Review of Brian Skyrms' *Signals*

Peter Godfrey-Smith
City University of New York

The final version appears in *Mind* 120 (2012): 1288-1297. doi: 10.1093/mind/fzs002

Signals: Evolution, Learning, and Information, by Brian Skyrms. Oxford: Oxford University Press, 2010. Pp. vii+195. H/b £30.00, $60.00; P/b £14.99, $27.00.

Suppose there are two agents. A *sender* can see the world but not act except to create signs of some kind that can be seen by a second agent. This *receiver* can act, but can only see signs sent by the sender. Actions by the receiver have consequences for both parties, and the two parties agree on which acts are good in each state of the world. By means of rational choice and common knowledge, agents such as these can maintain a sign system that seems to have at least rudimentary semantic properties.

This is David Lewis' model of conventional signaling, developed in his dissertation and 1969 book *Convention*, and intended as a reply to W.V. Quine's skeptical treatment of meaning. The model had limited influence on naturalistic philosophy, in part because Lewis presupposes rational agents who have thoughts with intentional properties. In a brief chapter in his 1996 book *Evolution of the Social Contract*, Brian Skyrms showed that evolution by natural selection, as well as rational choice, can give rise to signaling systems of the kind Lewis described, and something like Lewisian conventional signaling can exist in agents who are much simpler than humans.

This model was followed up in Skyrms' next book, *The Stag Hunt* (2004), a sustained look at the evolution of cooperation. There, signaling was treated as a means to coordination in joint projects. In his third book in this series, *Signals*, Skyrms puts the Lewis model at center stage. The book is entirely about the natural emergence of signaling behaviors – and with them, signs with content – as a result of biological evolution, learning, and other adaptive processes.

Signals is an excellent book. It is compact, clear, adventurous, and wide-ranging. Though it is full of models, it does not assume technical knowledge on the part of the reader. Skyrms distills the models' results into an accessible form (with some exceptions, especially in cases where the models are newer). Skyrms also looks back over the centuries to make connections to earlier thinkers, sometimes just to pick up an echo and sometimes because he thinks the earlier writer was glimpsing something that can now be understood in more detail. So *Signals* is dotted with quotes from Democritus, Rousseau, Hume, Smith, and Kant, alongside anecdotes about monkeys, birds, and bacteria. In contrast, a kind of impatience is occasionally visible when the book encounters themes that preoccupy contemporary philosophers interested in content. The book is very focused on the models, and quick to shoulder through philosophical niceties. Skyrms seems to think that with some exceptions such as Dretske, recent philosophers have not done a good job on these issues and it is better to start from scratch.

This approach has been very fruitful for Skyrms. But in a review it is appropriate to examine connections to other work, and perhaps revisit a few shouldered niceties. This also leads, I will suggest, to new possibilities within the modeling project. Before discussing details, though, I will make some general comments about the importance of the book.

Skyrms' book both contributes to, and throws into focus, a body of work that has developed and coalesced in an unobvious way over several decades. This body of work has dual origins: Shannon (1948) and Lewis (1969). When Shannon introduced mathematical information theory, he did so within what amounts to a *sender-receiver* framework. What Shannon called a "general communication system" has a "transmitter" and a "receiver," where the transmitter is sensitive to an information "source" of some kind. For Shannon, information is carried by a signal whenever the signal reduces uncertainty about the source. Lewis, in 1969, gave the model of conventional signaling outlined in my first paragraph. If a sender and a receiver have shared interests, informative signaling can exist as an equilibrium state maintained by rational choice. These two contributions fit together in a way that becomes clear in retrospect: Shannon took for granted the sender and receiver roles, and gave a theory of the properties of channels that could 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.

A "second generation" can be recognized in the philosophical literature in the 1980s, with something like the same division. Dretske (1981) brought Shannon's concept of information into philosophy of mind, but with little attention to the sender and receiver roles. Millikan (1984) gave a naturalistic account of sender-receiver (or "producer-consumer") coordination, with no formal component and little attention to information. Skyrms puts many of the pieces together: we have a Lewis-style model of the stabilization of sender-receiver behaviors, but the model is embedded within information theory, and recast in a naturalistic framework.[1]

Much of *Signals* is about the details of the processes by which sender and receiver behaviors are produced and stabilized. The book compares how this works in biological evolution and learning, and moves from simple sender-receiver configurations to more complicated networks of interaction. Much attention is paid to factors that obstruct the evolution of signaling systems, such as situations where one state of the world is much more probable than another, reducing the benefits of signaling. Here I will focus on the themes most directly relevant to philosophical work on meaning.

Suppose we have a signaling system in place. A sender is producing distinctive signals in different states of the world, making it possible for the receiver to coordinate actions with these states, leading to payoffs for both sides and the consequent stabilization of these behaviors by an evolutionary process. Assume that the agents involved are cognitively unsophisticated. What sorts of properties do the signals have just as a result of this set-up? Do they have *content* of some kind? And if so, what is that content?

Skyrms draws on the information theory of Shannon and others. People writing about information theory often say that it is antagonistic to the idea of meaning or content. For example, Freeman Dyson in a review of James Gleick's book *The Information* (both 2011) says that the "central dogma" of information theory is that "meaning is irrelevant." It is true that much of information theory can proceed without paying attention to the specific messages being sent over an information channel, but

[1] The connection between Millikan's and Skyrms' frameworks is discussed in Harms (2004).

there is no point in maintaining and using the channel unless the messages sent do bear on something in the world, and can guide actions or inferences of some kind. Skyrms thinks that information theory provides us with a way of directly explaining meaning. For a signal to *carry information about* the state of some part of the world is for it to *change the probabilities* of states of that part of the world. Suppose the world can be in three different states, S_1, S_2, and S_3. These states have probabilities independent of the state of the signal. Given a sender's dispositions – both the rule the sender follows in making signals and any imperfections in their perception or production – the probabilities of the states given the signal can differ from the unconditional probabilities of the states. (This is all Skyrms means when he talks of signals "changing" those probabilities.) For Skyrms, the *quantity of information* in a signal is measured by how far it moves the probabilities. He favors the "Kullback-Leibler divergence" as a measure of how much the probabilities are moved. Skyrms thinks this is not only a good representation of the *size* of the effect that a signal has on probabilities, but also provides a way to analyze the *content* of particular signals. For Skyrms, the *informational content* of a signal is the vector (or list) of all the changes made to the probabilities of states of the world by the signal. If $P(S_i)$ is the probability of state S_i independent of the signal, and $P_{sig}(S_i)$ is the probability of that state conditional on the signal, then in a simple three-state example, the informational content of a signal is given by:

$$<\text{Log}_2[P_{sig}(S_1)/P(S_1)], \text{Log}_2[P_{sig}(S_2)/P(S_2)], \text{Log}_2[P_{sig}(S_3)/P(S_3)]>$$

The amount of information in the signal is given by the weighted average of these, which is the Kullback-Leibler divergence between the distribution in the light of the signal and the distribution independent of the signal: $\Sigma_i P_{sig}(S_i)\text{Log}_2(P_{sig}(S_i)/P(S_i))$. What is important for Skyrms is not so much the way the Kullback-Leibler measure operates, with its fractions and logs to the base 2, but the idea that informational content is not always propositional – not always expressible with a "that..." clause. Propositional content is a special case, present when the signal reduces the probability of some states to zero.

For Skyrms, signals also not only carry information about states of the world, in virtue of the sender's dispositions, but carry information about the acts performed as well, in virtue of the *receiver's* dispositions. The same kind of content is present.

There are several things that seem problematic, or at least might be improved, here. First I will look at a signal's content about the state of the world. A natural reaction to Skyrms' view is that it seems wrong to say that the content of a signal is a matter of how the signal has *changed* the probabilities of states. Instead, a signal's content is a matter of how the world *is*, or *probably is*. This might seem to be merely an intuitive judgment, but I think it is more than that. We can ask: what is the *point* of signaling? What does a receiver *get* from a signal? What is it about a signal that makes it a good guide to action? If a receiver can learn from a signal how the world probably is, it is of no use to also learn how the probabilities have changed. And if a receiver does not learn from a signal how the world probably is, the signal cannot be used to guide action. Within this view, Skyrms' point about vectors and propositional content can still be made. The content could be seen as the list of all the probabilities about states of the world that hold in the light of the signal.

There is also a problem with this alternative view. If the content is the list of probabilities of states in the light of the signal, then a signal can apparently carry *no information* (if the new probabilities are same as old), but still have a definite *informational content*. One response to this would be to say that if *none* of the signals that might be sent on an occasion change the probabilities of the states, then these are not "signals" at all, and if only *some* of the signals that might be sent leave the probabilities unchanged, that is not a problem for the analysis. If the chance of rain independent of the weather report is 50/50, that does not stop a weather report from sometimes telling you the chance of rain is 50/50, as long as other possible reports would tell you something else.

Skyrms' view, where the content of a signal is given by the changes it makes to probabilities, has what he sees as a "seamless integration" with information theory (p. 42), but this is a view of content that does not fit well with what receivers get from senders and what guides their action. The alternative view, that the content is the list of post-signal probabilities, may have a more awkward connection to the amount of

information in a signal. There is probably no need to choose one view, saying that such-and-such is *the* content. The familiar philosophers' language of "content" is potentially misleading, as it suggests that a meaning is either *in* a sign or not in it. The metaphor of containment is probably a bad one. A signal may have many relations to the world, relevant in different contexts. These will include relations that involve more than just changes to probabilities, such as relations that have a special role in the maintenance of the sender-receiver configuration. A signal might do little to the probability of S_2, and raise the probability of S_3 much more, while it is the link to S_2 that explains why the relevant sender and receiver dispositions have been stabilized. Then the signal has a special kind of involvement with S_2, despite the weak probabilistic connection.[2]

Skyrms, as I noted, recognizes informational content about acts as well as about states of the world. But again, the content of a signal should have something to do with what makes it useful to a receiver, and what is the point of telling a receiver how the probabilities of its own acts have changed? Instead, it seems that signals in simple system like these have an *imperative* content. They say: do X! In Millikan's theory (1984, 2004), simple signs have "pushmi-pullyu" contents, they both tell the consumer of the sign how the world is and instruct them what to do. Harms (2004), Huttegger (2007) and Zollman (2011) have all offered accounts of how a signal in a Lewis-Skyrms game might acquire content that is purely indicative or imperative.

A propositional content can be true or false in a particular situation. On some of the days the newspaper says it will rain, it rains, and on other days it doesn't. This might be a peculiarity of propositional content, but something like this might be applicable generally. In Skyrms' framework, can a signal's content be right or wrong, accurate or inaccurate, on an occasion?

Skyrms recognizes a notion of *mis*information. If a signal moves the probabilities of states "in the wrong direction" — either by diminishing the probability of the actual state, by raising the probability of a non-actual state one, or both, then it contains misinformation. So there are two kinds of information in the story, though one kind is not named by Skyrms. *Accurate* information (or *eu-information*?) is present when the probability of the actual state is increased by the signal or the probability of a non-actual

[2] See Shea, Godfrey-Smith, and Cao (in preparation) for further development of these ideas.

one is reduced; there is *misinformation* if the probability of the actual state is reduced or of a non-actual state increased. Many signals will contain both kinds of information.

Another approach is possible. Suppose the content of the signal is given by the probabilities of states of the world in the light of the signal, as I said above. The content is a list of entries like: <0.2, 0.5, 0.3>, for states 1, 2, and 3. The actual state of the world can also be given as a distribution in the same form. If state 2 is actual, this would be: <0, 1, 0>. The distance between these two distributions can give a measure of how close the content of the signal is to the state of world. This measure could be taken in various ways. Skyrms favors the Kullback-Leibler divergence, and that could be used here. Starting from the truth and asking how far away the content of a signal was, the Kullback-Leibler divergence reduces to
$-\text{Log}_2 P_{Sig}(S_A)$, where S_A is the state of the world that is actual on that occasion. This has a minimum value of zero when $P_{Sig}(S_A)=1$, and no upper bound. In my example above, the distance between the truth and the signal's content would be 1. What the signal said was one bit of information away from the truth. This in effect treats a signal's content as an *estimate* of the truth. [3]

I will discuss one other part of Skyrms' treatment of content. The Lewis model assumed *common interest* between sender and receiver. For each state of the world, the two parties have the same preferences about actions the receiver might perform. Signaling cannot be stabilized if sender and receiver have c*omplete conflict of interest*: for each state of the world, sender and receiver have reversed preference orderings for acts the receiver might produce. If the sender provides an informative signal, the receiver will use it to get results opposed to the sender's interests. And if the receiver makes their behavior contingent on a signal, they can be exploited by the sender. Any equilibrium will be one where the sender does not send anything informative and the receiver does not attend to signals.

[3] In James Joyce's non-pragmatic justification of Bayesian belief-management (1998), a subject's degrees of belief in various propositions are treated as estimates of the truth values (1 or 0) of those propositions.

With complete common interest and complete conflict of interest marked out, we see there are many ways to have *partial* common interest. Sender and receiver might agree on what is the best action in every state, but not on what is the worst. They might agree on their preference orderings of actions for some states of the world, but not others. Complete common interest can be transformed into complete conflict of interest along many different paths.

Here is a general idea that Skyrms looks at, but in a way that might be improved: informative signaling is viable to the extent that there is common interest. This idea has been explored in economics. Crawford and Sobel, summarizing a famous 1982 model using a different framework, say "equilibrium signaling is more informative when agents' preferences are more similar." As the varieties of partial common interest distinguished above show, however, the extent of common interest in a Lewis-Skyrms model is not measured on a simple scale. So it is possible to explore different ways in which partial conflict of interest gives rise to different kinds of imperfect signaling outcomes (Blume et al. 2001).

Skyrms discusses this in the context of *deception*. He says that in situations of "mixed" interests, "what we should expect is some combination of information and misinformation." A signal carries misinformation, as noted above, if it moves probabilities of states in a way that makes the probability of the actual state lower or that of a non-actual state higher. "Deception" is present "when misinformation is sent systematically and benefits the sender at the expense of the receiver" (p. 80). An example of what Skyrms has in mind is the interesting case shown in Table 1 (from Skyrms p. 81). Assume the states of the world are equally probable.

		Acts		
		A_1	**A_2**	**A_3**
	S_1	2,10	0,0	10,8
States	**S_2**	0,0	2,10	10,8
	S_3	0,0	10,10	0,0

Table 1: A case of partial common interest, from Skyrms.

In each entry of Table 1, the first number gives the payoff to the sender for that action in that state of the world, and the second number gives the payoff to the receiver. Assume an agent is equally likely to take the sender or receiver role on any occasion. Here is a strategy: if you are the sender, then send message 1 in states S_1 and S_2 and send message 2 in S_3; if you are the receiver, then do act A_3 when you see message 1 and A_2 when you see message 2. This is an equilibrium. Skyrms says: "In this equilibrium, the occupant of the sender's role always manipulates the occupant of the receiver's role. In state 1, the sender's signal is a half-truth in that it raises the probability of state 2. In state 2, the sender's signal is a half-truth in that it raises the probability of state 1. These half-truths induce the receiver to choose act 3 in states 1 and 2, whereas accurate knowledge of the state would lead her to choose either act 1 or act 2." So we have "universal deception at equilibrium" (p. 81).

I don't think message 1 is a "half-truth," in the sense usually associated with that term. What the sender is doing is refusing to tell the whole truth, but what is said is simply true. The sender is saying something logically weaker than what they know. Expressed propositionally, the signal message 1 says "State 1 or 2 is actual." To say something logically weaker than what you might say is not to deceive. To tell half the truth is not to tell a half-truth.

Is this is a mere fact about how we habitually describe some cases? It is more than that. Skyrms' treatment is guided by examples involving animals. He discusses a case (drawn from Cheney and Seyfarth 1990) where a male vervet monkey gives fake alarm calls to interfere with the arrival of a rival male. But in a case like that, a standardized signal is maintained for a definite reason, while there is also leeway for exploitative uses. There is a difference between the *maintaining* and the *non-maintaining* uses of the signal. Some uses contribute to stabilization of the sender-receiver configuration and some, if more common, would undermine it. Those ones are deceptive. The same applies to another case, where fireflies of one species use the mating signal of another species to lure males in to be eaten; again there are maintaining and non-maintaining uses of the signal. This distinction does not exist in the Table 1 case, though. All signals sent in S_1 and S_2 play a "maintaining" role as much as any others do.

Setting deception aside, Skyrms' case is one where, due to partial common interest, signals sent at equilibrium are less informative than they might be. We can make a comparison of signals *within* this case, as well as the comparison *between* this case and others. When the world is in S_3, the sender's and receiver's interests are aligned, and the sender sends a unique signal which enables perfect coordination of action with the world. When the world is in S_1 or S_2, a worst option is being avoided in each case, but common interest is partial and a less-informative signal is sent. Expressed propositionally, that signal says "S_1 or S_2 obtains." In state 3 a higher-content signal is sent: "S_3 obtains." The mixture of matched and mis-matched interests is reflected in the amount of information contained in different signals.

The density of ideas in the 180 pages of *Signals* is such that an adequate review might be as long as the book. It discusses the evolution of systems which perform logical operations, the evolution of networks in which a single receiver makes use of many signalers, and the evolution of network structure itself – the forming and breaking of bonds of information transmission. Skyrms' book is relevant to debates about mental representation. He makes brief comments here, saying that neural interactions count as signaling in his sense. It is not clear that neurons in brains like ours do fit the Lewis-Skyrms model (Cao 2011), but the sender-receiver division corresponds at least roughly to the division between perception and action, between sensors and effectors, in a simple cognitive system. A further dimension might be recognized: *memory* is the sending of messages across time rather than space, the sending of messages from a present to a future self. For some kinds of memory this link to the sender-receiver model might be merely a weak analogy, but Gallistel and King (2010) have recently argued for the empirical importance of a form of memory – the "read-write memory" – that I think makes the connection between memory and the sender-receiver framework quite substantial.

Signals opens up many projects and theoretical directions. A slogan might be offered: a theory of meaning is a theory of sender-receiver coordination. From this point of view, many earlier approaches to meaning have been one-sided, focusing on either the expressive side or the interpretive side of an essentially two-sided set-up. Skyrms'

naturalization and extension of the Lewis sender-receiver model is one of the most exciting developments in recent philosophy.

* * *

Acknowledgment: For helpful discussions I am grateful to Rosa Cao and Nick Shea, participants at an "Author Meets Critics" session on Skyrms' book at the 2011 Pacific APA, and the CUNY Graduate Center discussion series on the evolution of communication.

References

Blume, A., DeJong, D. V., Kim, Y., and Sprinkle, G. B. (2001). "Evolution of Communication with Partial Common Interest." *Games and Economic Behavior* 37: 79-120.

Cao, R. (2011). "A Teleosemantic Approach to Information in the Brain." *Biology and Philosophy* 15: 1-23.

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

Cheney, D. L., and Seyfarth, R. M. (1990). *How Monkeys See the World.* Chicago, IL: University of Chicago Press.

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.

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.

Gleick, J. (2011). *The Information*: *A History, a Theory, a Flood.* New York, NY: Pantheon Press.

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. (2007). "Evolutionary Explanations of Indicatives and Imperatives." *Erkenntnis* 66 (3): 409-436.

Joyce, J. (1998). "A Non-Pragmatic Vindication of Probabilism," *Philosophy of Science* 65: 575-603

Lewis, D. K. (1969). *Convention*. Cambridge, MA: Harvard University Press.

Millikan, R. G. (1984). *Language, Thought and Other Biological Categories.* Cambridge, MA: MIT Press.

---. (2004). *Varieties of Meaning: The Jean Nicod Lectures.* Cambridge, MA: MIT Press.

Papineau, D. (1984). "Representation and Explanation," *Philosophy of Science* 51: 550-572.

Shannon, C. (1948). “A Mathematical Theory of Communication.” *The Bell System Mathematical Journal* 27: 379-423.
Shea, N., Godfrey-Smith, P., and Cao, R. (in preparation). “Content in Simple Signaling Games.”
Skyrms, B. (1996). *Evolution of the Social Contract*. Cambridge, MA: Cambridge University Press.
---. (2004). *The Stag Hunt and the Evolution of Social Structure.* Cambridge, MA: Cambridge University Press.
---. (2010). *Signals: Evolution, Learning, & Information.* New York, NY: Oxford University Press.
Zollman, K. (2011). "Separating Directives and Assertions Using Simple Signaling Games." *Journal of Philosophy* 108: 158-169.