Neural Signalling of Probabilistic Vectors

Nicholas Shea

A paper from the PSA 2012 Symposium
Signalling Within the Organism

Nicholas Shea
Department of Philosophy
King's College London
Strand
London WC2R 2LS
U.K.

Neural Signalling of Probabilistic Vectors

Nicholas Shea

Abstract

Recent work combining cognitive neuroscience with computational modelling suggests that distributed patterns of neural firing may represent probability distributions. This paper asks: what makes it the case that distributed patterns of firing, as well as carrying information about (correlating with) probability distributions over worldly parameters, *represent* such distributions? In examples of probabilistic population coding, it is the way information is used in downstream processing so as to lead to successful behaviour. In these cases content depends on factors beyond bare information, contra Brian Skyrms' (2010) view that representational content can be fully characterised in information-theoretic terms.

1.   Introduction

In *Signals* Brian Skyrms argues for the existence of "non-propositional" contents (2010, 34-42). A "propositional" content asserts just that a particular state of affairs obtains, or that one out of a proper subset of available states of affairs obtains. For example, in a world with three states, the content *S1 or S2 obtains* would count as propositional. Non-propositional contents don't merely rule out some states of affairs and rule in others; they also assign probabilities to those states (or some other measure related to probability). So a non-propositional content for a signal M might say that the probability of state S1 given M is 0.6, the probability of state S2 given M is 0.4, and the probability of every other state given M is zero. Skyrms' particular version of the non-propositional idea is an informational content vector that captures how much the signal has increased or decreased the probabilities of each world state from their unconditional probabilities.

> '[on the propositional view] … the information content 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.' (Skyrms 2010, 34)

[The informational content vector tells us how the probability of each state is moved by the signal.] '… 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. [The informational content vector] specifies more than the propositional content. Furthermore, some signals will not have propositional content at all.' (Skyrms 2010, 42)

There is nothing novel in the idea that representations can represent probabilities of world states. We do that frequently in natural language, as well as in the formal languages of mathematics and decision theory. But it is much less common to treat a single undifferentiated non-conceptual representation as having a content that concerns probability, much less a probability distribution over a range of possible states. Skyrms argues that standard philosophical treatments of content are too restrictive, in that they focus on contents of the rule-in/rule-out sort. They should also be giving an account of contents that consist of probability distributions of world states, of which rule-in/rule-out contents are just a special case. It seems to me that Skyrms has correctly identified an aspect of content that has been under-explored in philosophical discussions of content to date.

This paper argues that Skyrms' insight is apposite beyond the mathematical models of signalling with which he and his collaborators were concerned. Recent evidence from cognitive neuroscience makes a strong case for probabilistic representations in the brain. It is by calculating over representations of probability distributions that agents are able to obtain near-optimal levels of reward when making perceptual decisions about noisy stimuli. Section 2 argues that probabilistic population coding in the brain is an instance of the phenomenon that Skyrms has labelled "non-propositional content". In probabilistic population codes a single non-conceptual representation, lacking any semantically-significant constituent structure, plausibly represents a probability distribution over a worldly parameter (like direction of motion).

While it supports one of Skyrms' claims, the neuroscientific data on probabilistic population coding also undermines another claim made by Skyrms, namely that facts about information, in the sense of information theory (conditional probabilities and the like) are adequate to give a complete account of everything content-relevant that is going on in cases of signalling. The cognitive neuroscientific models only appeal to a very limited subset of the information carried by internal states in order to explain behaviour. Section 3 argues that, in order to answer the question of why it is that particular information that is explanatorily relevant, rather than all the other information also carried by internal states, we need to appeal to the way the internal states are used in downstream processing; in particular to aspects of the information they carry that are relied on in generating behaviour, that is in harvesting near optimal rewards for the task in which the agent is engaged. Therefore payoffs or stable task performance also have an ineliminable role to play in fixing

representational content in the cognitive neuroscientific case. That conclusion might also apply back to the mathematical models of signalling with which Skyrms was concerned.

2.   Probabilistic population coding in the brain

What is it to represent ('code for') a probability distribution? An influential example is the model in Ernst & Banks (2002), which is based only on behavioural data. Subjects in their experiment were asked to judge the height of a bar relying on both visual and haptic (touch-based) information. Ernst & Banks argued that subjects represent that the bar has a range of possible heights, representing a normal distribution of the probability of different heights by representing the mean and variance of that distribution. Ernst & Banks obtained behavioural evidence that such probability distributions are first represented separately for visual and haptic information, before being optimally combined based on the respective mean and variance of the two distributions. So one form of possible probabilistic representation is to have variables separately representing the mean and variance of a normal distribution of possible values of a parameter in the world.

A more recent example of probabilistic coding is a 'mechanistic model', in that it is partly based on data about how the putative representations are realised in the brain (in neural firing rates), and so goes beyond Ernst & Banks' 'descriptive model'. Beck et al. (2008) offer an account of how macaques make a perceptual decision about the preponderant direction of motion of a collection of moving dots. In each stimulus a proportion of the dots all move in the same direction (e.g. 51% may move right at 45° to the horizontal), with the rest of the dots moving in random directions. The monkey's task is to make an eye movement in the direction of the coherent motion, for which it is rewarded if correct.

Single unit recording has shown that individual neurons in various neural areas have receptive fields that correspond to a particular direction of motion in the visual stimulus. Other neurons have receptive fields that correspond to the reward associated with a particular direction of motion. For example, a given neuron might be most likely to fire when the monkey is viewing movement at 45%, with a decreasing probability for nearby angles dropping off to zero within 120° to either side. The standard interpretation is that a distributed pattern of firing across a collection of neurons in a given brain area represents a single direction of motion, the most likely direction of motion. The direction of motion (a single scalar) can be decoded by means of a weighted sum of instantaneous firing rates. The reliability and tractability of this form of coding has been studied extensively.

More recently Pouget et al. (2003) have argued that a distributed pattern of firing of neurons with such receptive fields is actually representing something more complicated: not just a single scalar value of a worldly parameter, but a probability distribution of that parameter *s* across the range of values that it can take (e.g. *s* could

be direction of motion from $+180^O$ to -180°). They show that the posterior probability that the worldly parameter s lies within any given range can be decoded from the distributed pattern of firing, for all possible values of s. That is, we can generate a probability density function giving the probability over all values of s, based just on a linear combination of instantaneous firing rates (with weights that depend on s).

Beck et al. (2008) obtained evidence that representations of this form were carried in the neural firing that is responsible for monkeys' near-optimal performance in a perceptual decision making task. At the first stage neurons in area MT represent the instantaneous probability that the stimulus right now is moving in direction s (for all directions s). Downstream area LIP integrates over time, representing the probability that the stimulus direction is s (for all s) based on the whole time that the monkey has been observing the stimulus. Beck et al.'s major contribution was to show how the calculation of accumulated evidence, which is potentially very complex mathematically, could be implemented in neural firing rates. (Provided the firing patterns of individual neurons are 'Poisson-like', i.e. the mean is proportional to the variance, and uncorrelated, the calculation can be achieved by individual LIP neurons simply adding the activity of individual MT neurons.) At the final stage the representation of the integrated probability distribution of s is used to calculate the most likely value of s, and an eye movement is programmed in that direction. The animal's optimal performance in obtaining rewards in the task is explained by its having an internal mechanism which processes representations of probability distributions, performing calculations on those distributions in order to work out the most likely direction of motion, which in these experiments is identical to the direction of stimulus motion that is most likely to be rewarded (and to the direction of eye saccade that is most likely to be rewarded).

Pouget et al. (2003)'s probabilistic population coding is just one way in which probability distributions could be encoded in the brain. The analysis in Beck at al. (2008) was not aimed at testing which particular probabilistic coding scheme is in use, but in deciding between probabilistic coding and the standard model in which only representations of scalar quantities are involved.

A natural but mistaken thought is that the pattern of firing rates distributed across a collection of neurons with different receptive fields always did represent probabilities: each cell's firing rate representing the relative probability that the value of the stimulus is in its preferred direction. Or perhaps the highest firing rate just directly represents the most likely direction. However, that is not how the standard model understood the distributed code, and for good reason. To rely on the firing rate of just one cell to calculate the probability of a particular direction s would be to throw away a large amount of useful information about the probability of s, namely information carried by other cells whose receptive fields also overlap with s (because their preferred direction is some different nearby direction s'). Indeed, to discriminate s from very nearby values, the firing rates of other neurons are more useful, since

their rates change more for small changes of direction than do cells for which s is the preferred direction.

The presence of noise also highlights the importance of information carried by cells with other preferred directions. Because of noise a single cell will vary a lot in how it responds from occasion to occasion, even to stimuli exactly in its preferred direction. The effect of that noise can be reduced by considering information from a large array of noisy cells, each of which has some sensitivity to stimuli in direction s. So compare the noisy distribution of firing rates in Pouget et al. (2003) with the smooth probability distribution that is decoded from them (Fig. 1 below). This point is also apparent if we compare the firing rates and probability distributions illustrated in Beck et al. (2008) – Fig. 2 below. Plotted in terms of preferred direction, neural firing covers a wide range of angles, whereas the probability distribution decoded from that distributed pattern of firing is narrower, and narrows further as evidence accumulates over time.

[Insert Figure 1 about here]

Fig. 1. From Pouget et al. (2003). The noisy distribution of population activity in the top panel (green, blue and red open circles) is transformed by a Bayesian decoder into the smooth posterior probability distributions in the lower panel (green, blue and red curves, respectively).

[Insert Figure 2 about here]

Fig. 2. From Beck et al. (2008). The firing rates of model neurons in panel A are transformed by a Bayesian decoder into the probability distributions in panel B. At each time the width of the probability distribution is narrower than the range of preferred directions which show non-zero firing rates. (Additionally, the variance of the distribution decreases over time as increasing evidence accumulates.)

In short, Beck et al. (2008) offer a representational explanation of how monkeys manage to harvest near optimal rewards in a task that involves judging the preponderant direction of motion of a visual stimulus. They offer behavioural and neural evidence for the conclusion that a three stage computation is responsible for this behaviour, and that at two stages distributed patterns of neural firing represent a probability distribution across the range of possible directions of motion s. Each distributed pattern of firing is a single non-conceptual representation, without semantically-significant constituent structure, analogous to a single signal in Skyrms'

models. There is nothing in this representation that corresponds separately to states and probabilities (as there is when we represent probabilities of world states mathematically, for example). Each different pattern of firing represents a different probability distribution, and the way that they do so explains how a calculation can be performed that first represents an instantaneous probability distribution, then transforms those into an integrated probability distribution, and finally calculates from that a maximum likelihood estimate of the stimulus direction. This part of cognitive neuroscience therefore offers strong support for Skyrms' idea that some simple signals can do more than just rule-in and rule-out various world states but also represent probabilities over world states. It shows that that idea is a plausible hypothesis about the working of some real information processing systems underlying primate behaviour.

3.   These contents depend on more than just information

In pointing to the existence of "non-propositional" contents, Skyrms argues that informational treatments of content are more general than standard philosophical approaches, in that they extend to more cases. He also argues that they cover all the cases – they capture everything that standard accounts of content do and more. The force of the arguments and analyses collected in *Signals* is that information-theoretic properties capture everything that there is to say about the meaning or content of signals, at least in these simple cases. To put it another way, facts about information, in the broadly correlational sense of information theory, are adequate on their own to ground facts about meaning or content.

> '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. … I believe we can do better by using a more general account of informational content.' (Skyrms 2010, 34)

> '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. … 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 quantity of information in a signal at that time.' (Skyrms 2010, 35)

> 'There is no mystery about the meaning of signals. The object of primary interest is *information*. Signals naturally acquire information of a variety of types, and they can carry information both in and out of equilibrium. The Platonist's propositional notion of meaning is an abstraction from one kind of

> information in perfect equilibrium in an idealized signalling game.' (Skyrms 2010, 177)

While probabilistic population coding in the brain supports the suggestion about non-propositional representation, it also cuts against the idea that facts about information, taken alone, are adequate to ground facts about meaning or content. Facts about the information carried by distributed patterns of firing in various neural areas are not, taken alone, sufficient to make it the case that the representations in the system have the contents picked out by Beck et al. (2008).

Part of Beck et al.'s motivation was indeed the finding, based on neural data, that distributed patterns of activation in the brain probably do carry information about probability distributions. But that was not enough on its own to support their argument that probability distributions are being coded for. After all, many rival coding theories also concern things about which the activation patterns carry information. For example they carry information about the scalar values adverted to in standard non-probabilistic population coding. The argument that what is represented are probability distributions also relies facts about which representations could be used to perform certain tasks. Performing calculations over represented probability distributions offers a general account of how the brain could manage to perform such decision making tasks near-optimally, and so is plausibly how the brain manages to perform the decision making task in the experiment in question near-optimally. Not all the information carried counts as content, only that information the reliance on which forms part of an account of how the agent manages to obtain the kinds of rewards it does.

This point is just an instance of the familiar general point that information is ubiquitous. For example, the distributed pattern of firing in LIP carries information, not just about the integrated probability distribution over motion directions, but also about lots of local facts: the patterns of firing in upstream area MT, the patterns of firing in downstream area SCb, the patterns of light on the animal's retina, the patterns of light passing between the screen and the animal, the direction in which the animal is likely to move its eyes, and so on. It also carries some information, albeit less reliably, about other properties of the stimuli: the number of coherently-moving dots present, the total surface area of dots moving in each direction, and so on. Are all these other pieces of information, which are carried by distributed patterns of firing in area LIP, also contents or meanings carried by those signals? It is a familiar philosophical observation that not all of them can be (the disjunction problem). The neuroscientific case study backs this up, because only a very particular aspect of the information carried by a signal is picked out and identified as the content which is being coded for.

The constraint that appears to be operating in the neuroscience case is that the correct account of what is being coded in various areas should feed into an explanation of how the animal manages to behave as it does – which in these cases is

to achieve near-optimal rewards. The monkeys come to have these useful configurations of internal states through a process of learning how to perform in a way that maximises the delivery of rewarding feedback (e.g. of volumes of juice). That fits well with an approach to content that combines teleological considerations (Millikan 1984, Papineau 1987) with facts about the correlational information carried by internal states (Shea 2007).

In this case a kind of teleology is underpinned by learning: the monkeys have come to have an internal mechanism configured as it is in virtue of a learning mechanism designed to optimise the delivery of reward (cf. Dretske 1988). Learning is also the process that stabilises sender-receiver configurations in some of Skyrms' models. The more general idea is that the operation of the system (internal processing in the neuroscience case, sender-receiver configurations in the signalling models) is subject to change and arrives at a stable configuration in virtue of how that configuration gives rise to results that lead to feedback that stabilises the configuration. In the replicator dynamics the result is replication and the feedback is selection. In reward-guided learning the result is the delivery of reward and the feedback is a learning rule that selects amongst internal configurations on the basis of maximising reward.

Does the population code for a raw probability distribution or for relative expected gain? In the experiment all directions are rewarded equally, so the two are confounded. However, the firing rates of LIP neurons are strongly affected by relative rewards (Platt & Glimcher 1999). The orthodoxy in the past decade has been that LIP neurons code for rewards (of either an action-based or goods-based sort).[1] This would suggest that the distribution identified by Beck et al. (2008) is in fact a distribution of expected rewards: probability times reward magnitude.

The fact that the monkeys manage to perform the task near-optimally given the quality of the information available to them acts as a further constraint on content. A content assignment according to which some of the representations were very inaccurate would be inadequate for explaining successful behaviour. For example, one might claim that all that is represented at every stage of this three stage internal process is the saccade direction that will be rewarded. Even the instantaneous evidence about motion direction available in the first stage of processing in MT correlates a bit with this parameter. So one could claim that neural activity in MT represents rewarded direction, only does so very inaccurately. That candidate content does not, however, contribute to an explanation of the animal's successful behaviour. Incorrect representation does not explain success. That is not to say that verificationism is operating. The probability distributions represented according to

---

[1] That orthodoxy may change in the light of recent evidence that LIP activity is more closely related to salience than to expected rewards, when the two are dissociated: Leathers & Olson (2012). (Negatively-associated stimuli provoke increasing responses as their negative importance increases (hence increasing salience) and their relative reward thereby decreases.)

Beck et al. need not match the objectively-correct probability distribution. But departures from accuracy are due to noise and other random processes. They are not a matter of making systematic errors.

Skyrms argued that in his signalling models the content or meaning of a signal is fully determined by facts about the probabilities of world states and the facts about sender and receiver strategies and their probabilities (2010, 35). Transposed to the neuroscience case, that becomes the claim that facts about content are fully determined by facts about how the internal states are produced and how they issue in behaviour. We saw good reasons to reject that conclusion in the case of probabilistic population coding. Facts about the external world task which that internal configuration is being used to achieve were also important to singling out one particular species of correlational information as being the content carried by distributed neural firing in an area. We can draw out Skyrms' claim slightly in the following simple argument:

(P1)   All the varieties of information carried by a signal at a time are fully determined by the probabilities of world states and the probabilities of sender and receiver strategies.

(P2)   The content or meaning of a signal at a time is fully determined by one or more of the varieties of information that it carries.

(C)   The content or meaning of a signal at a time is fully determined by the probabilities of world states and the probabilities of sender and receiver strategies.

We reject the analogue of (C) in the neuroscience case. Content facts are constituted by more than facts about inputs, outputs and probabilities of world states. But we can accept the analogue of (P1). Therefore we must reject (P2). Information on its own is not adequate to specify content. Facts about how reward is produced, and how feedback based on reward acts to stabilise internal configurations, also play an ineliminable role in making it the case that one part of the information carried by a state constitutes its representational content.

This approach to content fits well with Peter Godfrey-Smith's idea that representational content is a matter of system making use of an 'exploitable relation' between a set of internal states and a set of states of the world (Godfrey-Smith 2006). All the various varieties of information carried by a distributed pattern of firing are exploitable relations that could be made use of. As it turns out, the system is actually making use of the relation between patterns of firing rates and probability distributions over the motion parameter.

That also fits well with an account of the explanatory role of content I have argued for elsewhere (Shea 2013). For these simple systems, realist representational contents arise when the system is embedded in the context of some externally-specified task which it manages to achieve and is designed in some way to achieve.

The system performs the task by breaking down the externally-specified function into a series of steps, and the internal configuration shows how the system manages to go through those steps in order to achieve the outcome it does. Each internal step bears an exploitable relation to a state that is relevant to the task, and the whole system relies on those relations so that the internal steps faithfully instantiate stages that constitute a calculation of the externally-specified function. (In these cases 'design' is a matter of performing the task in a way that contributes to the stabilisation of that very internal configuration.)

Skyrms does not in fact make the strong claim that all meaning reduces to facts about information. He thinks that some complex systems (thinkers) may have a kind of mentalistic intentionality. Furthermore, in some cases there is teleosemantic intentionality. But these are both add ons – things that can be added to the theory, where they apply (2010, 43). Skyrms is clear that neither mental intentionality nor teleosemantic intentionality is a requirement for meaning or content. Facts about information are sufficient on their own to constitute a substantive kind of meaning or content, and do so whether or not teleosemantic intentionality is present. The neuroscientific case study above furnishes a reason for resisting that conclusion.

These observations apply both to Skyrms' signalling models and to the three-stage internal calculation in Beck et al. (2008). But note that the neuroscientific computation is not a straightforward instance of a Skyrms-type sender-receiver game. Nor are its states getting content in the standard teleosemantic way, by considering how they are used by a consumer system (and the stabilising functions and success conditions of such uses). It is sometimes suggested that in the brain each subsequent stage of processing is the consumer for the representations tokened at the immediately preceding stage. But something different is underpinning the contents in the Beck et al. case, according to my account. It is the function of the complex system as a whole that provides the foundation for content. An interacting network of internal components manages to achieve that function and is suited to doing so because of the exploitable relations each component bear to relevant world states. Each element is exploiting the fact that the other elements with which it interacts bears such exploitable relations.

A potential challenge to that view, however, is that rival incompatible computational variants are available, each of which would be equally good as an instrumentalist account of how the system manages to perform its externally-specified function, and each of which is equally compatible with the facts about how internal processing within the system gives rise to behaviour. I don't know of a good argument that such rival variants will always be available (which would imply a worrying degree of content indeterminacy). Whether the actual data in cognitive neuroscience permits rival variant contents in some fields is an empirical question. It's still early days in this field, but so far it looks as if debates between rival neural coding theories are substantive and empirically tractable.

One of the things blocking rival incompatible contents is a fairly strong assumption about how representations should be realised in neural vehicles. The probabilistic representational contents are relatively explicitly represented, in that their values can be decoded from the representational vehicle via a linear combination of firing rates. That is not as direct as cases where the firing rate of neuron correlates directly with the value of the parameter it represents. This raises the question of how tight the connection between vehicles and contents has to be in order for vehicles to count as neural realisations of those contents. It can't be that anything goes, otherwise there is no way to distinguish instrumentalism from a commitment to real internal vehicles of content. Linear decodability is plausibly a direct enough link between vehicle and content. But are looser connections permissible? There is not space here to address that important question.
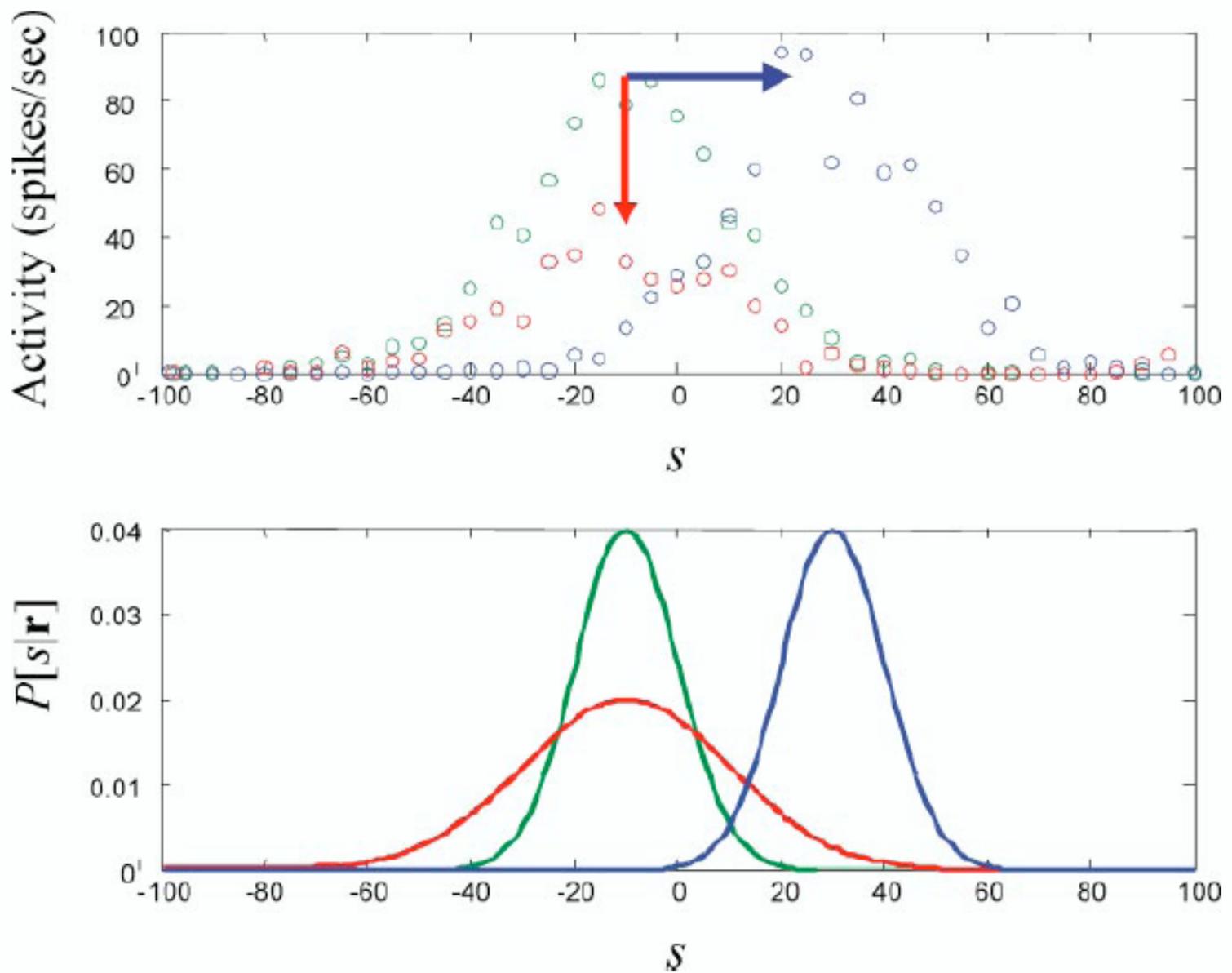
4.   Conclusion

Evidence from cognitive neuroscience for probabilistic population coding in the brain suggests that Skyrms' insight about the importance of non-conceptual contents that are probabilistic, and do more than just rule-in and rule-out world states, applies to real cases beyond those modelled by sender-receiver signalling games. However, this case study also implies that the way internal processing is stabilised by an externally-specified process is important to fixing content, in particular a connection between internal processing and payoffs plays a central role. Meaning in these cases is not just a matter of information of the right kind. The same conclusion may be equally apposite within Skyrms' models themselves.

References

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., et al. 2008. "Probabilistic population codes for Bayesian decision making." Neuron 60(6): 1142-1152.

Dretske, Fred (1988), *Explaining Behaviour: reasons in a world of causes*. Cambridge, MA: MIT Press.

Ernst, M. O., & Banks, M. S. 2002. "Humans integrate visual and haptic information in a statistically optimal fashion." Nature, 415(6870): 429-433.

Godfrey-Smith, Peter 2006. "Mental representation, naturalism and teleosemantics." In *New Essays on Teleosemantics*, ed. David Papineau and Graham Macdonald, 42-68. Oxford: Oxford University Press.

Leathers, M.L. and Olson, C.R. 2012. "Monkeys Making Value-Based Decisions, LIP Neurons Encode Cue Salience and Not Action Value." Science 338:132-35.

Millikan, Ruth Garrett 1984. *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.

Papineau, David 1987. *Reality and Representation*. Oxford: Blackwell.

Platt, M.L. and Glimcher, P.W. 1999. "Neural correlates of decision variables in parietal cortex." Nature 400:233-38.

Pouget, A., Dayan, P., & Zemel, R. S. 2003. "Inference and computation with population codes." Annual review of neuroscience 26(1): 381-410.

Shea, Nicholas 2007. "Consumers Need Information: supplementing teleosemantics with an input condition." Philosophy and Phenomenological Research 75:404-35.

Shea, Nicholas 2013. "Naturalising Representational Content." Philosophy Compass 8(5): 496-509.

Skyrms, Brian 2010. *Signals: Evolution, Learning, & Information*. Oxford: Oxford University Press.

Fig. 1. From Pouget et al. (2003). The noisy distribution of population activity in the top panel (green, blue and red open circles) is transformed by a Bayesian decoder into the smooth posterior probability distributions in the lower panel (green, blue and red curves, respectively).
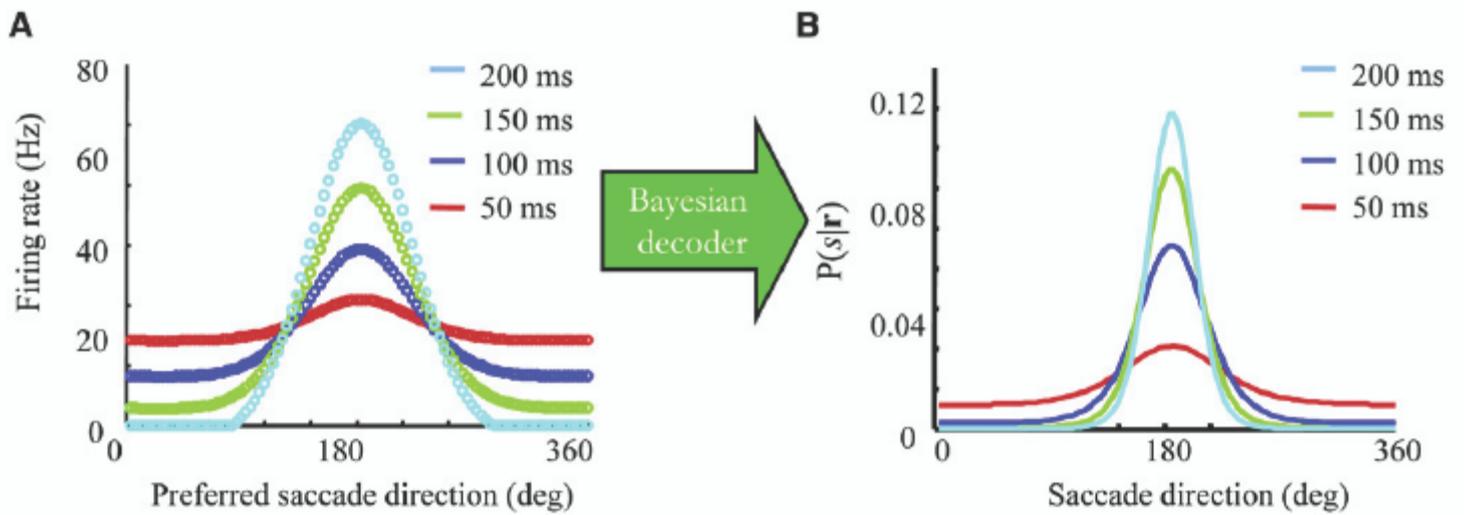
Fig. 2. From Beck et al. (2008). The firing rates of model neurons in panel A are transformed by a Bayesian decoder into the probability distributions in panel B. At each time the width of the probability distribution is narrower than the range of preferred directions which show non-zero firing rates. (Additionally, the variance of the distribution decreases over time as increasing evidence accumulates.)