Signals that make a Difference

Brett Calcott, Paul Griffiths, Arnaud Pocheville

Abstract

Recent work by Brian Skyrms offers a very general way to think about how information flows and evolves in biological networks — from the way monkeys in a troop communicate, to the way cells in a body coordinate their actions. A central feature of his account is a way to formally measure the quantity of information contained in the signals in these networks. In this paper, we argue there is a tension between how Skyrms talks of signalling networks and his formal measure of information. Although Skyrms refers to both how information *flows* through networks and that signals *carry* information, we show that his formal measure only captures the latter. We then suggest that to capture the notion of *flow* in signalling networks, we need to treat them as causal networks. This provides the formal tools to define a measure that does capture flow, and we do so by drawing on recent work defining *causal specificity*. Finally, we suggest that this new measure is crucial if we wish to explain *how* evolution creates information. For signals to play a role in explaining their own origins and stability, they can't just carry information about acts: they must be difference-makers for acts.

1   Signalling, Evolution, and Information

During the American Revolution, Paul Revere, a silversmith, and Robert Newton, the sexton of Boston's North Church, devised a simple communication system to alert the countryside to the approach of the British army. The sexton would watch the British from his church and place one lantern in the steeple if they approached by land and two lanterns if they approached by sea. Revere would watch for the signal from the opposite shore, and ride to warn the countryside appropriately. Revere and the sextons' use of lanterns was famously captured in Longfellow's poem with the phrase: 'one if by land, two if by sea'.

This warning system possesses all the elements of a simple signalling game as envisioned by David Lewis ([1969]). We have a *sender* (the sexton), and a *receiver* (Revere). The sender has access to some state of the world (what the British are doing), and the receiver can perform an act in the world (warn the countryside). Both sender and receiver have a common interest: that the countryside learns which way the British are coming. Together, they devise a set of signals and coordinate their behaviour to consistently interpret the signals.
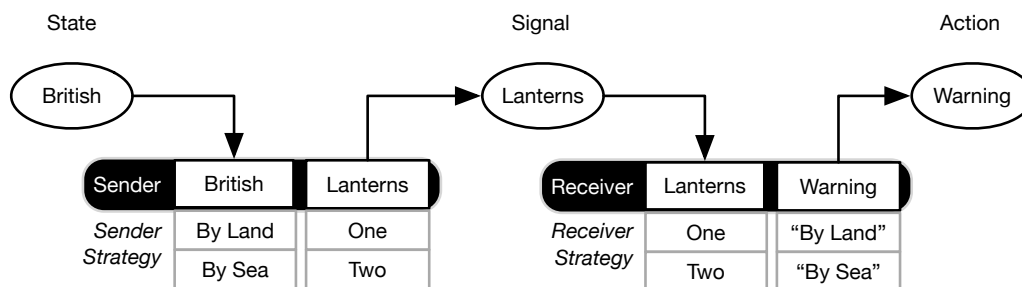

Figure 1: A simple signalling game, warning the countryside of the arrival of the British.

For Lewis, these coordination games showed how arbitrary objects (in this case, the lanterns) could acquire conventional meaning. Revere and the sexton needed to assign meaning to some signals in order to achieve their goal, but the warning system would have worked equally well if Revere and the sexton had decided to employ the opposite meanings: 'one if by sea, two if by land'.

Lewis treated the players in these games as rational agents choosing amongst different strate-

gies. But Skyrms, in his 1996 book, *The evolution of social contract*, extended Lewis's framework, showing that these conventions could arise in much simpler organisms, with no presumption of rational agency (Skyrms [1996]). In a population of agents with varying strategies, where the agents' fitnesses depend on coordinating their behaviour using signals, repeated bouts of selection can drive the population to an equilibria where one particular signalling convention is adopted.

With the requirement for rationality gone, the signalling framework can be applied to a broad range of natural cases — from the calls monkeys make to the chemicals exuded by bacteria. This generalisation also permits signalling to be applied not only where signalling occurs *between* individuals, but also when signalling occurs between subsystems *within* a single individual (Skyrms [2010b], pp. 2–3). This shift of perspective to internal signalling maintains the same formal structure, but shifts the focus to such things as networks of molecular signals, gene regulation, or neural signalling (Calcott [2014]; Godfrey-Smith [2014]; Planer [2013] Cao [2014]). We mention these cases because, although we intend our arguments here to apply generally to all cases of signalling, we think the most compelling examples of the complex networks we use to drive our arguments can be found *inside* organisms.

In his 2010 book on signalling, Skyrms connected these ideas about signalling to information theory, outlining a way to measure the information in a signal in any well defined model of a signalling network. By providing the formal tools to measure information *at a time* within a signalling network, and linking it to the previous work about how signalling on these networks may evolve *over time*, Skyrms delivers a framework in which he can clearly and justifiably claim that 'Evolution can create information' (Skyrms [2010b], p. 39).

Two key ideas recur throughout Skyrms's discussion of signalling networks: information *flows*, or is *transmitted*, through these networks, and signals *carry* information. In this paper, we argue that these two ideas are distinct, and that Skyrms's approach to measuring information only captures the latter. In simple networks, these two ideas may appear equivalent, so we provide some example networks where these two notions come apart. We then suggest that to capture the notion of *flow* in signalling networks, we should treat them as causal networks. This

provides the formal tools to define a measure that does capture the flow of information, and we connect this approach to recent work on defining *causal specificity*. Finally, with both measures in place, we suggest that this new measure is crucial if we wish to explain *how* evolution creates information.

2   Skyrms's Measure of Information

We begin with a brief overview of Skyrms's approach to measuring information in signals. The quantity of information in a signal, according to Skyrms, is related to how signals change probabilities (Skyrms [2010b], p. 35).[1] For example, if the probability of the British coming by land, w_1, was initially 0.5 and the probability conditional on seeing one lantern in the steeple, s_1, was 1, Then the signal (seeing one lantern) *changes* the probability from 0.5 to 1.[2] Skyrms proposes we look at the ratio of these probabilities (he dubs this ratio a *key quantity*):

$$\frac{p(w_1|s_1)}{p(w_1)} = \frac{1}{0.5} = 2.0 \tag{2.1}$$

If we take the logarithm (base 2) of this ratio, we get a quantity measured in *bits*. In this case, the amount of information is $\log_2(2.0) = 1$ bit. If the signal failed to change our probabilities, then the ratio would equal 1, and the logarithm would instead give us 0 bits.

This quantity (1 bit) tells us how much information a particular signal (one lantern, s_1) has about one state (the British coming by Land, w_1). It is sometimes known as the *point-wise mutual information* between single events. If we want to know how much information this particular signal has about *all* world states, then we take the weighted average over those states:

---

[1]In this paper we focus on Skyrms's definition of the *quantity of information* in a signal. Skyrms also defines a related semantic notion—the *informational content* of a signal. We avoid discussing the more controversial semantic issues in this paper.

[2]Here, and throughout the paper we mean *objective* probabilities. Recall that we're dealing with models here, and we can stipulate what all the probabilities are. Whether the model is a good one or not is another question.

$$\sum_w p(w|s) \log_2 \frac{p(w|s)}{p(w)} \qquad (2.2)$$

Skyrms identifies this quantity as a Kullback-Leibler distance.[3] The Kullback-Leibler distance measures the difference between two probability distributions, in this case the probability of the two alternative attacks before and after the signal. It is also known as the information gained, or the relative entropy.

What if we are interested in how much information, on average, we expect the lanterns to provide? To calculate this, we need to look at how much information each signal (one lantern or two lanterns) provides, and weight the probability that each will occur:

$$\sum_s p(s) \sum_w p(w|s) \log_2 \frac{p(w|s)}{p(w)} \qquad (2.3)$$

We shall refer to this as the *information in a signalling channel* to distinguish it from the *information in a signal*. The information in a signalling channel is the *mutual information*, $I(S;W)$, between the signalling channel and the world states, and it will be the focus of our inquiry for the remainder of this paper.

We focus on the information in a signalling channel (rather than a single signal) as it allows us to easily relate these ideas to the work on causal graphs we introduce in the following sections. This should be no cause for alarm, for mutual information is straightforwardly related to the Kullback-Leibler distance, and forms part of the 'seamless integration' of signalling theory with classical information theory that Skyrms emphasises (p. 42). The issues we identify with mutual information also translate seamlessly to Skyrms's claims about the Kullback-Leibler distance and his 'key quantity', the ratio mentioned above.

We just saw how Skyrms measures the information that a signal (and thus, a signalling channel) carries about the states of the world. But signals carry information about the *acts* being chosen

---

[3]We are following Skyrms's terminology here by using 'distance' rather than 'divergence', though it is not a true distance (as Skyrms himself notes in [2010b], p. 36).

too. Skyrms treats the information a signal carries about acts and cues as 'entirely analogous' (Skyrms [2010b], p. 38, [2010a]). If the probability that Revere would warn the countryside 'By Land', a_1, was originally 0.5, and the probability conditional on seeing one lantern in the steeple s_1 was 1, then the signal changes our probability from 0.5 to 1. Skyrms applies the same formalism as above, and thus the information in a signalling channel about acts can be measured using mutual information in exactly the same fashion that it was used to measure information about states:

$$I(S;A) = \sum_s p(s) \sum_a p(a|s) \log_2 \frac{p(a|s)}{p(a)} \qquad (2.4)$$

For reasons that shall become plain later in the paper, our examples will focus on information about *acts*, rather than information about world states, so it is this last equation that we use as a contrast in the following examples.

3   Carrying Information vs Information Flow

In this section we present three examples that reveal a tension between how Skyrms talks about signalling networks and how he measures information in these networks. We argue that although Skyrms's use of information theory can capture how much information a signal *carries* about an action, this measure alone misses something important, as it fails to capture the idea of *information flow* in a network. This becomes apparent when we construct signalling networks where signalling pathways can branch and merge.

Our examples build on the basic structure of the signalling game used to represent the warning system of Revere and the sexton. To aid in the exposition, however, we will make a number of modifications. First, we recast this model as an internal signalling system. To do this, we simply sketch a boundary around the two-player signalling game described. The result is a model of a plastic organism that encounters two different environments, and responds to each environment with a different behaviour. To further simplify, all signals, acts, and states will take on binary

values—so they're either ON or OFF. The world state consists of some environmental cue that is ON or OFF, the signal sent is either ON or OFF, and the act is likewise a behaviour that is either ON or OFF (see Figure 2). Our examples build on this signalling network, gradually increasing their complexity.
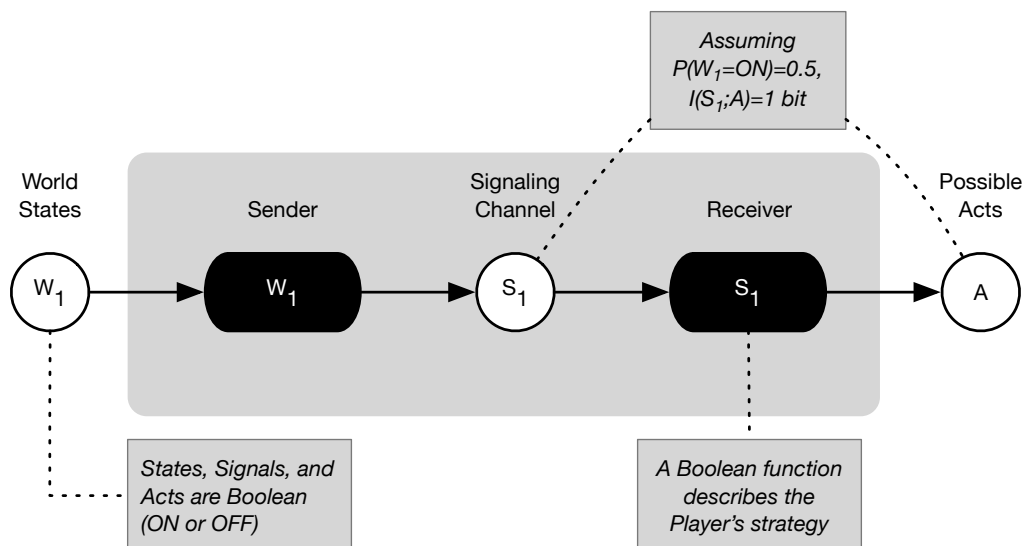

Figure 2: The behavioural plasticity of a simple organism, modelled as an internal signalling system.

3.1   Example 1.

The organism described consists of a single signalling channel S_1. Now we assume that, as a by-product of producing a signal in channel S_1, the sender simultaneously transmits another signal along a second signalling channel S_2. This signal can also be either ON or OFF, and our sender is wired so that when S_1 is ON, S_2 is also turned ON, and when S_1 is OFF, S_2 is also turned OFF. Pathway S_2, unlike channel S_1, doesn't go anywhere. It's not that the receiver ignores the signal from channel S_2, the signal simply doesn't reach it (see Figure 3).

What can we say about the information in signalling channel S_2, using the measure Skyrms provides? Because we stipulated that the signalling channel S_2 was perfectly correlated with that on S_1, the new signalling channel *carries* precisely the same amount of information as the original channel S_1, both about the state of the world, and about the act being performed.
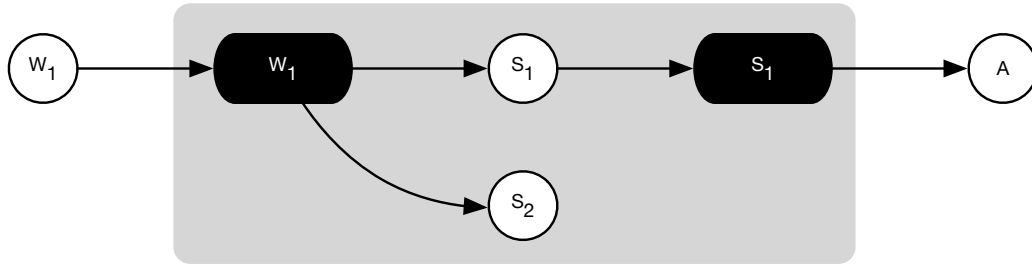
8

Figure 3: Adding a second signalling pathway, that is a by-product of the first, and perfectly correlated with it.

You would be right to think that the information in channel S_2 is redundant: once we know the information carried by channel S_1, the information in channel S_2 tells us nothing new. Formally, we can capture this using *conditional mutual information*. The mutual information, $I(S_2; A)$, is 1 bit, but the conditional mutual information, $I(S_2; A|S_1)$, is 0 bits. But notice that the reverse is also true: If we already know about S_2, then S_1 tells us nothing new—$I(S_1; A|S_2)$ is also equal to 0 bits. There is redundant information in the two channels, but if we look solely at the information measures, we're not in a position to pick out either channel as *the* redundant one. Skyrms's information measure cannot distinguish between these two signals.

Should Skyrms's measure distinguish between these two signals? That depends on what this information measure is meant to capture. Let us first state what Skyrms's measure *does not* capture.

One stated aim of Skyrms is to study the *flow of information* in signalling networks (Skyrms [2010b], pp. 32–3). What does he mean by *flow*? A flow implies *direction*, and indeed Skyrms talks of information being transmitted 'from a sender to a receiver' (p. 45), and of information flowing in one direction (p. 164) and sometimes in both directions (p. 163). Information also flows *through* networks by passing through one node and to the next. For example, it can flow along a signalling chain (p. 45), moving from from sender to receiver via an intermediary, who both sends and receives signals. Cutting a node out of the network can also disrupt this directed flow (p. 163). Thus, the flow of information in a network is dependent on the directed structure of the network, and this directed structure is an essential part of the networks depicted in the diagrams used throughout the book. This structure is not all there is to information flow: for

example, an intermediary player in a signalling chain that always does the same thing will not transmit any information, or information might *decay* as it passes through the nodes (p. 171). But the directed structure does place a restriction on how information flows: if we cannot trace a path between two nodes by following a series of arrows, then there cannot be any information flow between them.

In the network we have outlined above, there is clearly *no flow* from S_2 to A, as there is no arrow connecting the two nodes, nor is there any path, or combination of arrows, that travels from S_2 to A. Yet, according to Skyrms's measure, S_2 does carry information about A. So we conclude that Skyrms's measure of information does not capture the flow of information. As further evidence of this, we note that *mutual information*—which captures the amount of information in a signalling channel about the acts—is a symmetric measure and thus is insensitive to the direction of flow.

Although Skyrms's measure does not capture the flow of information in the network, it clearly captures something important. An observer, seeing the signals in channel S_2, gains information about the action, A, the organism will perform. Perhaps the observer is a parasite or predator, and can use this information to exploit the organism. Notice than an observer could equally exploit the organism if it observed channel S_1, so the fact Skyrms's measure does not distinguish between these two channels is a virtue if our goal was to explain how the organism was exploited.

A signalling channel like S_2 may also play a role in the organism itself. For example, in many organisms, a copy of the neural signals for movement are routed to the sensory structures, a phenomenon known as corollary discharge (Crapse and Sommer [2008]). This copy of the signal can enable an organism to distinguish whether it bumped into something, or whether something bumped into it. So, even if information does not flow from a signal to an action, the fact that a signal *carries* information about an action can play a role in explaining something about the organism.[4]

---

[4]We thank an anonymous reviewer for clarifying the role both measures play, and for supplying this intriguing example.

What about information *flow* then? Although Skyrms's measure cannot distinguish between channels S_1 and S_2, there are certainly reasons we want to keep them separate. For example, if we want to explain *why* the organism responds differently to the two environments, we will appeal to signalling channel S_1, for information flows from the world state to the action through this channel S_1, and not through S_2.

So there are two distinct things we might want to capture about signals and acts in a signalling network:

1. The information *flow* from a signal to an act.

2. The information a signal *carries* about an act.

A number of objections might be made at this point. You might think we've simply misunderstood the modeling exercise, as we've added on a channel that serves no purpose. You might even complain that channel S_2 is not a signalling channel at all, for if no one is listening, then whatever is being sent doesn't count as a signal. We think these objections are not good ones, and that there are valid reasons to model channels like this. For instance, once we turn to signalling networks where part of what evolves may be the topology of the signalling network itself (Skyrms [2010b], p. 3), then there are good reasons to model and measure information in channels that are as yet unconnected, for future evolutionary changes may connect them (Calcott [2014]). Rather than pursuing this line of support, however, we shall strengthen our case by showing that the distinction between carrying information and the flow of information does not require unconnected channels. To do this we need to introduce some more complex examples.

3.2 Example 2.

In our second example, the signalling channel S_2 flows to the receiver, but indirectly, via a third signalling channel S_3. As we mentioned above, Skyrms refers to this as a signalling chain. We shall add a twist to this, however. Our intermediary also receives a second cue, W_2, from the

world. So our world state now consists of two cues, W_1 and W_2. They are both binary, so the complete world state now consists of four possibilities.

Our intermediary will simply copy the signal it gets from the original sender, W_1, *but only when this second cue is present (when W_2 is ON)*. Our intermediary effectively acts like an AND gate[5], sending the ON signal only when both W_2 and A are ON.

Our receiver also now gets two signals, one from the original channel S_1, and another from channel S_3 (the end of the signalling chain). Our receiver behaves like an OR gate, acting when either of the signals it receives is ON. Figure 4 shows a diagram of the entire signalling network.
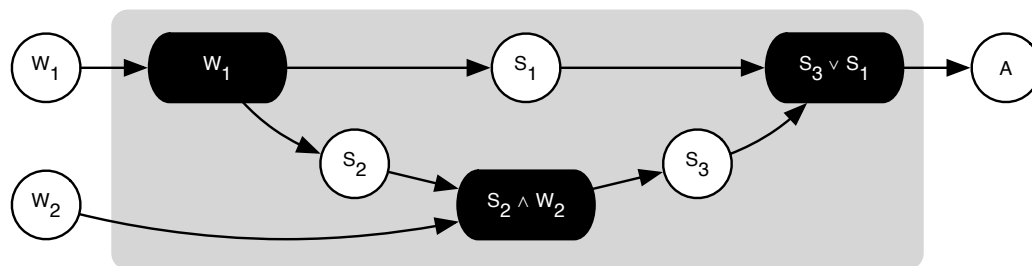

Figure 4: Adding a second signalling pathway that includes a signalling chain, mediated by another world state. Note that $I(S_1; A)$ and $I(S_2; A)$ are always equivalent, regardless of the probability distribution of W_2

How does this network behave? When the environmental cue W_2 is absent ($W_2 = OFF$), the signalling chain always transmits OFF to the receiver. When W_2 is present, however, the signalling chain delivers the same message as the direct path, via S_1. If the cue W_2 is *never* present, then the signalling chain never transmits the value from W_1. In contrast, however, if W_2 is *always* present, then channel S_3 will always take on the same value as channel S_2 (which, by stipulation always takes on the same value as S_1).

Clearly, W_2 controls how likely it is that information flows along the signalling chain consisting of S_2 and S_3. If W_2 is absent, then the network is effectively the same as the previous example, for there is never any flow of information from S_2 to the act, and thus behaves as though this particular signalling channel does not exist.

---

[5]A digital logic gate that implements the AND function.

But notice that although the probability of W_2 controls how much information flows from signalling channel S_2 to the act, A, it does not affect the information that S_2 *carries* about the act, A. For example, if we assume that the probability of W_1 being ON is 0.5, then no matter how we vary the probability of W_2, channel S_2 always carries 1 bit of information about the acts.

3.3   Example 3.

Our second example provided two pathways for the receiver to get information from the environment. The first pathway was direct, via channel S_1. The second pathway was via a signalling chain that was mediated by another cue from the environment. But this signalling chain added nothing new to the information gained by the receiver; removing the signalling chain would have no effect on the fitness of the organism.[6] Perhaps that explains (or justifies) why manipulating the way information flows down this chain had no effect on the information measure.

To see why this is not the case, we can extend this model again, to ensure that both channels have an effect on fitness. Now we're going to break our original pathway through signalling channel S_1, and make it a signalling chain too. Like the other signalling chain, it will be mediated by this second cue from the environment, and hence send a fourth signal, S_4. We shall make this new signal (attached to the end of the original pathway) only turn ON when the pathway A is ON and the second cue from the world is OFF (see Figure 5).

We can describe our new organism in the following way. When W_2 is present, the signalling chain that goes via channel S_2 is active, and it transmits the value of W_1 to the act, A. When W_2 is not present, the signalling chain that goes via channel S_1 is active instead. So our organism succeeds in reacting to W_1, but it does so by making use of two distinct signalling chains, and the particular chain that is active depends on the cue W_2. Assuming W_2 is sometimes present and sometimes absent, then removing *either signalling chain* will now affect the fitness of the

---

[6]Assuming that we disregard the idea that multiple channels might provide a more robust signalling mechanism—this idea is important, but beyond the scope of our current modeling endeavour.
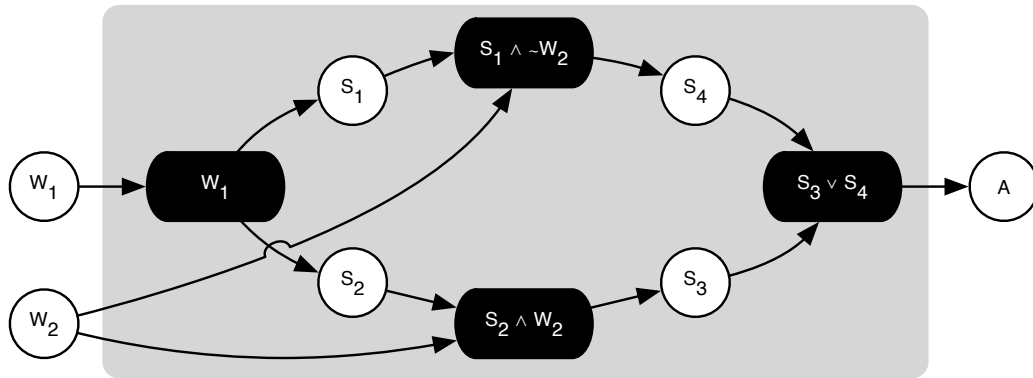
Figure 5: The acts are now driven by two different signalling chains. Each transmits the value of W_1, but which one successfully does so is dependent on W_2. The information in the two channels, S_1 and S_2, remains the same, regardless of the probabilities of W_2.

organism.

Yet again, however, varying the probability of W_2 *has no effect* on the information that the channels S_1 or S_2 carry about the act. For example, if W_2 is present 99% of the time, then signalling channel S_1 will only be active a trivial 1% of the time. In a situation like this, it seems intuitive to say that more information is *flowing* from S_2 to A than is *flowing* from S_1 to A, while the information *carried by* both S_1 and S_2 is equal. So even in networks where all signals lead somewhere and impact fitness, carrying information and the flow of information remain distinct.

These examples are manufactured, of course. But the idea that there may be multiple signalling channels that may be active under different conditions seems like a very generic and useful capacity. For example, the chemotactic abilities of cellular slime mould cells (*Dictyostelium discoideum*) that guide them to aggregate in times of stress appears to depend on multiple internal signalling pathways. Each of these internal pathways is active in different conditions, one in shallow chemical gradients, one in steep gradients, and another that acts in later stages of aggregation (Van Haastert and Veltman [2007]).

4 Signalling Networks are Causal Networks

In the last section we argued that the *information flow* from a signal to an act and the information *carried by* a signal about an act are distinct, and that Skyrms's measure only captures the latter of these ideas. Our aim now is to provide a formal measure of information flow.

First, we argue that signalling networks should be treated as causal graphs. This makes explicit the directionality of signalling flows in these networks, and identifies signals as points of intervention, whose manipulation has the power to change acts. Our strategy will be to suggest that the flow of information from a signal to an act should be understood as a causal notion, equivalent to the causal influence that the signal has over the act.

Our approach to formalising this measure will be to connect these ideas to recent work on formalising *causal specificity*, which uses information theory to precisely measure how much influence a cause has over an effect and, importantly, provides a means to distinguish the differential contribution of multiple causes of a single effect. We then extend and adapt this work to analyse the signalling framework, and outline a way of measuring information about acts that agrees with Skyrms's measure in simple cases, but adequately deals with the problem cases we've outlined in this section.

Perhaps the notion that information flow is causal strikes some readers as odd. We think there are many reasons for interpreting signalling networks as causal graphs. Signals, like causation, are *directed*—information flows in a particular direction. The notion of an intervention and the ability to evaluate counterfactuals is also implicit in signalling networks. The British actually came by sea, but the setup of the signalling system devised by Revere and the sexton tells us what *would have* happened if they had come by land. Importantly for our discussion below, *signals are points of intervention*. A mischievous choir-boy could have derailed Paul Revere's historic ride by removing one of the lanterns from the belfry of the north church. Furthermore, if we look at biological examples of signalling, a causal interpretation seems entirely natural: the bark of one vervet *caused* another to run up a tree; one neuron firing *caused* another to fire. Lastly, interventions are the key method by which biologists discover and document actual

signalling channels in molecular biology and elsewhere.

The translation between a signalling network and a causal graph is also straightforward. The world states, signalling channels, and sets of actions make up the variables in the causal graph. These variables take on different values corresponding to particular states, signals, and acts that are occurring. The world states in a signalling network lie upstream of both signals and acts, so these will be the *root* variables in the causal graph. The strategies of the players in the signalling network generate the conditional probabilities that relate one or more parent variables to one or more child variables. Given the probabilities of the world states (our root variables), the strategies of the players (which generate the conditional probabilities of all non-root variables), and the structure of the signalling network (the graph), we can calculate the probabilities of all other variables in the graph.

Transforming the signalling network into a causal graph also allows us to connect the structure of a signalling network to existing work on *causal explanation*. We have in mind the influential work by Woodward ([2003]), and in particular the insight that 'causal relationships are relationships that are potentially exploitable for purposes of manipulation and control' (Woodward [2010], p. 314). For example, treating the signalling network in our first example as a causal graph gives us the means to clearly state *why* the signalling channel S_2 is not explanatory: manipulating the variable S_2 will have no effect on the act variable A.

Once we transform the signalling network into a causal graph, we see that the cues, signals, and actions in a signalling game are just a special case of the more general notion of a set of variables in a causal system.[7] Furthermore, the distinction between *information flow* and *carrying information* is transformed into something familiar: the distinction between causation and correlation[8]. Two variables (a signal and an act) may be correlated not because one causes another, but because they are affected by a common cause.

---

[7]Not *all* of Skyrms's signalling networks can be easily treated as causal graphs, because they are not all Directed Acyclic Graphs (See the networks in chapter 14 in Skyrms [2010b]). But the ones where information flows from world state to act are. These are the ones that concern us here.

[8]The distinction we are interested here is often phrased as causation versus correlation, but it is more accurate to describe it as causation versus *association*, as correlation is often reserved for linear relationships between two variables, rather than the use of mutual information as is deployed here.

We can now see why we have focused on information about acts, rather than information about world states. As we mentioned in the introduction, Skyrms treats information about acts and states as 'analogous'. If the goal is to capture the information *carried* by signals, then this correlative measure, using mutual information, will do just fine. But if our goal is to explain how the signalling network makes the organism respond as it does, then there is clearly an asymmetry. A signalling channel need only be correlated with the world states to represent them, but for a signalling network to make the organism responsive, the signals in a channel must be the *causes* of acts.

Treating signalling networks as causal graphs also allows us to make use of a set of formal tools for distinguishing between merely observing the statistical relationship between two variables and measuring the causal effect of one variable on another (Pearl [2000]). The causal effect of setting a variable X to some particular value x amounts to something intuitive. We intervene on the graph, ignoring all incoming edges to a variable, and hold its value fixed at x. The resulting model, when solved for the distribution of another variable Y, 'yields the causal effect of X_i on X_j, which is denoted $P(x_j|do(x_i))$.' (Pearl [2000], p. 70). We use a more concise symbolism where $do(x_i)$ is replaced by $\widehat{x_i}$. The causal effect $P(x_j|\widehat{x_i})$ is to be contrasted with the observational conditional probability $P(x_j|x_i)$.

Using the *do* operator with the information-theoretic measures, we will be able to take the causal structure of the networks into account. In the next section, we outline how we can do that by connecting these ideas to recent work formalising *causal specificity*.

4.1   Causal Specificity

In complex systems, and especially in biology, an effect may have many upstream causes, and there is often heated debate about which causes are most important (for example, the nature–nurture controversy can be partly seen as one long, extended fight about this). In these cases, the problem is not what counts as a cause, but rather why some causes are more significant than others. We might put it this way: identifying causes tells us which variables are *explana-*

tory, whereas distinguishing amongst causes tells us *how explanatory* those different variables are.[9]

The difference between these two tasks is reflected in our examples. In the first example, channel S_2 is correlated with the action, but does not cause it. This is because manipulating channel S_2 makes no difference to the action, A. Hence we can say that channel S_2 plays *no* role in explaining why the action takes a particular value. When we turn to examples 2 and 3, however, channel S_2 is no longer merely correlated with the action. There are conditions under which manipulating this S_2 *would* change the action. But we still need to distinguish between S_1 and S_2 to address the different contributions they make to determining the action.

One prominent proposal to distinguish amongst causes concerns the degree to which they are *specific* to an effect. Interventions on a highly specific causal variable produce a large number of different values of an effect variable, providing what Woodward terms 'fine-grained influence' over the effect variable (Woodward [2010], p. 302). The intuitive idea behind causal specificity can be illustrated by contrasting the tuning dial and the on/off switch of a radio. Both the tuning dial and on/off switch are causes (in the interventionist sense) of what we are currently listening to. But the tuning dial is a more specific cause, as it allows a range of different music, news, and sports channels to be accessed, whilst flipping the on/off switch simply controls whether we hear something or nothing.

4.2 Formalising Causal Specificity

Philosophical analyses of causal specificity have been mainly qualitative, but Woodward has suggested that the upper limit of fine-grained influence is a one-to-one (bijective) mapping between the values of the cause and effect variables: every value of an effect variable is produced by one and only one value of a cause variable and vice versa. Griffiths *et al.* ([2015]) showed that this idea can be generalised to the whole range of more or less specific relationships using an information-theoretic framework. They suggest that causal specificity can be measured by

---

[9]Assuming an interventionist account of explanation.

the mutual information between the cause variable and the effect variable.[10] This formalises the idea that, other things being equal, the more a cause specifies a given effect, the more knowing how we have intervened on the cause variable will inform us about the value of the effect variable.

At first glance, this suggestion looks problematic, for the mutual information between two variables is symmetric, and thus typically only employed as a measure of correlation. Indeed, this is the very problem that we've encountered in the signalling examples, a straightforward measure of mutual information between two variables in a causal graph (or signalling network) takes no account of the structure of the graph. The required asymmetry of causation can be regained, however, by measuring the mutual information between cause and effect when we *intervene* on the cause variable, rather than simply observing it.

Measuring mutual information under interventions changes the core calculation in the mutual information equation from an observational conditional probability to a conditional probability that captures the *causal effect* of one variable on another:

$$I(\widehat{S};A) = \sum_s p(\widehat{s}) \sum_a p(a|\widehat{s}) \log_2 \frac{p(a|\widehat{s})}{p(a)} \tag{4.1}$$

Recall that a hat on a variable indicates that it's values are determined by intervention rather than observation. Adding a hat to a variable in an equation to turn mere correlation into causation may seem like magic, but it amounts to something intuitive: performing an experiment on a causal graph. We manipulate the cause variable, setting it to different values, and then record the ensuing probabilities of the different values of the effect variable. Recording these values generates a joint probability distribution *under intervention*. We can then measure the mutual information in this modified probability distribution, and it will reflect how much information our interventions give us about their effects.

---

[10]The approach developed in Griffiths *et al.* ([2015]) was anticipated by Korb *et al.* ([2009]). Pocheville *et al.* ([In Press]) extend this approach to measure the proportionality and stability of causal relationships in addition to their specificity.

This causal information theoretic approach does more than capture the notion of specificity, however, for the measure is zero in cases where the interventionist framework tells us that a variable is *not a cause*. Thus, the use of this information measure can capture a range of relationships between two variables, from no causal control at all, to fine-grained, highly specific causal control. This makes it an appropriate measure for contrasting the causal contribution that many upstream putative causes might have over an effect.

Intervening, rather than simply observing, does introduce an extra burden, however. Because we can no longer simply observe the probability distribution over the cause variable as it naturally occurs, we need to stipulate a probability distribution over the values of the cause variable. How do we decide what probabilities these interventions take? There are a number of valid approaches, depending on our aims.

One option is to assume all values of the cause variable are equiprobable (a maximum entropy distribution). This approach tells us something about the *potential control* of one variable over another.

Another option is to use the *natural distribution* of the cause variable. The natural distribution is the probability distribution that the cause variable takes when no interventions are made. This can be attained by observing the system without intervening, and recording the probability of each occurrence of the value of the cause variable. We then intervene on the system to mimic this distribution over the cause variable. This approach measures the *actual control* of the cause variable (see Griffiths *et al.* [2015] for discussion).

5   Information Flow as Causal Control

Our suggestion is to treat a signalling network as a causal graph, and to measure how causally specific a signal is for an act. We use the natural distribution of the signalling variable as this will tell us how much actual control the variable has given its normal range of variation. We'll also need to measure specificity in each world state, for the specificity of the signal may differ

across the different world states. We can combine these specificity measures using a weighted average, based on probability of each world state. We'll call the result the *average control* that a signal has over the act.

Formally, the measure is the expectation of causal specificity over all world states:

$$\mathbb{E}_W\left(I(\hat{S};A)|\hat{W}\right) \tag{5.1}$$

which is equivalent to:

$$I(\widehat{S};A|\widehat{W}) = \sum_w p(\widehat{w}) \sum_s p(\widehat{s|w}) \sum_a p(a|\widehat{s},\widehat{w}) \log \frac{p(a|\widehat{s},\widehat{w})}{p(a|\widehat{w})} \tag{5.2}$$

Calculating this quantity amounts to doing a series of intervention experiments. We place our organism in one world state, wiggle the signal, and measure the specificity it has for the act. We then place it into a second environment, wiggle the signal, and again measure the specificity. Finally, we sum these results weighting each specificity measurement by the probability of the corresponding world state.

Let us see how this works with our first example. We shall assume that the probabilities of the two world states are $P(W_1 = OFF) = 0.8$ and $P(W_1 = ON) = 0.2$, and that both players' strategies simply map the incoming signal or cue to the corresponding act or signal. Thus when the world state $W_1 = ON$, the signals will be $S_1 = ON$ and $S_2 = ON$, and the action will be $A = ON$; similarly for when $W_1 = OFF$. Given the strategies above, it follows that the probabilities of the signals map directly to those of the world states: $P(S_1 = OFF) = 0.8$ and $P(S_1 = ON) = 0.2$. These are the natural probabilities without any interventions, and we'll use these same probabilities to manipulate the channel S_1 in each of the world states. The probabilities of the acts are likewise $P(A = OFF) = 0.8$ and $P(A = ON) = 0.2$.

Given this setup, the mutual information in channels S_1 and S_2 is ≈ 0.72 bits. To do the work we want, our new measurement should provide a different value for these channels. We can get a

sense of how measuring the information in S_1 and S_2 will differ by looking at how manipulating the signals moves probabilities. Recall that to construct his information measure Skyrms began with a 'key quantity', which was how much *seeing* a signal moves the probabilities of a state or an act. Here we look at how the signals move the probabilities of the acts *when they are manipulated*. Our key quantity is this ratio (which can be found in the definition above):

$$\frac{p(a|\widehat{s_n})}{p(a)} \tag{5.3}$$

We only need examine a subset of these to see how differently they treat the two signalling channels. Suppose we fix the world state to $W_1 = OFF$, and look at the effect of manipulating S_1, setting it to ON (for simplicity, we'll drop conditioning on the world state, assuming it is fixed to OFF). We want to see how it changes the probability of A, by looking at the ratio:

$$\frac{p(A = ON|\widehat{S_1} = ON)}{p(A = ON)} = \frac{1}{0.8} \tag{5.4}$$

When $W_1 = OFF$, manipulating the signalling channel S_1 so that S_1=ON raises the probability of A from 0.8 to 1.

Now consider doing the same thing with channel S_2. Again, we fix the world state to $W_1 = OFF$, and manipulate W_2=ON:

$$\frac{p(A = ON|\widehat{S_2} = ON)}{p(A = ON)} = \frac{0.8}{0.8} \tag{5.5}$$

Manipulating W_2 to ON makes no difference to the probabilities of the act, and this is reflected in the fact that the value of ratios is 1. In the full equation of specificity given above, we take the logarithm of this ratio, and obtain zero bits. Indeed, when we compute the full equation across both world states, the amount of information in signalling channel S_2 is zero, for manipulating S_2 never changes the probabilities. In contrast, the same equation computed on channel S_1

gives us, ≈ 0.72, the exact amount that Skyrms's information measure gave.

5.1   Examples 2 and 3

How does this measure perform in our other examples, where the distinction between correlation and causation cannot be simply read off the structure of the network? Recall that, in examples 2 and 3, Skyrms's information measure was insensitive to changes in how information flowed through different channels in the network. These changes were driven by the probability of a second cue from the environment. Let's look at the effect that varying the probability of W_2 has on our information measure in the different signalling channels.

First, with example 2, we measure the information in both S_1 and S_2 as the probability of W_2 increases from zero to one.
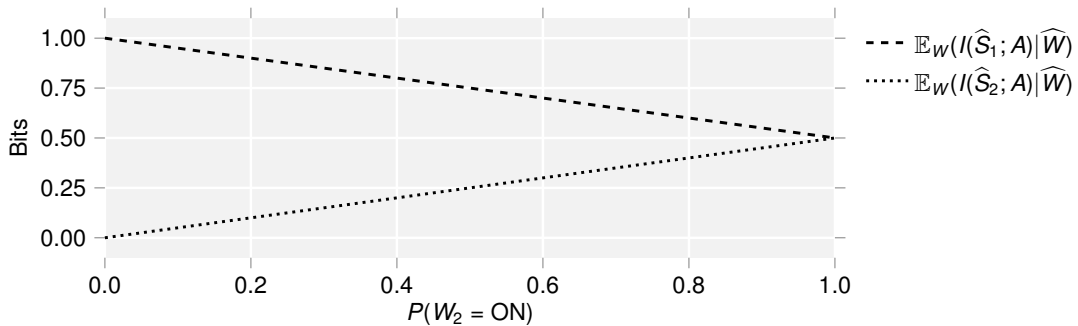
Figure 6: The result of gradually modifying the probability that $W_2 = ON$, using our suggested information measure for acts. Both channels now carry information that changes as $p(W_2)$ is modified.

Recall that, with the simple mutual information measure, both channels (S_1 and S_2) contain the same information regardless of the probability of W_2. Using our modified information measure, we see that W_2 affects both of these channels. As the probability of W_2 increases, the quantity of information transmitted by S_2 increases, and at the same time, the quantity of information transmitted by S_1 *decreases*. From the perspective of mutual information, we saw that our second channel was *redundant*. But our new measure doesn't show redundancy. Rather, information is spread across both channels. Eventually, when W_2 is always ON, both signalling channels have exactly the same amount of information about A (0.5 bits each).

In our third example we witness a similar effect. Recall that, in this case, both channels were causally relevant in different contexts, and there was no redundancy. Here, we see that the information in channel S_2 increases as the probability of W_2 increases whilst the information in S_1 decreases. Now, however, S_2 eventually reaches 1, and S_1 eventually goes to 0.
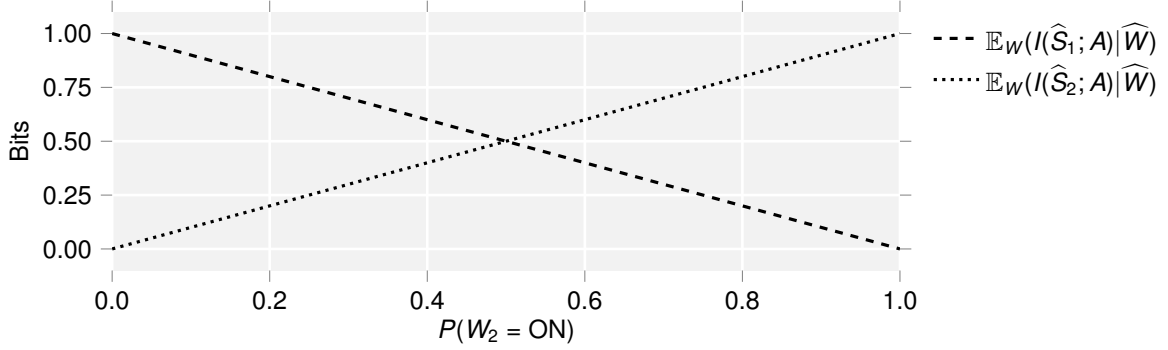

Figure 7: Modifying the probability that $W_2 = ON$ switches control from one signalling chain to another, a fact reflected in the information measure we suggest is appropriate for acts.

In both of these examples, we see that our information measure is sensitive to the structure of the network. This reflects how much information is flowing through that channel to the act, or how much control that channel has over the act.

5.2 Average Control Implicitly 'Holds Fixed' other Pathways

The information measure we have proposed tells us how much control, on average, a signalling channel has over the acts that a signalling network produces (assuming we limit our interventions to the natural distribution of the signalling variable). We can gain further insight into this measure by exploring the relation it has to another approach to measuring causality in complex networks: Ay and Polani's *information flow* (Ay and Polani [2008]).

Ay and Polani were interested in capturing how information flows and is processed in complex systems. They noted that a number of previous attempts that mention *flow* in complex networks really only capture correlations in the system, and that:

> ... a pure correlative measure does not precisely fit the bill. Different parts of a
> system may share information (i.e. have mutual information), but without infor-

24

mation flowing between these parts. Rather, the joint information stems from a common past. (Ay and Polani [2008], p. 17).

This is precisely the problem that our examples highlighted, and Ay and Polani's solution is to provide a mutual information measure that builds in interventions, much as we have done above. Ay and Polani's approach is to ask how much causal influence one variable has over another, given you are already controlling for, or holding fixed, a further set of variables.[11] They write this measurement as $I(X \rightarrow Y|\widehat{Z})$, which can be read as 'the information flow from X to Y given that we *do Z*' (see Appendix 1 for further details).

Let us assume we want to apply their measure to capture the causal flow from a signal channel to the act, where there are multiple causal pathways between the world state variables and the act variables (as in examples 2 and 3). Because the information transmitted along these other pathways may interact, we hold fixed (or control for) all channels that lie on other pathways between world states and acts that don't pass through our focal signalling channel. So, in example 2, we would measure the flow of information from S_1 to A, whilst controlling for S_3: $I(S_1 \rightarrow A|\widehat{S_3})$. This would tell us how much control S_1 has over A, after we've excluded the control this second pathway has (the signalling chain that connects W_1 and W_2 to A).

This particular way of measuring information flow, where we control for all other pathways, is equivalent to average control that S_1 has over A (see Appendix 1 for details). By simply averaging specificity over the different world states, we effectively control for all other signalling channels that can affect the behaviour. Given the structure of these signalling networks— where information flows from world states to actions via signals, Ay and Polani's measure is equivalent to average control. This equivalence makes it clear that the information flow from a signalling channel to the actions is sensitive to more than just changes to channels that lie on the pathway between it and act: it can also be affected by changes to other parts of the network.

---

[11]It is possible to condition on *no* other variables (the empty set), or multiple other variables. Note that multiple variables can be collapsed to single variable by taking the Cartesian product over the states of the various variables and using their joint probability (Pearl [2000], p. 9).

Our aim was to construct an information measure that captured the idea of *flow* within signalling networks. We've argued that this notion is equivalent to the *average control* that manipulating a signal has over an act, and this averaging effectively provides a way of holding fixed, or controlling for, other signalling pathways. An important feature of this measure is that it delivers precisely the same quantity in simpler networks (those without multiple pathways) as Skyrms's measure does. So it both tells why these measures are distinctive and why, in simpler networks, we may not recognise that these two ideas are distinct.

6   How Does Evolution Create Information?

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 (Skyrms [2010b], p. 44).

Our focus thus far has been to separate two distinct ideas about information in signalling networks: the *flow* of information, and *carrying* information. A key claim in Skyrms's book, however, is that evolutionary dynamics acting on signalling networks can *create information*. We now show how our distinction can be brought to bear on these evolutionary claims as well.

Consider our first example again, where S_2 is a signalling channel that flows nowhere, but is correlated with a second signalling channel S_1 connected to the act. We suggested that a key difference between these two channels is that S_1 *explains* why the organism acts as it does in the different world states, and S_2 does not. If we think of the signalling network as a causal graph, this idea can be borne out, because intervening on S_2 will not affect the act. Our suggested measure of average control also reflects this causal reading, telling us that there is zero flow of information from S_2 to A, but 1 bit of information flowing from S_1 to A.

If we assume the signalling network in this organism was the result of some evolutionary process, then we could offer a Skyrms-style explanation for how selection had created information

in signalling channel S_1: it was the result of a symmetry-breaking process in which some conventional information-carrying signals evolved between the sender and receiver. Note, however, that given our stipulation that S_2 is correlated with S_1, the information carried by signalling channel S_2 was also created by evolution. Clearly, evolution can create information that is carried by some signalling channels even if those channels themselves don't participate in the coordination game between the sender and receiver.

If we wanted to *explain* what signalling channel drove the evolutionary change, however, we would refer to channel S_1, for that is the channel that is responsible for connecting world states with the acts, and plays a role in generating the organism's fitness.[12] It is also the channel which connects the sender and receiver that are playing the game. So whilst using Skyrms's information measure informs us about results of adaptation, it cannot distinguish between the different roles that these two signals played in the adaptive process. These different roles are reflected in a well-known distinction in philosophy of biology: there is *selection-for* channel S_1 but merely *selection-of* channel S_2 (Sober [1984]).

The measure of information flow we have constructed—what we have called *average control*—can distinguish between these two roles, for it tells us which signal was selected-for. The same point extends to other examples, where it is the flow of information from a signal to an act that tells you how causally relevant that signal is in driving or maintaining the selection on the organism, rather than simply coming along for the ride. Evolution may result in information being *carried* in numerous signals, but for any information to be created at all, there must be a flow of information from some signals to the act. From a causal perspective, at least some signals must be *difference-makers* for the act in order for selection to be effective.

---

[12]As we discussed previously, the fact that a signalling channel correlates with, but does not flow to, the acts of the organism can be explanatory in some contexts (such as how the organism was exploited). But in the evolutionary model we are focused on here, it plays no role in explaining how the system was selected.

7 Conclusion

We have argued that there are two distinct uses of information at play in Skyrms's work, and have provided a new measure that captures the flow of information in signalling networks by drawing on recent work on causal specificity.

This measure has some straightforward, practical implications. If you analyse complex networks where there are multiple channels from world states to acts, and hence where signals may share information, then you should use a causal measure if you want to capture the flow of information from signals to acts. If you do not, you may fail to distinguish the different contributions that various signalling channels make to the success or failure of a network, and thus fail to accurately reflect the role that signals play in generating the behaviour of the network, and the role signals have in driving selection.

In networks with a single channel that Skyrms and others have analysed, these situations don't arise. In such simple cases (which are easy to identify by simply inspecting the network) you could continue to use mutual information, as it delivers exactly the same result. But this would miss the point. Our examples show that talk of flow in signalling networks is a causal concept. This is a crucial addition to a naturalistic theory about signalling and information. For if biological information is not to be merely 'driven by adaptive dynamics' (Skyrms [2010b], p. 44), but actually play an explanatory role in driving these dynamics, then the information in these biological systems cannot sit idly by, it must actually *do something*.

Appendix A   Average control and information flow.

In this appendix we explain how averaging the control of S for A over the values of the world-state W amounts to controlling for the variables in the signalling network which are not on the $W \rightarrow S \rightarrow A$ path. We start by building a canonical causal graph representing a signalling network, then we show the equivalence between the measures of average control and information flow.

A.1   A canonical causal graph for signalling networks

For ease of presentation, we consider only the paths which end up affecting the variable A. (If the paths don't affect A, then by definition they don't affect the average control for A or the information flow to A.) In this appendix the variable S is by definition upstream to A; and W represents the set of all root variables. W may affect A through affecting S and/or through another path. To reduce the graph to its simplest form (without loss of generality), other variables on these paths are not represented explicitly and are contained within the causal arrows (recall that these arrows represent mappings between values of the cause and values of the effect, and are thus blind to the existence or not of intermediary variables in a more detailed causal graph). See Figure 8A for our canonical signalling network (note that for the reasoning below to apply, the arrows from W need not necessarily exist).

A.2   Measuring average control and information flow

The measure of average control described in this paper consists of a two step procedure:

1. We fix (by an ideal intervention) the world-state W with its natural probability distribution,

2. In this world-state, we look at the causal specificity of S for A by intervening on S using the natural probability distribution for S. The causal specificity of S for A can be altered

by the value of W.

The formula reads as follows:

$$I(\widehat{S}; A|\widehat{W}) = \sum_w p(\widehat{w}) \sum_s p(\widehat{s|w}) \sum_a p(a|\widehat{s}, \widehat{w}) \log \frac{p(a|\widehat{s}, \widehat{w})}{p(a|\widehat{w})} \tag{A.1}$$

where, by hypothesis, $p(\widehat{w}) = p(w)$ and $p(\widehat{s|w}) = p(s)$. By definition of causal specificity, we have $p(a|\widehat{w}) = \sum_s p(\widehat{s|w})p(a|\widehat{s}, \widehat{w})$; that is, A is observed in a set-up where both S and W are subject to interventions. Thus we also have: $p(a|\widehat{w}) = \sum_s p(s)p(a|\widehat{s}, \widehat{w})$.

This average control $I(\widehat{S}; A|\widehat{W})$ is equivalent to the information flow from S to A when controlling for the path (if any) from W to A which does not go through S. To control for this path, we have to slightly modify our canonical network, for the only way to control for the direct $W \to A$ path would be, for the moment, to control for the variable W, which may in turn also affect S.[13]

To circumvent this obstacle, we introduce a *ghost* variable W' in the network. This ghost variable takes the value of W and affects all variables which are not on the path $W \to S \to A$ exactly as if it were W, but it does not affect the path $W \to S \to A$. This ghost variable W' is a purely theoretical entity introduced in the graph to ease calculus, and introducing such a variable is always possible in a causal graph. Ghosting the variable W into W' can be thought of as applying an operator like Pearl's *do*() operator, with the difference that this ghost operator is defined with respect to a variable (here W) and a path (here $W \to S \to A$). Controlling the variable W' enables us to control all information flowing through the (previously direct) $W \to A$ path (for a similar approach on controlling paths, see Janzing *et al.* [2013]). The new causal graph now appears in Figure 8B.

By definition of information flow (Ay and Polani [2008]), the formula of the information flow from S to A conditional on W' reads as follows:

---

[13]An easy calculation shows that when W fully determines S, the information flow conditional on W is null, whatever the influence of S on A: $I(S \to A|W) = 0$. This is because knowing W already tells us everything that S could tell us about A.

$$I(S \rightarrow A|W') = \sum_{w'} p(w') \sum_{s} p(s|\widehat{w'}) \sum_{a} p(a|\widehat{s}, \widehat{w'}) \log \frac{p(a|\widehat{s}, \widehat{w'})}{\sum_{s'} p(s'|\widehat{w'}) p(a|\widehat{s'}, \widehat{w'})} \qquad (A.2)$$

By hypothesis, we have the following equalities: $p(w') = p(w)$ (W' mimics W), $p(s|\widehat{w'}) = p(s)$ (W' does not affect S), $p(a|\widehat{s}, \widehat{w'}) = p(a|\widehat{s}, \widehat{w})$ (since W' mimics W with respect to A). It is therefore easy to see that the formulas A.1 and A.2 are equivalent: $I(\widehat{S}; A|\widehat{W}) = I(S \rightarrow A|W')$.
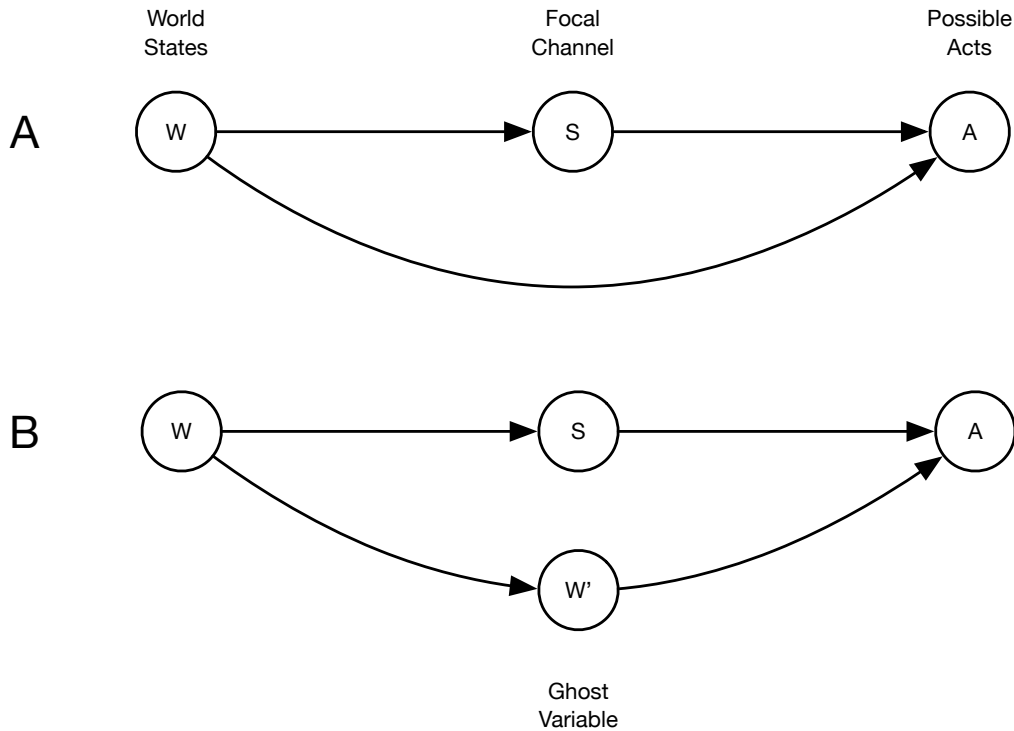

Figure 8: A. The canonical signalling graph. B. Adding a *ghost* variable separates the focal channel from all other channels stemming from W.

Funding

Acknowledgements

Brett Calcott

Department of Philosophy and Charles Perkins Centre

University of Sydney

NSW, Australia

brett.calcott@gmail.com

Paul E. Griffiths

Department of Philosophy and Charles Perkins Centre

University of Sydney

NSW, Australia

paul.griffiths@sydney.edu.au

Arnaud Pocheville

Department of Philosophy and Charles Perkins Centre

University of Sydney

NSW, Australia

arnaud.pocheville@sydney.edu.au

References

Ay, N. and Polani, D. [2008]: 'Information flows in causal networks', *Advances in complex systems*, **11**(01), pp. 17–41.

Calcott, B. [2014]: 'The Creation and Reuse of Information in Gene Regulatory Networks', *Philosophy of Science*, **81**(5), pp. 879–890.

Cao, R. [2014]: 'Signaling in the Brain: In Search of Functional Units', *Philosophy of Science*, **81**(5), pp. 891–901.

Crapse, T. B. and Sommer, M. A. [2008]: 'Corollary discharge across the animal kingdom', *Nature Reviews Neuroscience*, **9**(8), pp. 587–600.

Godfrey-Smith, P. [2014]: 'Sender-Receiver Systems within and between Organisms', *Philosophy of Science*, **81**(5), pp. 866–878.

Griffiths, P. E., Pocheville, A., Calcott, B., Stotz, K., Kim, H. and Knight, R. [2015]: 'Measuring Causal Specificity', *Philosophy of Science*, **82**(4), pp. 529–555.

Janzing, D., Balduzzi, D., Grosse-Wentrup, M. and Schölkopf, B. [2013]: 'Quantifying causal influences', *The Annals of Statistics*, **41**(5), pp. 2324–2358.

Korb, K. B., Hope, L. R. and Nyberg, E. P. [2009]: 'Information-Theoretic Causal Power', in F. Emmert-Streib and M. Dehmer (*eds*), *Information Theory and Statistical Learning*, Boston, MA: Springer US, pp. 231–265.

Lewis, D. [1969]: *Convention: a philosophical study*, Harvard: Harvard University Press.

Pearl, J. [2000]: *Causality: models, reasoning and inference*, vol. 29 Cambridge Univ Press.

Planer, R. J. [2013]: 'Replacement of the "genetic program" program', *Biology & Philosophy*, **29**(1), pp. 33–53.

Pocheville, A., Griffiths, P. E. and Stotz, K. [In Press]: 'Comparing Causes: An Information-

Theoretic Approach to Specificity, Proportionality and Stability', in H. Leitgeb, I. Niiniluoto, E. Sober and P. Seppälä (*eds*), *Proceedings of the 15th Congress of Logic, Methodology and Philosophy of Science*, London: College Publications.

Skyrms, B. [1996]: *Evolution of the Social Contract*, Cambridge University Press.

Skyrms, B. [2010a]: 'The flow of information in signaling games', *Philosophical Studies*, **147**(1), pp. 155–165.

Skyrms, B. [2010b]: *Signals: Evolution, Learning, and Information*, Oxford ; New York: Oxford University Press, 1st edition.

Sober, E. [1984]: *The Nature of Selection: Evolutionary Theory in Philosophical Focus*, University of Chicago Press.

Van Haastert, P. J. M. and Veltman, D. M. [2007]: 'Chemotaxis: navigating by multiple signaling pathways', *Science's STKE: signal transduction knowledge environment*, **2007**(396), pp. pe40.

Woodward, J. [2003]: *Making things happen: A theory of causal explanation*, Oxford University Press.

Woodward, J. [2010]: 'Causation in biology: stability, specificity, and the choice of levels of explanation', *Biology & Philosophy*, **25**(3), pp. 287–318.