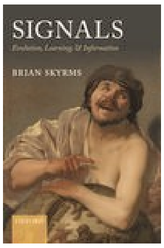

Signals: Evolution, Learning, and Information

Brian Skyrms

https://doi.org/10.1093/acprof:oso/9780199580828.001.0001
Published: 08 April 2010 **Online ISBN:** 9780191722769 **Print ISBN:** 9780199580828

Search in this book

CHAPTER

3 3 Information

Brian Skyrms

https://doi.org/10.1093/acprof:oso/9780199580828.003.0004 Pages 33–47
Published: April 2010

Abstract

This chapter shows that information is carried by signals. It flows through signaling networks that not only transmit it, but also filter, combine, and process it in various ways. We can investigate the flow of information using a framework of generalized signaling games. The dynamics of evolution and learning in these games illuminate the creation and flow of information.

Keywords: signals, signaling games, information flow

Subject: Philosophy of Science, Epistemology, Philosophy of Language

Collection: Oxford Scholarship Online

> "In the beginning was information. The word came later."
>
> Fred Dretske, *Knowledge and the Flow of Information* (1981)

Epistemology

Dretske was calling for a reorientation in epistemology. He did not think that epistemologists should spend their time on little puzzles[1] or on rehashing ancient arguments about skepticism. Rather, he held that epistemology would be better served by studying the flow of information. Although we may differ on some specifics, I am in fundamental agreement with Dretske.

Information is carried by signals. It flows through signaling networks that not only transmit it, but also filter, combine, and process it in various ways. We can investigate the flow of information using a framework of generalized signaling games. The dynamics of evolution and learning in these games illuminate the creation and flow of information.

p. 34

Information

What is the information in a signal? There are really two questions: *What is the informational content of a signal?* and *What is the quantity of information in a signal?*

Some philosophers have looked at information theory and have seen only an answer to the question of quantity. They do not see an answer to the question of content—or, to use a dangerous word,*meaning*—of a signal. As a result they move to a semantic notion of information, where the informational content in a signal is conceived as a proposition. The information in a signal is to be expressible as "the proposition that___." Signals then, in and out of equilibrium, are thought of as the sorts of things that are either true or false. Dretske takes that road and, as he himself says, it reduces the role of information theory to that of a suggestive metaphor. Others have followed his lead.

I believe that we can do better by using a more general concept of informational content. A new definition of informational content will be introduced here. Informational content, so conceived, fits naturally into the m:mathematical theory of communication and is a generalization of standard philosophical notions of propositional content.

The *informational content* of a signal consists in how the signal affects probabilities. The *quantity of information* in a signal is measured by how far it moves probabilities. It is easy to see the difference. Suppose, for instance, that there are two states, initially equiprobable. Suppose that signal A moves the probabilities to 9/10 for state 1 and 1/10 for state 2, and that signal B moves the probabilities in exactly the opposite way: 1/10 for state 1 and 9/10 for state 2. Even without knowing exactly how we are going to measure quantity of information, we know by considerations of symmetry that these two signals contain the same amount of information. They move the initial probabilities by the same amount. But they do not have the same *informational content*, because they
p. 35 move the initial probabilities in different directions. ↳ Signal A moves the probability of state 1 up; signal B moves it down.

The key to information is moving probabilities. What probabilities? We use the framework of a sender-receiver signaling game with evolving strategies.[2] That means that we are interested in information not only in equilibrium, but also before interactions have reached equilibrium. It is part of the structure of the game that the states occur with certain probabilities. The probabilities of sender and receiver strategies change over time. In learning dynamics, these probabilities are modified by the learning rule; in evolution they are interpreted as population frequencies changing by differential reproduction. At any given time, in or out of equilibrium, all these probabilities are well defined. Taken together, they give us all the probabilities that we need to assess the content and the quantity of information in a signal at that time.[3] Informational content evolves as strategies evolve.

How should we measure the *quantity of information* in a signal? The information in the signal about a state depends on a comparison of the probability of the state given that this signal was sent and the unconditional probability of the state. We might as well look at the ratio:

$$\mathrm{pr}_{\mathrm{sig}}(\mathrm{state})/\mathrm{pr}(\mathrm{state})$$

where $\mathrm{pr}_{\mathrm{sig}}$ is the probability conditional on getting the signal. This is a key quantity.[4] The way that the signal moves the probability of the state is just by multiplication by this quantity.

But when a signal does not move the probability of a state at all—for instance if the sender sends the same
p. 36 signal in all states—the ratio ↳ is equal to one, but we would like to say that the quantity of information is zero. We can achieve this by taking the logarithm to define the quantity of information as:

$$\log\left[pr_{sig}(state)/pr(state)\right]$$

This is the information in the signal in favor of that state. If we take the logarithm to the base 2, we are measuring the information in *bits*.

A signal carries information about many states, so to get an overall measure of information in the signal we can take a weighted average, with the weights being the probabilities of the states conditional on getting the signal:

$$I_{states}(signal) = \sum_i pr_{sig}(state\ i) log[pr_{sig}(state\ i)/pr(state\ i)]$$

This is the average information about states in the signal. It is sometimes called the Kullback–Leibler distance,[5] or the information gained. All this was worked out over 50 years ago,[6] shortly after Claude Shannon published his original paper on information theory. It goes under a slightly different name, *the information provided by an experiment*, in a famous article by Dennis Lindley.[7] Receiving a signal is like looking at the result of an experiment. Alan Turing used almost the same concept in his work breaking the German Enigma code during World War II.[8]

For example, consider our simplest signaling game from Chapter 1, where there are two states, two signals and two acts, with the states equiprobable. A signal moves the probabilities of the states, and how it moves the probability of the second state is determined by how much it moves the probability of the first, so we can plot the average information in the signal as a function of the probability of the first state given the signal. This is shown in figure 3.1:

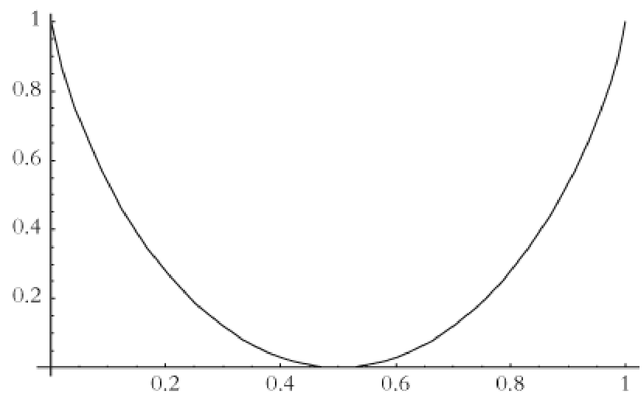

Figure 3.1: Information as a function of probability of state 1 given signal, state initially equiprobable.

 If the signal does not move the probability off one-half, the information is 0; if it moves the probability a little, there is a little information; if it moves the probability all the way to one or to zero, the information in the signal is one bit. In a signaling-system equilibrium, one signal moves the probability to one and the other moves it to zero, so each of the two signals carries one bit of information.

The situation is different if the states are not initially equiprobable. Suppose that the probability of state 1 is 6/10 and that of state 2 is 4/10. Then a signal that was sent only in state two would carry more information than one that only came in state one because it would move the initial probabilities more, as shown in figure 3.2:

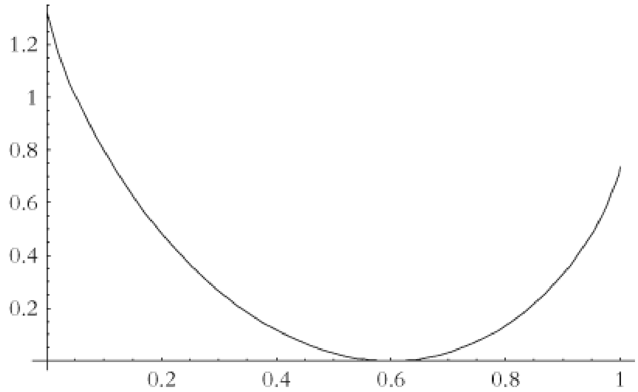

Figure 3.2: Information as a function of probability of state 1 given signal, state 1 initially at probability of .6.

In a game with four equiprobable states a signal that gives one of the states probability one carries two bits of information about the state. Compare a somewhat more interesting case from Chapter 1, where nature chooses one of four states by independently flipping two fair coins. Coin 1 determines up or down—let us say—and coin 2 determines left or right. The four states, up-left and so on, are equiprobable. There are now two senders. Sender 1 can observe only whether nature has chosen up or down; sender 2 observes whether it is left or right. Each sends one of two signals to the receiver.

p. 38

$\bullet\!\rightarrow\!\bullet\!\leftarrow\!\bullet$

The receiver chooses among four acts, one right for each state.

In an optimal signaling system equilibrium for this little signaling network, pairs of sender signals identify each of the four states with probability one—and the receiver makes the most of the information in the signals. In such a signaling system each signal carries one bit of information. One bit from each of the senders adds up to the two bits we had with one sender and four signals. This is a m:mathematical convenience of having taken the logarithms to the base 2.

Information about the act

All of the information discussed so far is defined by the probabilities with which nature chooses acts and the probabilities of the sender strategies. But there is also a different kind of information in the signals. We have been discussing *information about the state of nature*, but there is also *information about the act* that will be chosen. The definitions are entirely analogous to those of information about the state.

p. 39 Taken together, probabilities of the states, probabilities of sender's strategies, and probabilities of receiver's strategies give us unconditional probabilities of the acts. Just add up the probabilities of all combinations that give the act in question its initial probability. Probabilities of receiver's strategies alone give us probabilities of acts given a certain signal. The information in the signal is now measured by how much the signal moves the probabilities of the acts. The average information about the act in a signal is:

$$\mathrm{I_{acts}(signal)} = \textstyle\sum_{\mathrm{i}} \mathrm{pr_{sig}(act\ i)\ log\ [pr_{sig}(act\ i)/pr(act\ i)]}$$

The definition has just the same form and rationale as the definition of information about the state. There are thus two kinds of information in a signal, and two quantities summarizing amounts of information in a signal.

The two quantities need not be the same. For instance, suppose that the sender chooses a different signal for each state but the receiver isn't paying attention and always does the same act. Then there is plenty of information about the states in the signals, but zero information about the acts. Conversely, suppose that the sender chooses signals at random but the receiver uses the signals to discriminate between acts. Then there is zero information about the states in the signals, but there is information about the acts. There may be more states than acts or more acts than states. It is only in special cases where the two quantities of information are the same.

Creation of information in a signal

Let us reflect on what was shown in Chapter 1. Evolution can *create* information. It is not simply a question of learning to use information that is lying around, as is the case when we observe a fixed nature. With natural signs—smoke means fire—the information *about states* is just there, and we need to learn how to utilize it.
p. 40 Nature is *not* playing a game and does not have alternative ↳ strategies. Information *about acts* arrives on the scene when we learn to react appropriately to the information about states contained in the smoke. But in signaling games, there may be no initial information about acts or states in the signals. Senders and receivers may just be acting randomly. When evolution (or learning) leads to a signaling system, information is created. Symmetry-breaking shows how information can be created *out of nothing*. Figure 3.3 shows the creation of information about states by reinforcement learning in a two-state, two-signal, two-act signaling game.

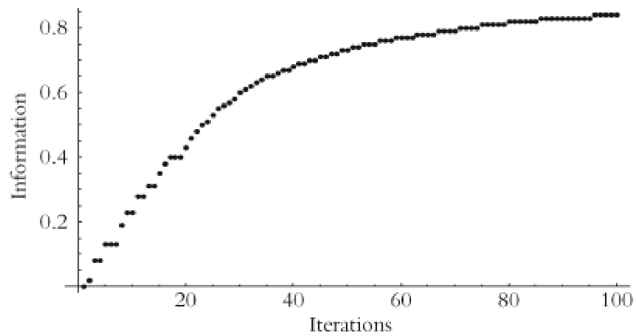

Figure 3.3: Creation of information *ex nihilo* by reinforcement learning.

Informational content

Now that we know how to measure the quantity of information in a signal, let us return to *informational content*. This is sometimes supposed to be very problematic, but I think that it is remarkably straightforward. Quantity of information is just a summary number—one bit, two bits, etc. *Informational content must be a vector*.[9]

p. 41 Consider the information in a signal about states, where there are four states. The informational content of a signal tells us how the signal affects the probabilities of each of the four states. It is a vector with four components, one for each state. Each component tells us how the probability of that state moves. So we can take the *informational content* of a signal to be the vector:

$$< \log[\mathrm{pr}_{\mathrm{sig}}(\mathrm{state}\ 1)/\mathrm{pr}(\mathrm{state}\ 1)], \log[\mathrm{pr}_{\mathrm{sig}}(\mathrm{state}\ 2)/\mathrm{pr}(\mathrm{state}\ 2)], \ldots >$$

The *informational content about acts* in the signal is another vector of the same form.

Suppose that there are four states, initially equiprobable, and signal 2 is sent only in state 2. Then the informational content about states of signal 2 is:

$$\mathrm{I}_{\mathrm{States}}(\text{Signal } 2) = < -\infty, 2, -\infty, -\infty >$$

The -∞ components tell you that those states end up with probability zero. (The -∞ is just due to taking the logarithm—no cause for alarm.) The entry for state 2 tells you how much its probability has moved. If the starting probabilities had been different, this entry could have been different. For instance, if the initial probability of this state had been 1/16 with everything else the same, the information about states in signal 2 would have been:

$$\mathrm{I}_{\mathrm{States}}(\text{Signal } 2) = < -\infty, 4, -\infty, -\infty >$$

"*Wait a minute*," someone is sure to say at this point. "*Something very important has been left out!*" What is it? "*But shouldn't the content—at least the declarative content—of a signal be a proposition? And isn't a proposition a set of possible worlds or situations?*"

Suppose a proposition is taken to be a set of states. (States can be individuated finely, and there can be lots of states if you please.) It asserts that the true state is a member of that set. A proposition can just as well be specified by giving the set of states that the true state is not in. That is what the -∞ components of the information vector do. If a signal carries propositional information, that information can be read off the p. 42 informational content vector. For instance, if the ↳ signal "tells you" that it is "state 2 or state 4" in our example, then the content vector will have the form:

$$\mathrm{I}_{\mathrm{States}}(\text{Signal } 2) = < -\infty, \_, -\infty, \_ >$$

with the minus infinity components ruling out states 1 and 3, and the blanks being filled by numbers specifying how the probabilities of state 2 and 4 have moved.

That is to say that the familiar notion of propositional content as a set of possible situations is a rather special case of the much richer information-theoretic account of content. This vector specifies more than the propositional content. Furthermore, some signals will not have propositional content at all. This will be typical in out-of-equilibrium states of the signaling game. It is the traditional account that has left something out.

Notice that the *quantity* of information in a signal—as measured by Kullback and Leibler—is just gotten by averaging over the components of the informational content vector. It is a kind of summary obtained from informational content.

If we average again we get the average quantity of information in the signals. This quantity is called *mutual information*. If we take the maximum of this over signaling system equilibria, we get a measure of the information transfer capacity in the signaling game. There is a seamless integration of this conception of content with classical information theory.

Intentionality and teleosemantics

Some philosophers take the view that real information presupposes *intentionality* and that consequently the m:mathematical theory of information is irrelevant to informational content. The semantic notion of information is conflated with the question of intentionality. What is intentionality? It is a said to be a kind of directedness towards an object. That doesn't tell us much, and doesn't explain why anyone should think it was not part of m:mathematical ↳ information theory. Signals, after all, do carry information directed toward the states and information directed toward the acts.

p. 43

The philosophical history of the concept of intentionality tells us more. It starts with Franz Brentano,[10] who held that intentionality was what distinguished the mental from the physical. If what is being left out is a model of the mental life of the agents, then I would say that it should be left out when the agents lack a mental life and put in when they do. I would not speculate on the mental life of bees; to talk of the mental life of bacteria seems absurd; and yet signaling plays a vital biological role in both cases. Some may want to define signals so that these are not "real" signals, but I fail to see the point of such maneuvers. Rather, I would treat the case where agents have a mental life as a special case. If we have a reasonable model of the relevant aspects of mental life, we can put them in the model. We move some way in this direction in the next section, where we consider subjective information.

Some have swallowed the requirement of intentionality or something quite like it, but have tried to let Mother Nature (in the form of evolution) supply the intentionality. As John Maynard Smith puts it: "In biology, the use of informational terms implies intentionality, in that both the form of the signal, and the response to it, have evolved by selection. Where an engineer sees design, a biologist sees natural selection."[11] This is roughly the idea behind Ruth Millikan's *teleosemantics*. An evolved signal has a directedness, or intentionality, in virtue of the Darwinian fitness accrued by its use.[12]

I say about teleosemantic intentionality the same thing I said about mentalistic intentionality. If we have a good model where it applies, it can be added to the theory. But neither intentionality nor ↳ teleosemantics is required to give an adequate account of the informational content of signals. Here I stand with Dretske. The information is just *there*. At this point some philosophers will say "You might as well say that Smoke carries information about fire." Well, doesn't it? Don't fossils carry information about past life forms? Doesn't the cosmic background radiation carry information about the early stages of the universe? *The world is full of information*. It is not the sole province of biological systems. What is special about biology is that the form of information transfer is driven by adaptive dynamics.

p. 44

Objective and subjective information

None of the probabilities used so far are degrees of belief of sender and receiver. They are objective probabilities, determined by nature and the evolutionary or learning process. Organisms (or organs) playing the role of sender and receiver need have no cognitive capacities.

But suppose that they do. Suppose that a sender and receiver are human and that they try to think rationally about the signaling game. Suppose that the sender has subjective probabilities over the receiver's strategies and the receiver has subjective probabilities over the sender's strategies, and that both have subjective probabilities over the states. These subjective probabilities are just degrees of belief; they may not be in line with the objective probabilities at all. Then each signal carries *two additional kinds of subjective information* . There is *subjective information about how the receiver will react*, which lives in the sender's degrees of belief. This is of interest to a sender who wants to get a receiver to do something. There is *subjective information about what state the sender observed*, which lives in the receiver's degrees of belief. This is of interest to a receiver who

wants to use the sender as a source of information about the states. Both sender and receiver use these kinds of
p. 45 information in decision making. Both sender and receiver strive (1) to act optimally given ↳ their subjective probabilities, and (2) to learn to bring subjective probabilities in concordance with the objective probabilities in the world. They may or may not succeed. When we are applying the account to beings that can reasonably be thought to have subjective probabilities, such as perhaps ourselves,[13] we now have at least four types of informational content—two objective and two subjective. If the signaling game is more complex, for instance if there is an eavesdropper, the informational structure becomes richer.

The flow of information

In the signaling equilibrium of a Lewis sender-receiver game, information is transmitted from sender to receiver, but it is only in the most trivial sense that we can be said to have a *flow* of information. As a preview of coming attractions (Chapters 11, 13, 14) and as an example of flow, let us consider a little signaling *chain*.

> •→•→•

There are a sender, an intermediary, and a receiver. Nature chooses one of two states with equal probability. The sender observes the state, chooses one of two signals, and sends it to the intermediary; the intermediary observes the sender's signal, chooses one of her own two signals, and sends it to the receiver. (The intermediary's set of signals may or may not match that of the sender.) The receiver observes the intermediary's signal and chooses one of two acts. If the act matches the state, sender, intermediary and receiver all get a payoff of one, otherwise a payoff of zero.

It is tempting to assume that these agents already have signaling for simpler sender-receiver interactions to build upon. But even if they do not, adaptive dynamics can carry them to a signaling system, as shown in figure 3.4:

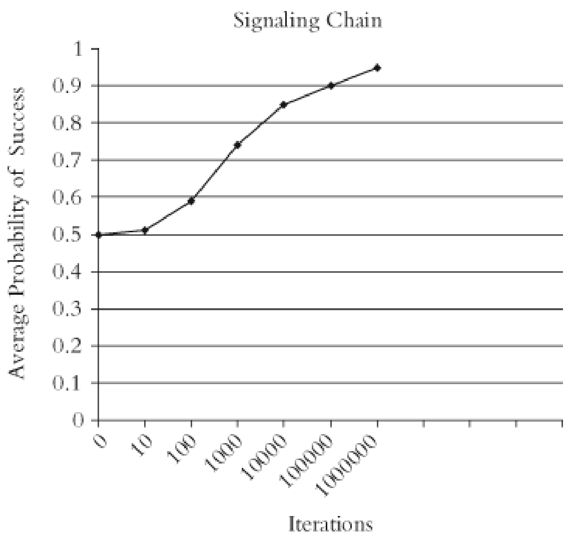

Figure 3.4: Emergence of a signaling chain *ex nihilo* by reinforcement learning.

p. 46 Although reinforcement learning succeeds in creating a signaling chain without a pre-existing signaling background, notice that it takes a much longer time than in the simpler two-agent model.

The speed with which the chain signaling system can be learned is much improved if the sender and receiver have pre-existing signaling systems. They need not even be the same signaling system. Sender and receiver can have different “languages” so that the intermediary has to act as a “translator”, or signal transducer. One could even consider an extreme case in which the sender and receiver used the same tokens as signals but with opposite meanings. "For example, sender's and receiver's strategies are:

SENDER	RECEIVER
State 1 ⇒ red	red ⇒ Act 2
State 2 ⇒ blue	blue ⇒ Act 1

A successful translator must learn to receive one signal and send another, so that the chain leads to a successful outcome.

SENDER	TRANSLATOR	RECEIVER
State 1 ⇒ red	see red ⇒ send blue	blue ⇒ Act 1
State 2 ⇒ blue	see blue ⇒ send red	red ⇒ Act 2

The translator's learning problem is really quite simple, and she can learn to complete the chain very quickly."

In this signaling chain equilibrium, the sender's signal to the translator contains one bit of information about the state and the translator's signal to the receiver contains one bit of information about the state. And on any play, the translator's signal to the receiver has the same *informational content* as the sender's signal to her. Information *flows* from sender through translator to receiver. The receiver then acts *just as she would have if she had observed the state directly*.

That is, of course, the ideal case. Some information can get lost along the way because of noise or error.[14]

Using our notion of the content of a signal, there is no difficulty in allowing for gradual degradation of content. Information can flow through longer signaling chains and through more complex signaling networks. Some informational content may get lost. This may even be beneficial if extraneous information needs to be filtered out. We will see how information from different sources may be integrated in ways that include logical inference and computation of truth values as special cases. Signaling networks of different kinds are the locus of information transmission and processing at all levels of biological and social organization. The study of information processing in signaling networks is a new direction for naturalistic epistemology.

Notes

1 I must admit to having done some of this, before I knew better.

2 As always, there is the question of whether the framework is being correctly applied to model the situation of interest. We assume here that it is.

3 The probabilities never really hit zero or one, although they may converge towards them. So conditional probabilities are well defined. We don't have to worry about dividing by zero. If it appears in an example that we are dividing by zero, throw in a little epsilon.

4 By Bayes' theorem, the same quantity can be expressed as:

pr(signal given state)/pr(signal).

5 Although not technically a metric because it is not symmetric.

6 Kullback and Leibler 1951, and Kullback 1959.

7 Lindley 1956.

8 See I. J. Good's preface to *Good Thinking* 1983.

9 This is information content within a given signaling game. It is implicit that this vector applies to the states or acts of *this game*. For a different game, the content vector shows how the signal moves probabilities of different states, or different acts. Content depends on the context of the signaling interaction. It is a modeling decision as to which game is best used to analyze a real situation.

10 Brentano 1874.

11 Maynard Smith 2000.

12 See Millikan 1984. For other teleosemantic theories that do not share Millikan's basic commitment to a picture theory of meaning see Papineau 1984, 1987.

13 Modern psychology details systematic departures from this idealized picture.

14 Here I part company with Dretske 1981: 57–8.