The Evolution of Meaning

Peter Godfrey-Smith
City University of New York

The Thacher Lecture at George Washington University, April 2012

Models of "sender-receiver" systems provide a new way of thinking about meaning and the evolution of representation. I describe this work and then look at the application of these models to internal representation and the content of thought.

1. Introduction

The last hundred years or so in philosophy have seen a huge amount of work on *meaning*, and related topics such as representation and truth. This began primarily in philosophy of language, and spread to philosophy of mind. Words and sentences have meaning, at least some of the time, and apparently so do thoughts. One thing we can apparently do with sign systems, both in language and in thought, is *represent* the world we are dealing with. As Frank Ramsey put it, in the 1920s, our beliefs are "maps by which we steer."

We can try to understand these phenomena with a very general theory, a theory covering *all* representations and signs, or in a more piecemeal way. For a while very ambitious theories were offered, with great enthusiasm. During the last part of the 20th century there was a move back towards less general ones. In part this was because of frustration, a felt lack of progress. This is certainly true within the naturalistic side of philosophy, which is my side, and it is also visible within some parts of "Continental" philosophy, in the rise and decline of semiotics.

Now I think things have changed; the shape of a general theory is becoming visible. I'll argue that progress is now possible through use of a particular family of models, combined with some new views about what the models *do*, how they relate to phenomena.

The first outlines of the model are familiar and intuitive. Any object which is a sign or representation has this status because of the way it is used in an interaction between two other things, which can be called a *sender* and a *receiver*. "Sending" is

understood to include any kind of creation, display, or inscription of a sign. "Receiving" involves the sign's reception and use.

Nearly everyone would agree that senders and receivers are *often* important, and *some* signs should be approached this way. It is more tendentious to put this in the center of *everything* here. It looks odd to apply the same model both to communication *between* agents and to things going on *inside* a single agent. That is something I will discuss. There is also another kind of generality that is controversial. A former colleague told me about a discussion at a conference on human action. One person said: "People are very complicated. In order to get a handle on human action, we should start with ants." Another person said: "People and ants are very different. In order to get a handle on human action, we should start with people." The discussion divided, unproductively, into ant-people and people-people. That is not how we'd like things to go. What should we hope for? Ideally, in the best case, we'd have an account of origins and simple cases, and an account of how on this basis the more sophisticated cases arise – a theory of the basics that makes sense of what comes later.

2. Senders, Receivers, and Signs

Here is a set-up found at many places in the living world.

Sender-Receiver Configuration: A *sender* produces a sign in a way that is responsive to something in the world. A *receiver* acts on the sign.

An appropriate example at George Washington University is the case of Paul Revere. One person, the Sexton of the Old North Church, can see the behavior of the British army. He produces a signal – lanterns in a church tower – that can be seen by Revere, who can do something in response. Sender and Receiver are Sexton and Revere.

Suppose we find two agents doing this. Why? Why does the sender bother to send, and why does the receiver pay any attention? A simple answer is: common interest. Sender and receiver have the same preferences for what they want done in each state of the world. The sender can see what state the world is in, and the receiver can act. Putting it metaphorically, the sender acts as the receiver's eyes, and the receiver acts as the sender's muscles. That keeps both sides doing what they

are doing, and leads to the production of those unusual objects between them – signs – small things that can have large effects.

In this way of thinking, to be a sign is to be located between two other things. A theory of signs is a theory of how distinctive behaviors on each *side* of the sign come to exist. It is an interesting fact that in the history of philosophy this has not been the main way people have approached these matters. But one theory that does do things this way is David Lewis's model of "conventional signaling," presented in his dissertation and his first book, *Convention* (1969). Lewis assumed a "communicator" and "audience" with shared interests and definite roles. The communicator (my "sender"), can see the world but cannot act except to make signals; the audience (my "receiver") can only see the signals, but can act in a way that affects both. Each side adjusts their behavior independently, through rational choice and with knowledge of what the other agent knows. Lewis showed that in a case like Paul Revere, a system of informative signaling can be maintained – can be an equilibrium. There are pairs of rules such that, if sender and receiver reach them, neither will have any reason to change their behavior. This includes the historical "one if by land, two if by sea" rule in the Revere case. One lantern *means* that the British are coming by land, because of how the one-lantern sign mediates between a seeing sender and an acting receiver.

The Lewis model did not have much influence on the part of philosophy looking for a ground-up theory of meaning and representation, because he assumed *rational* agents and common *knowledge* – the sorts of things that seem to need explaining. I'll come back to this shortly. First I will describe a bit more history.

I see the model I work with as having two origins. One is the Lewis model. The other is outside philosophy. Claude Shannon, in 1948, introduced *information theory* or *communication theory* in the mathematical sense. Here is a famous diagram from his 1948 paper.

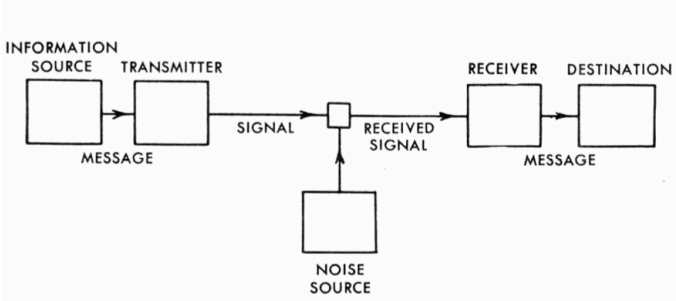

Figure 1. Shannon's diagram of a "general communication system"

This is a picture of something like a sender-receiver configuration. Shannon said that *information is carried* in a set-up like this whenever the state of the signal *reduces uncertainty* about the state of the source. It is a matter of physical correlation or dependence.

Shannon drew on no philosophy, and Lewis two decades later did not seem to draw on Shannon. But the two contributions fit together in a way that can be seen in retrospect: Shannon *took for granted* the sender and receiver roles, and gave a theory of the channels that could successfully achieve coordination between them; Lewis *took for granted* the possibility of a channel, and gave a first account of how agents could come to play the sender and receiver roles – how these roles could be stably occupied.

People writing about information theory, from Shannon onwards, often say that it is part of the theory that "meaning is irrelevant" (Shannon 1948, Bergstrom and Rosvall 2009, Dyson 2011). Meaningless strings can be transmitted over the channel as well as messages of great strategic import. It is true that meaning might not have to be considered once we take for granted the facts about senders and receivers that Shannon took for granted, but it is not irrelevant to the total project.[1]

The Lewis model had limited influence outside some technical discussions of language, as I said, until Brian Skyrms, in his 1996 book *Evolution of the Social Contract*, recast the Lewis model in an evolutionary framework. Skyrms showed that the model does not depend on the "intellectualized" framework in which Lewis presented it. Various kinds of selection processes, including biological evolution and learning by reinforcement, can stabilize the crucial sender and receiver behaviors, and these behaviors can be seen in organisms much simpler than humans. The model applies more broadly than Lewis realized (see also Skyrms' *Signals*, 2010).

My work in this area sets out from that point; from a Lewis-like specification of sender and receiver roles, the idea of a range of mechanisms that stabilize sending and receiving, and the idea that understanding the meaning of signs is

[1] In philosophy, there is a "second generation" of work in the 1980s, with complementary roles. Fred Dretske (1981) introduced information theory to philosophy of mind, but not in a sender-receiver framework. Ruth Millikan (1984) argued that any entity that is a representation has that status as a consequence of its relations to a "producer" on one side and an "interpreter" or "consumer" on the other. She did not appeal to information, but argued that signs can "map" the world in virtue of how their producers and consumers evolved.

understanding their role in set-ups like these.[2] I will move between different terminologies according to convenience. On one side, there are senders, producers, writers; on the other, receivers, readers, users, consumers. I will use the phrase *sender-receiver configuration* for systems that have at least an approximate fit to the Lewis model. There are also models of signaling that have a different structure, as the sender is not signing in response to some state of the world that is determined independently, and where instead the point of signing is purely to achieve coordination between behaviors (Robson 1990). Here I will stay with the Lewis-style models.

When we fix the sender-receiver configuration in our minds and go looking through nature, we find a great diversity of things that fit the pattern, sometimes approximately and sometimes in a more exact way. Biology uncovers them at an ever-increasing range of scales. Some are found between organisms: bee dances, mating calls, some alarm calls, chemical signals that mark trails and territories. Others are found within organisms. These include hormones and some gene regulation systems, and at least some events that go on within nervous systems. There is also another dimension of generality. The gaps that are bridged in most discussions of standard examples of the SRC are, roughly speaking, spatial gaps: Paul Revere, the bee dances. But there are also gaps in *time*. The theory applies to both kinds of bridging. The sender-receiver configuration is a *natural kind*, something that evolution builds over and over again, on different scales and from different materials.

Once we see the SRC in this way, as a recurring kind, the next question to ask is: why should we find things in this arrangement? The answer that Lewis gave was that both sides benefit. This is the core of the more general answer, too, though there is a lot of detail to add.

Being a bit more formal, the sender has a rule, f_S, by which they respond to states of the world by producing signs. The receiver has a rule, f_R, by which they respond to signs by producing acts. Senders' rules include things like: ignore what you see in the world and always send sign X, and they also include rules in which each state of the world is indicated with a unique sign. Receivers, also, can choose to attend to signs or choose to ignore them. What has to be explained is when and why the sender and receiver will settle on rules in which the sender sends signs that are associated with states of the world – signs that are *informative*, in the Shannon

[2] For other work along the same lines, see Harms (2004) and Huttegger et al. (2010), and additional works by those authors.

sense – and where the receiver will use them as guides to action. A sign itself can be *anything*, as long as it is related in the right way to the sender's and receiver's rules. What matters is how the rules on each side interlock. The role of these rules is represented in Figure 2.

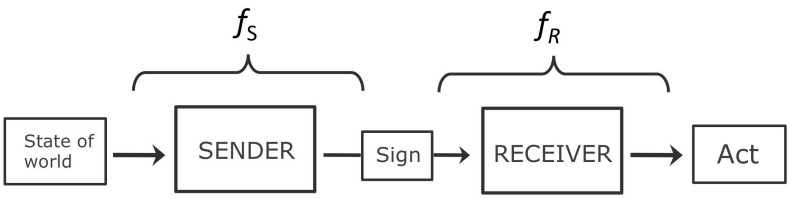

Figure 2: Sender-receiver configuration (SRC) with the sender's rule and receiver's rule

There is a family of mechanisms by which the consequences of sending and receiving can stabilize, or destabilize, the rules followed by senders and receivers. These are all, in a broad sense of the term, "feedback" mechanisms – the consequences of the pairings of acts with states feed back and affect the two rules that are together responsible for that pairing. This family of processes, which operate on different scales, include evolution by natural selection, learning by reinforcement, and some kinds of copying – the copying of successful individuals around you – along with rational choice (Skyrms 2010). These are all ways by which the consequences of the application of sender and receiver policies can stabilize or change those policies. In the evolutionary case, suppose that different individuals in a population follow different rules of sending and receiving, receive different payoffs as a result, and reproduce in a way such that offspring use the same rules as their parents. Then combinations of rules that work well together in achieving payoffs will proliferate in the population. In the rational choice case, an agent foresees the consequences of an policy and either applies it or changes it. These processes operate on different time-scales, and apply to different versions of the basic set-up. There are many variations: sometimes the individual who sends a message is not affected by the receiver's act, but some other agent of the same type as the sender, or some other member of the sender's lineage, is affected; sometimes there is more than one receiver, and sending is stabilized by many effects that each sign has. So while a simple summary says that signaling is stabilized by "common interest" across sender and receiver, this means different things in different cases and talk of common interest is really a shorthand for a description of how *some* feedback process is present that is sensitive to the outcomes of receiver actions in the right way.

Continuing in the same area: Lewis assumed *complete* common interest, or near enough, in his model. Sender and receiver both benefit from exactly the same pairings of acts with states of the world. Clearly this is a special case. If there is *complete conflict of interest*, such that sender and receiver have entirely opposed preferences, then signaling cannot be stabilized. Most cases lie between the extremes, and further modeling work shows that *partial* common interest is enough to stabilize signaling in some cases, though the signaling that results is often less informative or more fragile.[3] There are many kinds of partial common interest, and this is an area where more work could be done. But the picture emerging is that a range of feedback processes can act on sender and receiver rules in the right way, and stabilization of signaling requires at least partial common interest between the two sides.

3. Elaboration and submerging of the SRC

Earlier I said the SRC is a "natural kind," something that nature builds over and over again. It is a particular sort of natural kind, though. It is one that exists sometimes in clear, definite forms and at other times in partial or marginal forms. It is sometimes a backbone on which more gets built, and sometimes a structure that gets washed out or submerged, with a loss of the distinctive relations between sender, sign, and receiver. This can happen either because things fall apart, or because they stay together but change into something else. The elaboration and submerging that occurs depends on the context and on the raw materials. This can be expected to work differently across the cases of signaling between organisms and signaling within them.

Between organisms: A crucial factor here is the relationship between the interests on each side. In the basic Lewis-Skyrms model, complete common interest is assumed. When sender and receiver are different people or different animals, clearly this will often break down.

Partial common interest, as I said, is enough to stabilize signaling in some cases. In other cases there is no stabilization – signaling is lost or there may be an

[3] Crawford and Sobel (1982), Godfrey-Smith (forthcoming)

ongoing "arms race" between senders and receivers trying to use the others' sensitivity for their own ends.[4]

Some recent philosophy of language argues that once you have smart human agents involved, communicative behavior is full of innovation and context-sensitive modification of "conventional" language, even when interests are shared (Sperber and Wilson 1986). Stabilized "meaning" of signs is part of what is going on, but a smaller part than earlier theories have supposed. Lewisian signaling is submerged by the flexibility and the multifarious agendas of human beings.

Within organisms: In the within-organism case there is much less conflict of interest. In fact, that language seems odd. What does it mean to say that your arm has the *same* interests as your leg? Rather, these are parts of a single cooperating system. if "interest" talk is to be used, then this is a case of *joint* interests than *common* interests. But though evolution mostly aligns the biological interests of the parts of an organism, and an organism is generally also a single center of chosen or conscious goals, it is possible for common interest to break down to some extent within an organism. So I will keep using the language of "common" interest.

In any case, the primary source of both elaboration and submerging of the SRC in the within-organism case is different; it comes from the active powers of the intermediate structures.

In paradigm SRCs, a sign is just an intermediary. A good sign in the paradigm between-organism cases, for example, is something that is cheap, stable, and easily controlled. In the within-organism case, the "channel" linking the parts of an organism is made up of living material, 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 merely transmitting but modifying and "processing." 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 SRC pattern, in ways that make adaptive sense.

A *paradigm* SRC has clear separation between the roles of sender, receiver, and sign, and stabilization of informative signaling via common interest. *Marginal* cases have a less clear separation of roles, and owe their existence to something other than stabilization by common interest.

---

[4] Dawkins and Krebs (1978), Owren, Rendell, and Ryan (2010).

4. Thought

In this rest of this talk I will look at the application of the model to thought, and to debates over "mental representation."

For several reasons it can look like a mistake to apply the SR model to things like this. First, the "sender" and "receiver" in the model are treated as agent-like things. If agency is taken for granted in the model, it can't explain what it is to be an intelligent agent. But there is another way of looking at things. Earlier I said about the model: "the sender acts as receiver's eyes; the receiver acts as sender's muscles." That is metaphorical in the between-agent case, but also has a more literal application. The SRC is seen in the basic design of much of life, especially animal life, with respect to the causal flow from sensors to effectors. It is part of the "design skeleton" of any organism that has to adjust its activities to what is going on around it.

"Skeleton" is not an ideal term, though. In the case of skeletons, more can be added but the skeleton stays. Though the SRC is clearly present in the simplest control systems, including likely early stages in the evolution of nervous systems, once we get to more complex cases the characteristic roles seen in the SRC get lost, at least in part. Most neurons can only "see" what other neurons are doing, and if they are close enough to sensory periphery to be responsive to the world itself, they are too far from the other periphery to cause actions that have consequences (Cao 2012).

There is also a more philosophical point here. The kind of content analyzed by the SR model is a *communicative* kind; its point is the bridging of some gap. *Thinking*, it seems, is not like that. Various kinds of "bridging" are needed in the machinery of the brain, but that is not the central task of a brain-like system. Thinking more fundamentally involves other things – finding solutions, learning, inferring. Communication between parts of the system is part of how that is done, but if there are messages being sent between the parts in order to get the job done, these are not messages which have the kinds of contents that the whole person's beliefs and thoughts have. So there might be a theory of "sub-personal" signaling between neurons or other parts of a brain, but we'd also need a further theory of how this sub-personal activity adds up to a situation where the whole person believes that Washington is south of New York, or that Obama will win the next election.

Some of this criticism is right, I think. But it is not the whole story. The sender-receiver model has at least one further role to play here, perhaps two roles. When I was introducing the model, I said that sender-receiver systems can bridge space or time (or both). The problem faced in thought is not essentially one of bridging parts of a system – *except* in the temporal case. Bridging time *is* something that a brain has to do. This is memory. Memory is usually seen as a matter of *storing* information. But it can also, and perhaps more accurately, be seen as the sending of messages, from a present stage to a future stage of oneself. This is person-level signaling, not sub-personal, at least some of the time. It is not one part of your brain sending something to another part. Or rather, it *is* a case of that, but it is a "temporal part," not a spatial part, that is sending something to another temporal part.

This is a link between thought and the sender-receiver model. But the link between a real case and the model may be shallow or deep, may be trivial or may tell us something. Which is it in this case?

Seeing memory as akin to communication is in some ways a very familiar idea. Many early theories of memory were based on the idea of *inscription*. This goes back to Plato and Aristotle, and was carried forwards after them. You first inscribe a memory, and later you read or otherwise perceive it. Inscribing-and-reading is an sending-and-receiving process. According to Mary Carruthers (1992) and Kurt Danziger (2008), who have worked on the history, this has been the "master metaphor" in Western thinking about memory since the time of Plato. There is an obvious analogy between memory and writing notes to oneself, using reminder pads and diaries, and that is emphasized by the inscription view of memory.

But the communicative view of memory has not been *worked* very hard from a philosophical perspective. I think that is partly because the general theories of communication that might have be applied have been too rudimentary, or unsuitable because they take thought for granted. I think the sender-receiver model is both abstract enough and good enough to be applied to memory in this way. Let's see how much we can get out of it.

4.1. Common interest

In the model, stable signaling requires at least partial common interest. Memory is signaling within an organism, so we expect a lot of concordance of interests. We expect the parts of an organism to cooperate biologically, for the most part, and this

leads us to also expect that any parts that can have preferences will have preferences that line up fairly well with each other. But there are possibilities of divergence. In the context of conscious choice there can certainly be divergence of interests across stages. You might believe now that your future self will do something that you do not want it to do. One way to constrain your future self would be to "tie yourself to the mast," as Ulysses did. Another way might be to starve your future self of needed information. That also raises the possibility of sending *bad* information.

Given the way memory works in us, this is not easy to do. It is hard to send your future self misinformation without believing it yourself. When you create a memory trace at t_1 for use at a later time, t_2, it is present in the intervening time and, in the psychological machinery that we have, it is not apparently possible to stash the trace away so it is not part of your belief system in the meantime. To bring about normal use of a memory trace in the future without believing it now requires some rupture between the temporal stages.

Science fiction stories often make use of these ruptures. Examples include Philip K. Dick's 1966 story "We Can Remember It for You Wholesale," and the movies based on it – two movies to date, both called *Total Recall*, 1990 and 2012. As the plot of the first movie has it, earlier Douglas Quade wants later Quade to infiltrate a resistance movement, so that earlier Quade can achieve his desire of destroying that movement. The only way to achieve the infiltration is to erase and fake his own memories, and then intervene later by means of recordings and co-conspirators to regain continuity between earlier and later Quade.

Once there is breakdown of common interest, sender and receiver do appear as different agents, and the sender-receiver structure becomes clearer. As I said, this tends to require a psychologically abnormal rupture between stages, or recourse to some external memory trace, such as a written record, that can be created and set aside.

5.2. Separation and the role of the reader

In a paradigm SRC, there is good separation between sender, receiver, and sign. One way for this situation to be lost is for the sign to be "swallowed up" by the receiver. Hold that thought.

In the history of theories of how the mind works, one basic hypothesis has been internal representation. We saw this in Plato's wax tablet model of memory,

and in Ramsey's metaphor about belief: beliefs are "maps by which we steer." A continual problem with representation-based views of thought has been regresses. If there is an inner representation then there also has to be an inner reader. If there is a little agent inside reading the signs, this agent has to be intelligent, it seems. So no explanation of intelligence is achieved by positing inner signs, as intelligence is assumed in the mechanism.

In the mid 20th century, this argument was especially powerful through the influence of Wittgenstein. This changed around the 1970s, not because of philosophical arguments, but because of something from outside. The rise of computer technology changed the situation. *Somehow* computers showed that a representational view of the mind is OK in principle, though I think it was not really clear exactly how.

Jerry Fodor's 1975 book *Language of Thought* was an influential defense of a representationalist approach. Daniel Dennett in 1977 wrote a review of the book, where he grappled with the regress problem. Dennett noted the history of the problem, and the power of the argument. Then he took the plunge.

> [N]othing is intrinsically a representation of anything; something is a representation only *for* or *to* someone; any representation or system of representations requires at least one *user* of the system who is external to the system. Call such a user an exempt agent. Hence, in addition to a system of internal representations, neo-cognitivism requires the postulation of an inner exempt agent or agents. . . .
>
> Hume wisely shunned the notion of an inner self that would intelligently manipulate the ideas and impressions, but this left him with the necessity of getting the ideas to "think for themselves". His associationistic couplings of ideas and impressions, his pseudo-chemical bonding of each idea to its predecessor and successor, is a notorious non-solution to the problem. Fodor's analogous problem is to get the internal representations to "understand themselves", and one is initially inclined to view Hume's failure as the harbinger of doom for all remotely analogous enterprises. But perhaps the *prima facie* absurd notion of self-understanding representations is an idea whose time has come, for what are the "data structures" of computer science if not just that: representations that understand themselves? [underlining added]

Dennett goes on to qualify this a little.[5] But the message I see him as taking from computers is that one need not worry about the idea that for every representation

[5] "But perhaps the *prima facie* absurd notion of self-understanding representations is an idea whose time has come, for what are the "data structures" of computer science if not just that: representations that understand themselves? In a computer, a command to dig goes straight to the shovel, as it were, eliminating the comprehending and obeying middleman. Not *straight*

there must be a reader, and the regress that threatens, because readers are not required for a system to run on representations. This, I think, is the standard way that philosophers and many others have become comfortable with the idea of inner representation. But there is another view of the situation. Here I draw on Randy Gallistel's recent work, especially his book with Adam King, *Memory and the Computational Brain* (2009). The alternative view is that it is true that regress problems with internal representation hypotheses are always serious problems, and true that computers defused them. But the way they did this is not by showing the possibility of self-reading representations, but by showing how to *embrace* the role of readers in mechanistic systems. A computer of the ordinary kind has a distinction in the hardware between processor and memory. Marks are *written* into memory and then *read*. Computers respect the asymmetries in the sender-receiver model, with time as the gap being bridged. Computers did not show the viability of un-read representations, but the power of mechanical systems with large memories coupled to simple readers and processors.

Gallistel and King go further. They think that behavioral evidence, in animals as simple as ants, as well as humans, shows that brains must actually contain a read-write memory of roughly the sort seen in computers. But mainstream neurobiology holds that the brain works *without* a clear separation between representations and the devices that read or use them. Mainstream neurobiology thinks we are not much like computers in this respect. Gallistel and King think we *must* be like computers in this respect, and neurobiology has not found the crucial mechanisms inside us yet.

It is interesting to put this argument in a larger historical context. The Ancients, as I noted, were attracted to a *write-store-read* model of memory. Some "associationist" theories, from the 18th century through to the 20th, tried to get rid of the difficult hypothesis of an inner reader. Gallistel thinks that this is a mistake – it was a mistake in people like Hartley and Hume, and just as much a mistake in contemporary neurobiology. For Gallistel, the ancient view might look crude, but it was on the right track – closer to the right track than neurobiology textbooks today.

---

to the shovel, of course, for a lot of sophisticated switching is required to get the right command going to the right tools, and for some purposes it is illuminating to treat parts of this switching machinery as analogous to the displaced shovellers, subcontractors and contractors. The beauty of it all, and its importance for psychology, is precisely that it promises to solve Hume's problem by giving us a model of vehicles of representation that function without exempt agents for whom they are ploys."

Perhaps the Ancients got lucky here, as some of them (different ones) did in the case of atomism.

A rival view goes like this: the ancients brought to memory an analogy with writing – in Plato's time, a fairly new technology in Greek life. Computer designers, from Turing and Von Neumann onwards, used the read-write mechanism, not only as an explanatory metaphor, but as the basis for a technology of huge importance. But our brains were always different. They, working with the unusual raw materials of living cells, merged the roles of representing and processing.

On the first of these views, the operation of memory is a paradigm SRC, laid out in time. On the second view, it is a more marginal case, as there is no separation between representations and readers.

To conclude I will briefly discuss one further link between the sender-receiver model and human thought, one that is not about memory. This concerns the kind of "internal representation" that is most familiar to all of us, as it is part of ordinary conscious experience. I have in mind *inner speech*, the flow of monologue or commentary that accompanies much of what we do.

You might wonder why this fact alone did not make the idea of internal representation of *some* kind viable. Maybe it did, but for many years this phenomenon was not seen as a big deal in psychology, or in philosophy. In late 20th century psychology, the focus of much work was sophisticated *un*conscious processing. I think that many saw the inner chatter as froth on the surface of intelligent processing. Recently though, there has been interest in inner speech as an important part of our minds, one with a role in the explanation of distinctive features of human thought. One aspect of this role is the integration and organizing of information.

Integration is often a difficult thing for brains to achieve. Many animals seem to have less integrated nervous systems than ours – sometimes smart and powerful, but different from ours, and with limitations. Integration in humans is probably achieved in a range of ways, but one way may involve the internalization of forms of representation that are derived from tools developed for social interaction, from public speech.

This idea goes back to the Soviet Russian psychologist Lev Vygotsky (1932), though he is not given much credit for it, and has been developed in new ways by

people like Peter Carruthers, Liz Spelke, and Andy Clark.[6] Language has two roles when it is internalized. First, it is a very flexible representational medium; it is a means for bringing together and organizing information from different sources. Second, it is a medium for inner "broadcast." A sentence can be constructed *as if* for speech, but routed back to the input end of the system, so it appears in "auditory imagination." Here it can be made available to other parts of the system for further use, including use in deliberate conscious reasoning, and the slow, serial "thinking things through" that we can do when the stakes are high.

Internal representations of this kind have personal-level contents, like "OK, now disconnect the power supply," rather than subpersonal ones, and the way these signs are "broadcast" gives them a clear place in a sender-receiver structure, though we are sending these signs to ourselves.

So in broadcast inner speech we have a form of inner signaling that fits the sender-receiver model, that is an evolutionary late-comer, and that came to us through the development of tools for social interaction. It is a within-agent application of a tool originating in between-agent cooperative interaction, and one that may have a special role in the explanation of unified, conscious human thought.

At the beginning of this talk I described an unhappy choice that arose between ant-based and people-based approaches to cognition and action. The sender-receiver model has, I think, something to say about both the ants and the people too.

(Partial) References

Bergstrom, C. and M. Rosvall (2009). "The transmission sense of information." *Biology and Philosophy*. 26: 159-176

Cao, R. (2012). "A Teleosemantic Approach to Signaling in the Brain." *Biology and Philosophy.*

Carruthers, M. (1992). *The Book of Memory: A Study of Memory in Medieval Culture*. Cambridge University Press.

Carruthers, P. (2002). "The cognitive functions of language.'' *BBS*.

Crawford, V. P., and Sobel, J. (1982). "Strategic information transmission." *Econometrica* 50: 1431-1451.

Danziger, K. (2008). *Marking the mind: A history of memory.* New York: Cambridge University Press.

Dawkins. R. and Krebs. J. (1978). "Animal signals: information or manipulation?" In J. Krebs and R. Davies, eds., *Behavioural Ecology: an evolutionary approach*. Oxford: Blackwell, pp. 282–309.

Dennett, D. (1977). Review of Fodor's *The Language of Thought*, in *Mind*.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

Dyson, F. (2011). "How We Know" *New York Review of Books*. March 10th, 2011.

[6] I am indebted to Kritika Yegnashankaran (2010) for introducing me to a number of these ideas, including Vygotsky's contributions.

Fodor, J.A. (1975). *The Language of Thought*.
Gallistel, C. R. and King, A. P. (2009). *Memory and the Computational Brain: Why Cognitive Science Will Transform Neuroscience*. New York, NY: Wiley/Blackwell Press.
Godfrey-Smith, P. (forthcoming). "Information and Influence in Sender-Receiver Models, with Applications to Animal Behavior." To appear in *Animal Communication Theory: Information and Influence*, edited by Ulrich Stegmann, Cambridge UP.
Harms, W. (2004). "Primitive Content, Translation, and the Emergence of Meaning in Animal Communication," in D.K. Oller and U. Griebel (eds.), *Evolution of Communication Systems: A Comparative Approach.* Cambridge: MIT Press, 2004, 31-48.
Huttegger, S., Skyrms, B., Smead, R., Zollman, K. (2010). "Evolutionary dynamics of Lewis signaling games: signaling systems vs. partial pooling." *Synthese* 172: 177–191.
Lewis, D. K. (1969). *Convention*. Cambridge, MA: Harvard University Press.
Millikan, R. G. (1984). *Language, Thought and Other Biological Categories.* Cambridge, MA: MIT Press.
Owren, M., Rendall, D., and Ryan, M. (2010) "Redefining animal signaling: influence versus information in communication." *Biology and Philosophy* 25: 755–780.
Ramsey, F. P. (1929). "General Propositions and Causality."
Robson, A. (1990). "Efficiency in Evolutionary Games: Darwin, Nash and the Secret Handshake", *Journal of Theoretical Biology* 144: 379-396.
Shannon, C. (1948). "A Mathematical Theory of Communication." *The Bell System Mathematical Journal* 27: 379-423.
Spelke, L. (1996). "What Makes Us Smart."
Sperber, D. and D. Wilson (1986). *Relevance*: *Communication and Cognition*.
Skyrms, B. (1996). *Evolution of the Social Contract*. Cambridge, MA: Cambridge University Press.
–––––. (2010). *Signals: Evolution, Learning, & Information.* New York, NY: Oxford University Press.
Vygotsky, L. (1932/1986). *Thought and Language.*
Yegnashankaran, K. (2010). *Reasoning as Action*. PhD Dissertation, Harvard University.