts in each state, but some states are "forgiving," giving high payoffs to a range of actions other than the very best action. Then when all goes optimally, some signals have success conditions that are logically weaker than their informational content. The signals prompt acts that work well in states that are never actual when the signal is produced.

When there is only partial common interest, the "success conditions" for a sign, given the receiver's rule, are different for sender and receiver. The C=0 case discussed earlier is an extreme example, but any degree of divergence of interests leads to divergent success conditions. Informational content remains univocal, but the signals sent might not rule any states out, instead shifting probabilities in complicated ways.

Some might try to find a way through all this; the real content is the following combination of informational properties, action-related properties, and perhaps other things.... I think the lesson instead is this: the idea of content, in this sense, is a *tool for the simple cases*. It has a home in cases where the sender-receiver system is such that signs have a very simple kind of involvement with states in the world, as in the case of Revere. In the complicated cases, signs have various connections with states, actions, and success – they are still involved with things outside them – but they don't have the simple and definite kind of involvement with states of the world that some influential philosophical discussions invoke.

Appendix:

First, suppose the payoff matrix is like this, with only one payoff per cell because sender and receiver are in complete agreement:

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

An equilbrium pair of rules:

f_S: {S1 → m1, S2 → m1, S3 → m3}
f_R: {m1 → A1, m2 → A2, m3 → A3}

The sender has to pool S1 and S2, and the receiver will choose A1 in response to that signal, m1. A1 is not just the best on average in S1 and S2, but the *right* act in S1.

Second, suppose the payoff matrix is like this:

		States		
		S1	**S2**	**S3**
Acts	**A1**	2	2 - *e*	0
	A2	0	2	0
	A3	0	0	2

The sender can discriminate all states, and the receiver performs the right acts.

An equilbrium pair of rules:

f_S: {S1 → m1, S2 → m2, S3 → m3}
f_R: {m1 → A1, m2 → A2, m3 → A3}

Maximum payoff is achieved. The informational content of message 1 is *S1*, as it is always produced in S2. The adaptive content of m1 is *S1vS2*, a disjunction, and this is true even though A1 is never produced in S2, as the tiny *e* is enough for the receiver to prefer A1 always. Part of the adaptive content is idle, irrelevant.