

# Two kinds of information processing in cognition

Mark Sprevak  
*University of Edinburgh*

18 May 2019

What is the relationship between information and representation? Dating back at least to Dretske (1981), an influential answer has been that information is a rung on a ladder that gets one to representation. Representation is information, or representation is information plus some other ingredient. In this paper, I argue that this approach oversimplifies the relationship between information and representation. If one takes current probabilistic models of cognition seriously, information is connected to representation in a new way. It enters as a property of the represented content as well as a property of the vehicles that carry that content. This offers a new, conceptually and logically distinct way in which information and representation are intertwined in cognition.

## 1 Introduction

There is a new way in which cognition could be information processing. Philosophers have traditionally tended to understand cognition's relationship to Shannon information in just one way. This suited an approach that treated cognition as an inference over representations of single outcomes (*there is a face here, there is a line there, there is a house here*). Recent work conceives of cognition differently. Cognition does not involve an inference over representations of single outcomes but an inference over *probabilistic representations* – representations whose content includes multiple outcomes along with their estimated probabilities.

My claim in this paper is that recent probabilistic models of cognition open up new conceptual and empirical territory for saying that cognition is information

processing. Empirical work is already exploring this territory and researchers are drawing tentative connections between the two kinds of Shannon information in the brain. In this paper, my goal is not to propose a specific relationship between these two quantities of information, although some possible connections are sketched in Section 6. My goal is to convince you that there *are* two conceptually and logically distinct kinds of Shannon information whose relationship should be studied.

Before we proceed, some assumptions. My focus in this paper is only on Shannon information and its mathematical cognates. I do not consider other ways in which the brain could be said to process information.<sup>1</sup> Second, I assume a representationalist theory of cognition. I take this to mean that cognitive scientists find it *useful* to describe at least some aspects of cognition as involving representations. I focus on the role of Shannon information within two different kinds of representationalist model: ‘categorical’ models and ‘probabilistic’ models. My claim is that *if* one accepts a probabilistic model of cognition, then there are two ways in which cognition involves Shannon information. I do not attempt to defend representationalist theories of cognition in general.<sup>2</sup>

Here is a preview of my argument. Under probabilistic models of cognition there are two kinds of probability distribution associated with cognition. First, there is the ‘traditional’ kind: probability distributions associated with a specific neural state occurring in conjunction with an environmental state (for example, the probability of a specific neural state occurring when a subject is presented with a line at 45 degrees in a certain portion of her visual field). Second, there is the new kind, characteristic of probabilistic neural representation: probability distributions that are *represented by* neural states. These probability distributions are the brain’s guesses about the possible environmental outcomes (say, that the line is at 0, 35, 45, or 90 degrees).<sup>3</sup> The two kinds of probability distribution – one associated with *a neural/environmental state occurring* and the other associated with *the neural system’s estimate of a certain environmental state occurring* – are conceptually and logically distinct. They have different outcomes, different probability values, and different types of probability (objective and subjective) associated with them. They generate two separate measures of Shannon information in the brain. The algorithms that

---

<sup>1</sup>See Floridi (2011).

<sup>2</sup>Note that I define representationalist theories in terms of their *utility* for describing cognitive processes, not in terms of their *truth*. Some deny truth but accept utility: they endorse some form of instrumentalism about representationalist models in cognitive science (for example, Egan, 2010; Colombo and Seriès, 2012; Sprevak, 2013). On my view, this still falls within the representationalist paradigm. To the extent that it is legitimate, even if only on pragmatic grounds, to use a representationalist model of cognition, it is legitimate to say that cognition involves two kinds of information processing.

<sup>3</sup>This distinction is not that between ‘encoding’ and ‘decoding’ probability distributions (Elia-smith, 2005a). Encoding and decoding distributions are discussed in Section 3.

underlie cognition can be described as processing *either* or *both* of these Shannon quantities.

## 2 Shannon information

Before attributing two kinds of Shannon information to the brain, we first need to know what justifies attributing *any* kind of Shannon information. Below, I briefly review the definitions of Shannon information in order to identify sufficient conditions for a physical system to be ascribed Shannon information. The rest of the paper shows that these conditions are satisfied in two separate ways. Definitions in this section are taken from MacKay (2003), although similar points can be made with other formalisms.

In order to define Shannon information, one first needs to define the notion of a *probabilistic ensemble*:

- Probabilistic ensemble  $X$  is a triple  $(x, A_X, P_X)$ , where the outcome  $x$  is the value of a random variable, which takes on one of a set of possible values,  $A_X = \{a_1, a_2, \dots, a_i, \dots, a_I\}$ , having probabilities  $P_X = \{p_1, p_2, \dots, p_I\}$ , with  $P(x = a_i) = p_i$ ,  $p_i \geq 0$ , and  $\sum_{a_i \in A_X} P(x = a_i) = 1$

A sufficient condition for the existence of a probabilistic ensemble is the existence of a random variable with *multiple possible outcomes* and an associated *probability distribution*.<sup>4</sup> If the random variable has a finite number of outcomes, this probability distribution takes the form of a mass function, assigning a value,  $p_i$ , to each possible outcome. If the random variable has an infinite number of outcomes, the probability distribution takes the form of a density function, assigning a value,  $p_i$ , to the outcome falling within a certain range. In either case, multiple possible outcomes and a probability distribution over those outcomes is sufficient to define a probabilistic ensemble.<sup>5</sup>

If a physical system has multiple possible outcomes and a probability distribution associated with those outcomes, then that physical system can be treated as a probabilistic ensemble. If a neuron has multiple possible outcomes (e.g. firing or not), and a probability distribution over those outcomes (reflecting the chances of it firing), then the neuron can be treated as a probabilistic ensemble.

Shannon information is a scalar quantity measured in bits. It is predicated of at least three different types of entity: *ensembles*, *outcomes*, and *ordered pairs of ensembles*.

<sup>4</sup>The term ‘outcome’ here is not meant to imply that this is the output of a causal process.

<sup>5</sup>In principle, an ensemble might have only one outcome (necessarily, with probability 1). As we will see, this corresponds to an ensemble and its single outcome having 0 bits of Shannon information.

The definitions differ, so let us consider each in turn.

The Shannon information,  $H(X)$ , of an *ensemble* is defined as:

$$H(X) = \sum_i p_i \log_2 \frac{1}{p_i}$$

The only independent variables in the definition of  $H(X)$  are the possible outcomes of the ensemble (the *is*) and their probabilities (the  $p_i$ s). The Shannon information of an ensemble is a mathematical function of, and only of, these features. Therefore, merely being an ensemble in the sense defined above – having multiple possible outcomes and a probability distribution over those outcomes – is enough to define a  $H(X)$  measure and bestow a quantity of Shannon information. Any physical system that is treated as a probabilistic ensemble *ipso facto* has an associated measure of Shannon information. If a neuron is treated as an ensemble (because it has multiple possible outcomes and a probability distribution over those outcomes), then it automatically has a quantity of Shannon information attached.

The Shannon information,  $h(x)$ , of an *outcome* is defined as:

$$h(x = a_i) = \log_2 \frac{1}{p_i}$$

$H(X)$  is the expected value of  $h(x)$  taken across all possible outcomes of ensemble  $X$ . The only independent variable in  $h(x)$  is the probability of the outcome,  $p_i$ . This means that, again, the existence of an ensemble is a sufficient condition for satisfying the definition of  $h(x)$ . If an ensemble exists, each of its outcomes has a probability and *ipso facto* has a measure of Shannon information. No further conditions need to be met. If a neuron is treated as an ensemble, each of its outcomes (e.g. firing or not firing) has an associated probability, and hence each has a quantity of Shannon information attached.

There are many Shannon measures of information defined for *ordered pairs of ensembles*.<sup>6</sup> Common ones include:

*Joint information:*

$$H(X, Y) = \sum_{x, y \in A_X A_Y} P(x, y) \log_2 \frac{1}{P(x, y)}$$

*Conditional information:*

---

<sup>6</sup>One member of the pair is usually called the ‘sender’ and the other the ‘receiver’.

$$H(X | Y) = \sum_{y \in A_Y} P(y) \sum_{x \in A_X} P(x | y) \log_2 \frac{1}{P(x | y)}$$

*Mutual information:*

$$I(X; Y) = H(X) - H(X | Y)$$

These measures differ from each other in important ways, but again, a sufficient condition for satisfying any one of them is that a physical system has multiple possible outcomes and a probability distribution over their respective outcomes. Two ensembles,  $X$  and  $Y$ , have individual outcomes and probability distributions over those outcomes. The Shannon measures above assume that there is also a joint probability distribution,  $P(X, Y)$ , which describes the probability of any given pair of outcomes from the two ensembles occurring.<sup>7</sup> If ensembles  $X$  and  $Y$  exist, and if pairs of their respective outcomes have probabilities (even if some are 0), then the Shannon measures of joint information, conditional information, and mutual information are defined. Consequently, if two neurons are treated as two ensembles, and if there is a joint probability distribution over pairs of their respective outcomes, then those neurons have associated measures of joint information, conditional information, and mutual information.

A sufficient condition for a physical system to be ascribed Shannon information is that it *has multiple possible outcomes and a probability distribution over those outcomes* (or pairs of outcomes). The Shannon information of an ensemble, a single outcome, or a pair of ensembles is a function of, and only of, the possible outcomes and probability distribution associated with that ensemble, single outcome, or pair. If a physical system is treated as an ensemble (or a pair whose joint outcomes have probabilities), it *ipso facto* has Shannon information.

If a physical system *changes* the probabilities associated with its possible outcomes over time, its associated Shannon measures are likely to change too. Such a system may be described as ‘processing’ Shannon information. This change could happen in at least two ways. If a physical system modifies the probabilities associated with its *physical states* occurring (e.g. a neuron makes certain physical states such as

---

<sup>7</sup> $P(X, Y)$  defines conditional probability measures, such as  $P(X | Y)$ .

firing more or less likely), it can be described as processing Shannon information.<sup>8</sup> Alternatively, if the firing of the neuron *represents* a probability distribution over possible outcomes, and that represented probability distribution changes over time – perhaps as a result of learning or inference – then that neuron’s associated Shannon measures will change too. In both cases, probability distributions and Shannon information change. But distinct probability distributions and distinct measures of Shannon information change in each case. The remainder of this paper will unpack the distinction between the two.

### 3 The traditional kind of Shannon information

Traditionally, Shannon information has been used as a building block when *naturalising representation*. Many versions of information-theoretic semantics try to explain semantic content in terms of Shannon information. These accounts aim to explain how representation arises from Shannon information. Such theories often claim that Shannon information is a source of naturalistic, objective facts about representational content. Dretske formulated one of the earliest such theories.<sup>9</sup> Dretske’s (1981) theory aimed to entirely reduce facts about representational content to facts about Shannon information. More recently, other accounts – including Dretske’s later (1988; 1995) views – have proposed that an information-theoretic condition is only one part of a larger naturalistic condition on representational content. Additional conditions include variously conditions on teleology, instrumental (reward-guided) learning, structural isomorphism, and/or appropriate use.<sup>10</sup> In what follows, I will focus solely on the information-theoretic part of such a semantic theory.

Information-theoretic semantics attempts to explain representation in terms of one physical state ‘carrying information’ about another physical state. The relationship of ‘carrying information’ is assumed to be a precursor to, or a precondition for, certain varieties of representation. In the context of the brain, such a theory says:

---

<sup>8</sup>One way in which this could occur is during learning and other kinds of long-term plasticity. However, similar changes also occur during short-term processes. When a neuron fires, it makes a specific outcome – *firing* – certain. That will affect the probabilities associated with other neurons in the brain (making their respective outcomes of firing more or less probable), and hence change their associated Shannon measures. Neuroscientists can track how these Shannon measures change as a specific outcome propagates in the brain during cognition. Thanks to Nick Shea for this point.

<sup>9</sup>Prior to Dretske’s work, Shannon information had been linked to semantic content, although not always in reductive fashion (Bar-Hillel and Carnap, 1964; Wiener, 1961).

<sup>10</sup>See Millikan (1984); Papineau (1987); Dretske (1988); Shea (2007); Shea (2014a); Skyrms (2010); W. M. Ramsey (2016) for a range of such proposals. Note that some of these authors argue that mental representations sometimes gain their content solely on the basis of non-Shannon factors. Thanks to an anonymous referee for pointing this out.

(R) Neural state,  $n$  (from  $N$ ), represents an environmental state,  $s$  (from  $S$ ), only if  $n$  ‘carries information’ about  $s$ .

Implicit in R is the idea that neural state,  $n$ , and environmental state,  $s$ , come from a set of possible alternatives. According to R, neural state  $n$  represents  $s$  only if  $n$  bears the ‘carrying information’ relation to  $s$  and not to other outcomes. Different neural states could occur in the brain (e.g. different neurons in a population might fire). Different environmental states could occur (e.g. a face or a house could be present). Crudely, the reason why certain neural firings represent a face and not a house is that those firings, and only those firings, bear the ‘carrying information’ relationship to *face* outcomes; they do not bear this relationship to *house* outcomes. R implicitly assumes that we are dealing with multiple possible outcomes: multiple possible representational vehicles ( $N$ ) and multiple possible environmental states ( $S$ ). It names a special relationship between individual outcomes that is necessary for representation. Representation occurs only when  $n$  from  $N$  bears the ‘carrying information’ relation to  $s$  from  $S$ .

The primary task for an information-theoretic semantics is to explain what this carrying information relation is. Different versions of information-theoretic semantics do this differently.<sup>11</sup> Theories can be divided into roughly two camps: those that are ‘correlational’ and those that invoke ‘mutual information’.

The starting point of ‘correlational’ theories is that one physical state ‘carries information’ about another just in case there is a statistical correlation between the two that satisfies some probabilistic condition. This still leaves plenty of questions unanswered: What kind of correlation (Pearson, Spearman, Kendall, mutual information, or something else)?<sup>12</sup> How should physical states be typed so that a correlation can be measured? How much correlation is enough for information carrying? Does it matter if the correlation is accidental or underwritten by a law or disposition?

Rival information-theoretic semantics take different views. Consider the following three proposals:

1.  $P(S = s \mid N = n) = 1$

---

<sup>11</sup>The relation of ‘carrying information’ is also sometimes described as one physical state ‘having natural information’ about another (see Stegmann, 2015).

<sup>12</sup>Millikan (2001) suggests that one should look at the probabilistic relations that are ‘learnable’ for an agent: A is correlated with B, and hence carries information, if B is learnable (or inferable) from A. However, any degree of probabilistic dependence between A and B (no matter how slight) could, in principle, allow an agent to learn, or infer, one from the other. With suitable rewards, even the mildest degree of probabilistic dependence *could* be a target of learning as an agent could be arbitrarily incentivised to do so. The notion of a ‘learnable’ relation – if it is not merely a synonym for *not probabilistically independent* – is as much in need of explication as the notion of ‘correlation’.



2.  $P(S = s \mid N = n)$  is ‘high’
3.  $P(S = s \mid N = n) > P(S = s \mid N \neq n)$

Dretske (1981) endorses (1): a neural state carries information about an environmental state just in case an agent, given the neural state, could infer with certainty that the environmental state occurs (and this could not have been inferred using the agent’s background knowledge alone). Millikan (2000; 2004) endorses (2): the conditional probability of the environmental state, given the neural state, need only be ‘high’, where what counts as ‘high’ is a complex matter involving the correlation having influenced past agential use via genetic selection or learning.<sup>13</sup> Shea (2007) and Scarantino and Piccinini (2010) propose that the correlation should be understood in terms of probability raising, (3): the neural state should make the occurrence of the environmental state more probable than it would have been otherwise.

At first glance, there may seem nothing particularly Shannon-like about proposals (1)–(3). Probability theory alone is sufficient to express the relevant condition on representation. These theories are perhaps better described as ‘probabilistic’ semantics than ‘information-theoretic’ semantics.<sup>14</sup> Nevertheless, there is a legitimate way in which these accounts do entail that cognition is Shannon information processing. According to (1)–(3), ‘carrying information’ is a relationship between particular outcomes and those outcomes must come from ensembles that have probability distributions. Remember that a sufficient condition for a system to have Shannon information is that it *has multiple possible outcomes and a probability distribution over those outcomes*. (1)–(3) assure us that this is true of a cognitive system that contains representations. According to (1)–(3), the representational content of a neural state arises when that state is an outcome from an ensemble with other possible outcomes (other possible neural states) that could occur with certain probabilities (and probabilities conditional on various possible environmental outcomes). If cognition involves representation, and those representations gain their content by any of (1)–(3), then cognition *ipso facto* involves Shannon information. Shannon information attaches to representations because of the probabilistic nature of their vehicles. According to (1)–(3), that probabilistic nature is essential to their representational status. Therefore, to the extent that cognition can be described as processing representations, and to the extent that we accept one of these versions of information-theoretic semantics, cognition can be described as processing states with a probabilistic nature, and so, processing states with Shannon information.

‘Mutual information’ versions of information-theoretic semantics unpack ‘carrying

<sup>13</sup>See Stegmann (2015), pp. 873–874 for helpful analysis of Millikan’s view.

<sup>14</sup>Timpson (2013), pp. 41–42 makes a similar point with regard to Dretske’s (1981) theory, and related criticisms are raised by commentators for Dretske (1983).



information’ differently. They invoke the Shannon concept of mutual information – or, rather, pointwise mutual information, the analogue of mutual information for pairs of single outcomes. The familiar notion of mutual information  $I(X; Y)$  is the expected value of pointwise mutual information  $PMI(x, y)$  across all outcomes from a pair of ensembles.<sup>15</sup> Pointwise mutual information for a pair of single outcomes,  $x, y$ , is defined as:

$$PMI(x, y) = \log_2 \frac{P(x, y)}{P(x)P(y)} = \log_2 \frac{P(x | y)}{P(x)} = \log_2 \frac{P(y | x)}{P(y)}$$

Skyrms (2010) and Isaac (2019) propose that the information carried by a physical state,  $n$ , (its ‘informational content’), is a vector consisting of the  $PMI(n, s)$  value for every possible environmental state,  $s_i$ , from  $S$ , given that  $n$  from  $N$ :  $\langle PMI(n, s_1), \dots, PMI(n, s_n) \rangle$ . Isaac identifies the meaning or representational content of  $n$  with this  $PMI$ -vector. Skyrms says that the meaning or content is likely to be a more traditional semantic object, such as a set of possible worlds – this set is derived from the  $PMI$ -vector by considering the environmental states that generate high-value elements in the vector; the representational content is the set of possible worlds in which high  $PMI$ -value environmental states occur.

Like Skyrms and Isaac, Usher (2001) and Eliasmith (2005b) appeal to pointwise mutual information to define ‘carrying information’. Unlike Skyrms and Isaac, they define it as a relation that holds between a single neural state,  $n$ , and a single environmental state,  $s$ . They say that  $n$  carries information about  $s$  just in case  $s$  is the environmental state for which  $PMI(n, s)$  has its maximum value given neural state  $n$ . Neural state  $n$  carries information about the  $s$  that produces the peak-value element in its  $PMI$ -vector. Usher and Eliasmith connect this to what is measured in ‘encoding’ experiments in neuroscience. In an encoding experiment, many environmental states are presented to a brain and researchers look for the environmental state that best predicts a specific neural response – that yields the highest  $PMI(n, s)$  as one varies  $s$  for some fixed  $n$ . Usher and Eliasmith offer a second, conceptually independent definition of ‘carrying information’. This is based around what is measured in ‘decoding’ experiments. In a decoding experiment, researchers examine many neural states and classify them based on which one best predicts an environmental state – i.e. which neural state  $n$  yields the highest  $PMI(n, s)$  for a fixed  $s$ . Here, instead of looking for the maximum  $PMI(n, s)$  value as one varies  $s$  and keeps  $n$  constant, one looks for the maximum  $PMI(n, s)$  value as one varies  $n$  and keeps  $s$  constant. There is no reason why the results of encoding and decoding experiments should coincide: they pick out two different kinds of information-theoretic relationship between the brain and its environment. Usher

---

<sup>15</sup> $I(X; Y) = \sum_{x, y \in A_X, A_Y} P(x, y) PMI(x, y)$

and Eliasmith argue that they provide different, complementary, and equally valid accounts of representational content.

On each of these semantic theories, Shannon information is ascribed to a cognitive system because of the probabilistic properties of neural states *qua vehicles*. It is because a given neural state is an outcome from a set of possible alternative states, combined with the probability of various environmental outcomes, that the cognitive system has the Shannon information properties relevant to representation and hence to cognition. In the next section, I describe a different way in which Shannon information enters into cognition. Here, the relevant information-theoretic quantity arises not from the probabilistic nature of the physical vehicles and environmental states, but from its *representational content*. ‘Probabilistic’ models of cognition claim that the representational content of neural states is probabilistic. This means that Shannon information attaches to a cognitive system in a new way: via its content rather than via the probabilistic occurrence of its neural vehicles.

## 4 The new kind of Shannon information

Probabilistic models of cognition, like the accounts discussed in the previous section, ascribe representations to the brain. Unlike the previous accounts, these models do not aim to naturalise representational content. They help themselves to the existence of representations. Their claim is that these representations have a particular kind of content. They are largely silent about how these representations get this content. In principle, probabilistic models of cognition are compatible with a variety of underlying semantic theories, including versions of information-theoretic semantics.<sup>16</sup>

The central claim of a probabilistic model of cognition is that neural representations have probabilistic representational content. Traditional, ‘categorical’ approaches assume that neural representations have *single outcomes* as their representational content. Under a categorical approach, a neural state, *n*, represents a single environmental outcome (or a single set of outcomes). Thinking about neural representation in these terms has prompted description of neural states early in V1 as *edge detectors*: their activity represents the presence (or absence) of an edge at a particular angle in a portion of the visual field. The represented content is a particular outcome (*edge at ~45 degrees*). Similarly, neurons in the inferior temporal (IT) cortex are described as *hand detectors*: their activity represents the presence (or absence) of a hand. The represented content is a single outcome (*hand present*). Similarly, neurons in the fusiform face area (FFA) are described as *face detectors*: their activity represents the

---

<sup>16</sup>I discuss how information-theoretic semantics might interact with probabilistic models of cognition in Section 6.

presence (or absence) of a face. The represented content is a single outcome (*face present*) (for example, see Gross, 2007; Logothetis and Sheinberg, 1996; Kanwisher, McDermott and Chun, 1997).

There is increasing suspicion that representation in the brain is not like this. Content is rarely categorical (*hand present*); rather, what is represented is a probability distribution over many possible states. The brain represents many outcomes simultaneously in order to ‘hedge its bets’ during processing. This allows the brain to store, and make use of, information about multiple possible outcomes if it is uncertain which is the true one. Uncertainty may come from unreliability in the perceptual hardware, or from the brain’s epistemic situation that even with perfectly functioning hardware it only has incomplete access to its environment.

Ascribing probabilistic representations to a cognitive agent is not a new idea (de Finetti, 1990; F. P. Ramsey, 1990). However, there is an important difference between past approaches and new probabilistic models of cognition. In the past, probabilistic representations were treated as *personal-level* states of a cognitive agent – ‘credences’, ‘degrees of belief’, or ‘personal probabilities’. In the new models, probabilistic representations are treated as states of subpersonal *parts* of the agent – of neural populations, or single neurons. Their novel claim is that, regardless of whichever personal-level states that are attributed to an agent, various parts of that agent token diverse (and perhaps even conflicting) probabilistic representations. Thinking in these terms has prompted redescription of neural states early in V1 as probabilistically nuanced ‘hypotheses’, ‘guesses’, or ‘expectations’ about edges. Their neural activity does not represent a single state (*edge at ~45 degrees*) but a probability distribution over multiple edge orientations (Alink et al., 2010). The represented content is a probability distribution over how the environment stands with respect to edges. Similarly, neural activity in the IT cortex does not represent a single state of affairs (*hand present*) but a probability distribution over multiple possible outcomes regarding hands. The represented content is a probability distribution over how the environment stands with respect to hands. Similarly, neural activity in FFA does not represent a single state of affairs (*face present*) but a probability distribution over multiple possible outcomes regarding faces. The represented content is a probability distribution over how the environment stands with respect to faces (Egner, Monti and Summerfield, 2010).

Traditional models of cognition tend to describe cognitive processing as a computationally structured inference over specific outcomes – *if there is an edge here, then that is an object boundary*. Probabilistic models of cognition in contrast describe cognitive processing as a computationally structured inference over probability distributions – *if the probability distribution of edge orientations is this, then the probability distribution of object boundaries is that*. Cognitive processing is a series

of steps that use one probability distribution to condition, or update, another probability distribution.<sup>17</sup> Neural representations may conceivably maintain a probabilistic character right until the moment that the brain is forced to plump for a specific outcome in action. At that point, the brain may select the most probable outcome from its current represented probability distribution conditioned on all its available evidence (or some other point estimate that is easier to compute).

Modelling cognition as probabilistic inference does not mean modelling cognition as non-deterministic or chancy. The physical hardware and algorithms underlying the probabilistic inference may be entirely deterministic. Consider that when your electronic PC filters spam messages from incoming emails it performs a probabilistic inference, but both the PC's physical hardware and the algorithm that the PC follows are entirely deterministic. A probabilistic inference takes representations of probability distributions as input, yields representations of probability distributions as output, and transforms input to output based on rules of valid (or pragmatically efficacious) probabilistic inference. The physical mechanism and the algorithm for processing representations may be entirely deterministic. What makes the process probabilistic is not the chancy nature of vehicles or rules but that probabilities feature in the represented content that is being manipulated.

Perhaps the best-known example of a probabilistic model of cognition is the 'Bayesian brain' hypothesis. This says that brains process probabilistic representations according to rules of Bayesian or approximately Bayesian inference (Knill and Pouget, 2004). Predictive coding provides one proposal about how such inference could be implemented in the brain (Clark, 2013; Friston, 2009). It is worth stressing that the motivation for ascribing probabilistic representations to the brain, and for probabilistic models of cognition in general, is broader than that for the Bayesian brain hypothesis (or for predictive coding). The brain's inferential rules could, in principle, depart very far from Bayesianism and still produce adaptive behaviour under many circumstances. It remains an open question to what extent humans are Bayesian (or approximately Bayesian) reasoners. Probabilistic techniques developed in AI, such as deep learning, reinforcement learning, and generative adversarial models can produce impressive behavioural results despite having complex and qualified relationships to Bayesian inference. The idea that cognition is a form of probabilistic inference is a more general idea than that cognition is Bayesian. A researcher in cognitive science may subscribe to probabilistic representation in the brain even if they take a dim view of the Bayesian brain hypothesis.<sup>18</sup>

The essential difference between a categorical representation and a probabilistic

---

<sup>17</sup>Conditional probabilities tell the cognitive system how to update its estimate of one variable based on its knowledge about other variables.

<sup>18</sup>See Ma (2012).

one lies in its content. Categorical representations aim to represent a single state of affairs. In Section 3, we saw that schema R treats representation as a relationship between a neural state,  $n$ , and an environmental outcome,  $s$ . Representational content is typically specified by a truth, accuracy, or satisfaction condition. Meeting this condition is assumed to be largely an all-or-nothing matter. A categorical representation effectively ‘bets all its money’ that a certain outcome occurs. An edge detector represents *there is an edge*. Multiple states of affairs may sometimes feature in the representational content (for example, *there is an edge between ~43–47 degrees*), but those states of affairs are grouped together into a single outcome that is represented as true. There is no probabilistic nuance, or apportioning of different degrees of belief, to different outcomes.

In contrast, probabilistic representations aim to represent a probability distribution over multiple outcomes. The probability distribution is a measure of how much the system ‘expects’ that the relevant outcomes are true. Unlike categorical representations, the represented content does not partition the possible environmental states into only two classes (*true* and *false*). Representation is not an all-or-nothing matter but involves assigning a probability weight to various possible outcomes. As we will see in the next section, these outcomes need not coincide with the possible outcomes of  $S$ . Whereas categorical representational content is typically specified by a truth, accuracy, or satisfaction condition, probabilistic representational content is typically specified by a probability mass or density function over a set of possible outcomes.

In principle, probabilistic representations could use any physical vehicle, and any formal format. There is nothing about the physical make-up of a representational vehicle that determines whether it is categorical or probabilistic. Either type of representation could also, in principle, use any number of different formal formats to organise its structure and guide the algorithms that operate on it. Possible formal formats for a representation include being a setting of weights in a neural network, a symbolic expression, a directed graph, a ring, a tree, a region in continuous space, or an entry in a relational database (Griffiths, Chater et al., 2010; Tenenbaum et al., 2011). The choice of physical vehicle and representational format affects how easy it is to implement an inference with computation in a specific physical context (Marr, 1982). Certain physical vehicles and certain formal formats are more apt to serve certain computations than others. But in principle, there is nothing about the physical make-up or formal structure of a representation that determines whether the representation is categorical or probabilistic. That is determined solely by its represented content.

The preceding discussion should not be taken as suggesting that a model of cognition must employ only one type of representation (categorical or probabilistic). There is

no reason why both types of representation cannot appear in a model of cognition, assuming there are appropriate rules to take the cognitive system between the two. Neither does the discussion suggest that one type of representation cannot be reduced to the other. A variety of such reductions may be possible. For example, a cognitive system might use structured complexes of traditional representations to express the probability calculus and thereby express probabilistically nuanced content with categorical representations (maybe this is what we do with the public language of mathematical probability theory). Alternatively, a cognitive system might use structured complexes of probabilistic representations to represent all-or-nothing-like truth conditions. Feldman (2012) describes a proposal in which categorical representations are approximated by probabilistic ones with strongly modal (sharply peaked) probability distributions.<sup>19</sup> Categorical and probabilistic representations may mix in cognition, and perhaps, given the right conditions, one may give rise to the other.<sup>20</sup>

## 5 Two kinds of information processing

In Section 1, we assumed that cognition is profitably described by saying it involves representations. In Section 2, we saw that having multiple outcomes and a probability distribution over those outcomes is sufficient to have an associated measure of Shannon information. We have now seen, in Sections 3 and 4, two ways in which the representations involved in cognition can have multiple outcomes and probability distributions associated with them. Consequently, Shannon information may attach to cognition in two separate ways. What characterises the Shannon information of Section 3 is that it is associated with probability of the *vehicle* occurring (conditional on various environmental outcomes). What characterises the Shannon information of Section 4 is that it is associated with the probabilities that appear inside the represented *content*.

The degree to which these two quantities of Shannon information differ depends on the degree to which the two underlying sets of outcomes and probability distributions differ. In this section, I argue that they typically involve different *sets of outcomes*, different numerical *probability values*, and they must involve different *kinds of probability*.

*Different sets of outcomes.* In Section 3, the relevant set is the set of *possible neural and environmental states*. The outcomes are the objective possibilities – neural

---

<sup>19</sup>Feldman calls these ‘symbolic representations’, but his claim is about their content, not about the representational format of their vehicles.

<sup>20</sup>Also see Rahnev (2017) for models of cognition that are ‘intermediate’ between categorical and probabilistic representation.



and environmental – that could occur. What interests Dretske, Millikan, Shea, Skyrms, and others is to know whether a particular neural state from a set of alternatives ( $N$ ) occurs conditional on a particular environmental state from a set of alternatives ( $S$ ).<sup>21</sup> In contrast, in Section 4, the relevant outcomes are the *represented possible states of the environment*. These are the ways that the brain represents the environment could be. This set of represented environmental possibilities need not be the same as what is objectively possible. A cognitive system might make a mistake about what is possible just as it might make a mistake about what is actual: it might represent an environmental outcome that is impossible (e.g. winning a lottery that the agent never entered) or it might fail to represent an environmental state that is possible (e.g. that it is a brain in a vat). Unless the cognitive system represents all and only the objectively possible outcomes, there is no reason to think that its set of represented outcomes will be the same as the set of possible outcomes in Section 3. Hence, the set of outcomes represented by a neural state need not be the same as the set of outcomes  $S$ . Moreover, for the two sets of outcomes over which probabilities are ascribed to be the same, the brain would need to represent not only the possible environmental states ( $S$ ) but also its possible neural states ( $N$ ). Only in the special case of a cognitive system that (a) represents all and only the objectively possible environmental states and (b) represents all and only its own possible neural states would the respective sets of outcomes which are assigned probabilities coincide.

*Different probability values.* Suppose that a cognitive system, perhaps due to some design quirk, does represent all and only the objectively possible environmental and neural states. In such a case, the numerical probability values associated with the outcomes are still likely to differ. In the context of the projects of Section 3, these probability values measure the objective chances, frequencies, propensities, or some similar measure of a neural state occurring conditional on a possible environmental state. What interests Millikan, Shea, and others are these objective probabilistic relations between neural states and environmental states. In contrast, for the projects of Section 4, the probability values are the cognitive system's *estimation* of how likely each outcome is, not its objective probability. Brains are described as having 'priors' – probabilistic representations of various outcomes – and a 'likelihood function' or 'probabilistic generative model' – a probabilistic representation of the relationships between the outcomes. Psychologists are interested in how the brain uses its priors and generative model to make inferences about unknown events, or in how it updates its priors in light of new evidence. All the aforementioned probabilities are the brain's guesses about the possible outcomes and the relationships between them. Only a God-like cognitive agent, one who knew the truth about the objective

---

<sup>21</sup>Or whether an environmental state occurs conditional on some neural state occurring. Each can be exchanged for the other via Bayes' theorem.



probabilities of events and their relations, would assign the right probability values to the various outcomes and relations. Such a system would have a *veridical* (and a *complete*) probabilistic representation of its environment, its own neural states, and the relationships between them. This may be a goal to which a cognitive system aspires, but it is surely a position that few achieve.

*Different kinds of probability.* Assume for the sake of argument that we are dealing with a God-like cognitive agent who has a complete and veridical probabilistic representation of its environment and its neural states. Even for that agent, there are still two distinct types of Shannon information. This is because its respective probability values, even if they agree numerically, measure different *kinds* of probability. The  $P(\cdot)$ s measure something different in each case. In the context of the projects of Section 3, the  $P(\cdot)$  values measure objective probabilities. These may be chances, frequencies, propensities, or whatever else corresponds to the objective probability of the relevant outcome occurring.<sup>22</sup> In the context of the projects of Section 4, the  $P(\cdot)$  values measure subjective probabilities. These are the system's estimation of how likely it thinks the relevant outcomes are. Chances, frequencies, propensities, or similar are not the same as a system's representation of how likely an event is to occur. Even for a God-like cognitive agent – for whom the two are stipulated as equal in terms of numerical value – what is measured is distinct. Subjective probabilities, even if they agree in terms of numerical value with objective probabilities, do not become objective probabilities merely because they happen to accurately reflect them. No more than a description of a Komodo dragon becomes a living, breathing Komodo dragon if that description happens to be accurate. One is a representation, the other is a state of the world. In the case of our God-like agent, one is a distribution of objective probabilities and the other is the system's (veridical) representation of possible outcomes and their respective credences. Well-known normative principles connect subjective and objective probabilities. However, no matter which normative principles one endorses, and regardless of whether a God-like agent satisfies them, the two kinds of probability are distinct.<sup>23</sup>

Two kinds of probability distribution feature in cognition. Each generates an associated measure of Shannon information. The two Shannon measures are distinct:

<sup>22</sup>Different theorists in Section 4 take different views about the nature of these objective probabilities. Shea (2007) says the probabilities are chances (although he does not say what chances are); Millikan (2000) focuses on the idea that they are frequencies and she considers the consequent reference class problem. No one entertains the hypothesis that they are subjective probabilities.

<sup>23</sup>Skyrms agrees: 'objective and subjective information' may be carried by a neural state (2010, pp. 44–45). Skyrms' concern is with the objective probabilities that are associated with neural states and environmental states. However, he agrees that subjective probabilities (and, hence, subjective information) may be carried by a neural state *qua* content.

they are likely to involve different outcomes, different probability values, and must involve different kinds of probability. This allows us to make sense of two kinds of Shannon information being processed in cognition: two kinds of probability distribution change under probabilistic models of cognition. Processing involves changes in a system's representational vehicles and changes in a system's probabilistic represented content. Information-processing algorithms that govern cognition can be defined over either or both of these Shannon quantities.<sup>24</sup>

## 6 Relationship between the two kinds of information

My claim in the previous section was that the two kinds of Shannon information are distinct. This does not rule out all manner of interesting connections between them. That they are distinct does not mean that they can vary independently of each other. This section highlights some possible connections.

### 6.1 Connections via semantic theory

One is likely to be persuaded of deep connections between the two kinds of Shannon information if one endorses some form of information-theoretic semantics for probabilistic representations. The probabilistic models described in Section 4 are silent about how neural representations get their content. In principle, these models could be combined with a range of semantic proposals, including some version of the information-theoretic semantics described in Section 3.

Skyrms' or Isaac's theory looks the most promising approach to generate an information-theoretic account of probabilistic content. Both their theories already attribute multiple environmental outcomes plus a graded response for each outcome. However, it is not immediately obvious how to proceed. The probability distribution represented by  $n$  cannot simply be assumed to be the probability distribution of  $S$ . As we saw in Section 5, a probabilistic representation may misrepresent the objective possibilities and their probability values. A second consideration is that the represented probabilities appear to depend not only on the probabilistic relations between a representational vehicle and its corresponding environmental outcomes; they also depend on what else the system 'believes'. The probability that a system assigns to *there is a face* should not be independent of the probability

---

<sup>24</sup>One might object that there are not two *kinds* of Shannon information, but only two applications of a single mathematical kind of Shannon information to the brain. However, the same could be said of objective and subjective probabilities: both are applications of a single kind of mathematical probability (to measure objective chances and agents' uncertainties). To the extent that one is willing to say that there are two 'kinds' of probability (objective and subjective), one should be willing to say the same for Shannon information.

that it assigns to *there is a person*, even if the two outcomes are represented by different neural vehicles. A noteworthy feature of the information-theoretic accounts of Section 3 is that they disregard relationships of probabilistic coherence between representations in assigning representational content. They assign content piecemeal, without considering how the contents may cohere. How to address these two issues and create an information-theoretic semantics for probabilistic representations is presently unclear.<sup>25</sup>

If an information-theoretic semantics for probabilistic neural representations could be developed, it would provide a bridge between the two kinds of Shannon information. One kind of information (associated with the represented probabilities) could not vary independently of the other (associated with the objective probabilities). The two would correlate at least in the cases to which this semantic theory applied. Moreover, if the semantic theory held as a matter of conceptual or logical truth, then the connection between the two Shannon quantities would hold with a similar strength. An information-theoretic account of probabilistic representation offers the prospect of a conceptual or logical connection between the two types of Shannon information. In the absence of such a semantic theory, however, it is hard to speculate on exactly what the nature of that connection would be.

If one is sceptical about the prospects of an information-theoretic semantics for probabilistic neural representation, then one may be less inclined to see deep conceptual or logical connections between the two kinds of Shannon information. If one endorses Grice's (1957) theory of *non-natural* meaning, for example, then the two Shannon quantities may look conceptually and logically independent. Grice said that in cases of non-natural meaning, representational content depends on human intentions and not, for example, on the objective probabilities of a physical vehicle occurring in conjunction with environmental outcomes. There is nothing to stop a physical vehicle representing any content, provided it is underwritten by the right intentions. I might say that the proximity of Saturn to the Sun (appropriately normalised) represents the probability that Donald Trump will be impeached. Provided this is underwritten by the right intentions, probabilistic representation occurs. Representation is, in this sense, an arbitrary connection between a vehicle and a content that can be set up or destroyed at will, without regard for the probabilities of the underlying events.<sup>26</sup> If one endorses Grice's theory of non-natural meaning, there need be no connection between the probabilities of neural and environmental states and what those states represent, and one Shannon measure could vary independently of the other. This is not to say that the two measures

---

<sup>25</sup>See Shea (2014b); Shea (2018) for a promising approach.

<sup>26</sup>Skyrms (2010) argues against this that 'all meaning is natural meaning' (p. 1). All meaning depends on the physical probabilities that connect vehicles and their content.

would not correlate in the brain; just that, if they correlate, that would not flow from the semantic theory.

## 6.2 Connections via empirical correlations

Regardless of connections that may arise from one's semantic theory, there are likely to be other reasons why the two measures of Shannon information would correlate in the brain. The nature of these connections will depend on the strategy that the brain uses to 'code' its probabilistic content. This coding scheme describes how probabilistic content – which may consist of probability values, the overall analytical shape of the probability distribution, or summary statistics like the mean or variance – maps onto physical activity in the brain or onto physical relations between the brain and environment. The specific scheme that the brain uses to code its probabilistic content is currently unknown and the subject of much speculation. Suggested proposals include that the firing rate of a neuron, the number of neurons firing in a population, the chance of neurons firing in population, or the spatial distribution of neurons firing in a population is a monotonic function of characteristic features of the represented probability distribution (see, for example, Barlow, 1969; Averbeck, Latham and Pouget, 2006; Ma et al., 2006; Deneve, 2008; Fiser et al., 2010; Griffiths, Vul and Sanborn, 2012; Pouget et al., 2013). According to these schemes, the probability of various neural physical states occurring varies in some regular way with their represented probability distributions. This relationship may be straightforward and simple or it may be extremely complicated and vary in different parts of the brain or over time. The same applies to the relationship between the two Shannon quantities. If an experimentalist were to know the brain's coding scheme, she may be able to infer one Shannon measure from the other. But even granted this were possible, the two kinds of Shannon information would remain distinct, for the reasons given in Section 5.

Cognitive processing is sometimes defined over the information-theoretic properties of the neural vehicles. Saxe, Calderone and Morale (2018) describe how brain entropy during resting state, as measured by fMRI, correlates with general intelligence. Chang et al. (2018) describe how drinking coffee increases the brain's entropy during resting state. Carhart-Harris et al. (2014) describe the relationship between consciousness and brain entropy, and how this changes after taking the psychedelic drug psilocybin. Rieke et al. (1999) advocate a research programme that examines information-theoretic properties of neural vehicles (spike trains) and their relationships to possible environmental outcomes. They argue that information-theoretic properties of the neural vehicles and environmental outcomes allow us to infer possible and likely computations that the brain uses and the efficiency of the brain's coding scheme. In each of these cases, the Shannon measures are defined over

the possible neural vehicles and environmental states, not over their represented content (although several of the authors suggest that since the two are correlated by the brain's coding scheme, we can use one to draw conclusions about the other).

In contrast, Feldman (2000) looks at algorithms defined over the information-theoretic properties of the represented content. He argues that the difficulty of learning a new Boolean concept correlates with the information-theoretic complexity of the represented Boolean condition. Kemp (2012) and Piantadosi, Tenenbaum and Goodman (2016) extend this idea to general concept learning. They propose that concept learning is a form of probabilistic inference that seeks to find the concept that maximises the probability of the represented classification. This cognitive process is described as the agent seeking the concept that offers the optimal Shannon compression scheme over its perceptual data. Gallistel and Wilkes (2016) describe associative learning as a probabilistic inference concerning the most likely causes of an unconditioned stimulus given the observations. They describe it in terms of Shannon information processing: the cognitive system starts with priors over hypotheses about causes that have maximum entropy (their probability distributions are as 'noisy' as possible consistent with the data); the cognitive system then aims to find the hypotheses that provide optimal compression (that maximise Shannon information) of the represented hypothesis and observed data. In general, researchers who model cognition probabilistically move smoothly between probabilistic formulations and information-theoretic formulations when describing a cognitive process. In each of the cases described above, the Shannon information is associated not with the probabilities of specific neural vehicles occurring, but with the probability distributions that they represent (although, again, one might think that the two are likely to be related via the brain's coding scheme).

### 6.3 Two versions of the free-energy principle

Friston (2010) claims that the 'free-energy principle' provides a unified theory of how cognitive and living creatures work. He invokes two kinds of Shannon information processing and he effectively describes two separate versions of the free-energy principle.

First, Friston says that the free-energy principle is a claim about the probabilistic inference performed by a cognitive system. He claims that the brain aims to predict upcoming sensory activation and it forms probabilistic hypotheses about the world that are updated in light of its errors in making this prediction. Shannon information attaches to the represented probability distributions over which the inference is performed. Friston says that the brain aims to minimise the 'surprisal' of – the Shannon information associated with – new sensory evidence. When the brain is engaged in probabilistic inference, however, he says that it does not represent

the full posterior probability distributions as a perfect Bayesian reasoner would do. Instead, the brain approximates them with simpler probability distributions, assumed to be Gaussian. Provided the brain minimises the Shannon-information quantity ‘variational free energy’, it will bring these simpler probability distributions into approximate correspondence with the true posterior distributions that a perfect Bayesian reasoner would have (Friston, 2009; Friston, 2010). Variational free energy is an information-theoretic quantity, predicated of the agent’s represented probability distributions, that measures how far those subjective probability distributions depart from the optimal guesses of a perfect Bayesian reasoner. According to Friston, the brain minimises ‘free energy’ and so approximates an ideal Bayesian reasoner.

Friston makes a second, conceptually distinct, claim about cognition (and life in general) aiming to minimise free energy. In this context, his goal is to explain how cognitive (and living) systems maintain their physical integrity and homoeostatic balance in the face of a changing physical environment. Cognitive (and living) systems face the problem that their physical entropy tends to increase over time: they generally become more disordered and the chance increases that they will undergo a fatal physical phase transition. Friston says that when living creatures resist this tendency, they minimise free energy (Friston, 2013; Friston and Stephan, 2007). However, the free energy minimised is not the same as that which attaches to the represented probabilistic guesses of some agent. Instead, it attaches to the objective probabilities of various possible (fatal) physical states of the agent occurring in response to environmental changes. Minimising free energy involves the system trying to arrange its internal physical states so as to avoid being overly changed by probable environmental transitions. The system strives to maintain its physical nature in equipoise with likely environmental changes. The information-theoretic free energy minimised here is defined over the objective distributions of possible physical states that could occur, not over the probability distributions represented by an agent’s hypotheses.

Minimising one free-energy measure may help an agent to minimise the other: a good Bayesian reasoner is plausibly more likely to survive in a changing physical environment than an irrational agent. But they are not the same quantity. Moreover, any correlation between them could conceivably come unstuck. An irrational agent could depart far from Bayesian ideals but be lucky enough to live in an hospitable environment that maintains its physical integrity and homoeostasis no matter how badly the agent updates its beliefs. Alternatively, an agent might be a perfectly rational Bayesian and update its beliefs accordingly, but its physical environment may change so rapidly and catastrophically that it cannot survive or maintain homoeostasis. Understanding how Friston’s two formulations of the free-energy principle interact – that pertaining to represented subjective probabilities and that

pertaining to objective probabilities – is ongoing work.<sup>27</sup>

## 7 Conclusion

Shannon information has traditionally been seen as a rung on a ladder that takes one to naturalised representation. In this context, Shannon information is associated with the outcomes and probability distributions of neural and environmental states. This project, however, obscures a novel way in which Shannon information enters into cognition. Probabilistic models of cognition treat cognition as an inference over representations of probability distributions. This means that probabilities may enter into cognition in two distinct ways: as the objective probabilities of neural vehicles and/or environmental states occurring and as the subjective probabilities that describe the agent's expectations. Two types of Shannon information are associated with cognition accordingly: information that pertains to the probability of the neural vehicle occurring and information that pertains to the represented probabilistic content. The former is conceptually and logically distinct from the latter, just as representational vehicles are conceptually and logically distinct from their content. Various (conceptual, logical, contingent) relations may connect the two kinds of Shannon information in the brain, just as various such relations connect traditional categorical vehicles and their content. Care should be taken, however, not to conflate the two. For, as we know, much trouble lies that way.

## Acknowledgements

This paper has been greatly improved by comments from Matteo Colombo, Carrie Figdor, Alastair Isaac, Oron Shagrir, Nick Shea, Ulrich Stegmann, Filippo Torresan, and two anonymous referees. An early version of this paper was presented on 1 June 2016 at the 30th Annual International Workshop on the History and Philosophy of Science, Jerusalem. I would like to thank the participants and organisers for their encouragement and feedback.

## Bibliography

Alink, A., C. M. Schwiedrzik, A. Kohler, W. Singer and L. Muckli (2010). "Stimulus predictability reduces responses in primary visual cortex". In: *Journal of Neuroscience* 30, pp. 2960–2966.

---

<sup>27</sup>Colombo and Wright (2018) draw a similar contrast between the two formulations of the free-energy principle. They describe different versions of the principle as involving 'epistemic' and 'physical' probabilities.



- Averbeck, B. B., P. E. Latham and A. Pouget (2006). “Neural correlations, population coding and computation”. In: *Nature Reviews Neuroscience* 7, pp. 358–366.
- Bar-Hillel, Y. and R. Carnap (1964). “An outline of a theory of semantic information”. In: *Language and Information*. Reading, MA: Addison-Wesley, pp. 221–274.
- Barlow, H. B. (1969). “Pattern recognition and the responses of sensory neurons”. In: *Annals of the New York Academy of Sciences* 156, pp. 872–881.
- Carhart-Harris, Robin, Robert Leech, Peter Hellyer, Murray Shanahan, Amanda Feilding, Enzo Tagliazucchi, Dante Chialvo and David Nutt (2014). “The entropic brain: A theory of conscious states informed by neuroimaging research with psychedelic drugs”. In: *Frontiers in Human Neuroscience* 8, pp. 1–22.
- Chang, D., D. Song, J. Zhang, Y. Shang, Q. Ge and Z. Wang (2018). “Caffeine caused a widespread increase in brain entropy”. In: *Scientific Reports* 8, p. 2700.
- Clark, A. (2013). “Whatever next? Predictive brains, situated agents, and the future of cognitive science”. In: *Behavioral and Brain Sciences* 36, pp. 181–253.
- Colombo, M. and P. Seriès (2012). “Bayes on the brain—On Bayesian modelling in neuroscience”. In: *The British Journal for the Philosophy of Science* 63, pp. 697–723.
- Colombo, M. and C. Wright (2018). “First principles in the life sciences: The free-energy principle, organicism, and mechanism”. In: *Synthese*. DOI: [10.1007/s11229-018-01932-w](https://doi.org/10.1007/s11229-018-01932-w).
- de Finetti, B. (1990). *Theory of Probability*. Vol. 1. New York, NY: Wiley & Sons.
- Deneve, S. (2008). “Bayesian spiking neurons I: Inference”. In: *Neural Computation* 20, pp. 91–117.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
- (1983). “Précis of *Knowledge and the Flow of Information*”. In: *Behavioral and Brain Sciences* 6, pp. 55–90.
- (1988). *Explaining Behavior*. Cambridge, MA: MIT Press.
- (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.
- Egan, F. (2010). “Computational models: a modest role for content”. In: *Studies in History and Philosophy of Science* 41, pp. 253–259.
- Egner, T., J. M. Monti and C. Summerfield (2010). “Expectation and surprise determine neural population responses in the ventral visual system”. In: *Journal of Neuroscience* 30, pp. 16601–16608.

- Eliasmith, C. (2005a). "A new perspective on representational problems". In: *Journal of Cognitive Science* 6, pp. 97–123.
- (2005b). "Neurosemantics and categories". In: *Handbook of Categorization in Cognitive Science*. Ed. by H. Cohen and C. Lefebvre. Amsterdam: Elsevier, pp. 1035–1055.
- Feldman, J. (2000). "Minimization of Boolean complexity in human concept learning". In: *Nature* 407, pp. 630–633.
- (2012). "Symbolic representation of probabilistic worlds". In: *Cognition* 123, pp. 61–83.
- Fiser, J., P. Berkes, G. Orbán and M. Lengyel (2010). "Statistically optimal perception and learning: From behavior to neural representations". In: *Trends in Cognitive Sciences* 14, pp. 119–130.
- Floridi, L. (2011). *The Philosophy of Information*. Oxford: Oxford University Press.
- Friston, K. (2009). "The free-energy principle: a rough guide to the brain?" In: *Trends in Cognitive Sciences* 13, pp. 293–301.
- (2010). "The free-energy principle: A unified brain theory?" In: *Nature Reviews Neuroscience* 11, pp. 127–138.
- (2013). "Life as we know it". In: *Journal of the Royal Society Interface* 10, p. 20130475.
- Friston, K. and K. E. Stephan (2007). "Free-energy and the brain". In: *Synthese* 159, pp. 417–458.
- Gallistel, C. R. and J. T. Wilkes (2016). "Minimum description length model selection in associative learning". In: *Current Opinion in Behavioral Sciences* 11, pp. 8–13.
- Grice, P. (1957). "Meaning". In: *Philosophical Review* 66, pp. 377–388.
- Griffiths, T. L., N. Chater, C. Kemp, A. Perfors and J. B. Tenenbaum (2010). "Probabilistic models of cognition: exploring representations and inductive biases". In: *Trends in Cognitive Sciences* 14, pp. 357–364.
- Griffiths, T. L., E. Vul and A. N. Sanborn (2012). "Bridging levels of analysis for probabilistic models of cognition". In: *Current Directions in Psychological Science* 21, pp. 263–268.
- Gross, C. G. (2007). "Single neuron studies of inferior temporal cortex". In: *Neuropsychologia* 46, pp. 841–852.
- Isaac, A. M. C. (2019). "The semantics latent in Shannon information". In: *The British Journal for the Philosophy of Science* 70, pp. 103–125.

- Kanwisher, N., J. McDermott and M. M. Chun (1997). “The fusiform face area: A module in human extrastriate cortex specialized for face perception”. In: *Journal of Neuroscience* 17, pp. 4302–4311.
- Kemp, C. (2012). “Exploring the conceptual universe”. In: *Psychological Review* 119, pp. 685–722.
- Knill, D. C. and A. Pouget (2004). “The Bayesian brain: the role of uncertainty in neural coding and computation”. In: *Trends in Neurosciences* 27, pp. 712–719.
- Logothetis, N. K. and D. L. Sheinberg (1996). “Visual object recognition”. In: *Annual Review of Neuroscience* 19, pp. 577–621.
- Ma, W. J. (2012). “Organizing probabilistic models of perception”. In: *Trends in Cognitive Sciences* 16, pp. 511–518.
- Ma, W. J., J. M. Beck, P. E. Latham and A. Pouget (2006). “Bayesian inference with probabilistic population codes”. In: *Nature Neuroscience* 9, pp. 1432–1438.
- MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.
- Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.
- Millikan, R. G. (1984). *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.
- (2000). *On Clear and Confused Ideas*. Cambridge: Cambridge University Press.
- (2001). “What has natural information to do with intentional representation?” In: *Naturalism, Evolution and Mind*. Ed. by D. Walsh. Cambridge: Cambridge University Press, pp. 105–125.
- (2004). *The Varieties of Meaning*. Cambridge, MA: MIT Press.
- Papineau, D. (1987). *Reality and Representation*. Oxford: Blackwell.
- Piantadosi, S. T., J. B. Tenenbaum and N. D. Goodman (2016). “The Logical Primitives of Thought: Empirical Foundations for Compositional Cognitive Models”. In: *Psychological Review* 123, pp. 392–424.
- Pouget, A., J. M. Beck, W. J. Ma and P. E. Latham (2013). “Probabilistic brains: Knows and unknowns”. In: *Nature Neuroscience* 16, pp. 1170–1178.
- Rahnev, D. (2017). “The case against full probability distributions in perceptual decision making”. In: *bioRxiv*. DOI: [10.1101/108944](https://doi.org/10.1101/108944).
- Ramsey, F. P. (1990). *Philosophical Papers*. Ed. by D. H. Mellor. Cambridge: Cambridge University Press.

- Ramsey, W. M. (2016). "Untangling two questions about mental representation". In: *New Ideas in Psychology* 40, pp. 3–12.
- Rieke, F., D. Warland, R. R. van Steveninck and W. Bialek (1999). *Spikes*. Cambridge, MA: MIT Press.
- Saxe, G. N., D. Calderone and L. J. Morale (2018). "Brain entropy and human intelligence: A resting-state fMRI study". In: *PLoS ONE* 13, e0191582.
- Scarantino, A. and G. Piccinini (2010). "Information without truth". In: *Metaphilosophy* 41, pp. 313–330.
- Shea, N. (2007). "Consumers need information: Supplementing teleosemantics with an input condition". In: *Philosophy and Phenomenological Research* 75, pp. 404–435.
- (2014a). "Exploitable isomorphism and structural representation". In: *Proceedings of the Aristotelian Society* 114, pp. 123–144.
- (2014b). "Neural signaling of probabilistic vectors". In: *Philosophy of Science* 81, pp. 902–913.
- (2018). *Representation in Cognitive Science*. Oxford: Oxford University Press.
- Skyrms, B. (2010). *Signals*. Oxford: Oxford University Press.
- Sprevak, M. (2013). "Fictionalism about neural representations". In: *The Monist* 96, pp. 539–560.
- Stegmann, U. E. (2015). "Prospects for probabilistic theories of natural information". In: *Erkenntnis* 80, pp. 869–893.
- Tenenbaum, J. B., C. Kemp, T. L. Griffiths and N. D. Goodman (2011). "How to Grow a Mind: Statistics, Structure, and Abstraction". In: *Science* 331, pp. 1279–1285.
- Timpson, C. G. (2013). *Quantum Information Theory and the Foundations of Quantum Mechanics*. Oxford: Oxford University Press.
- Usher, M. (2001). "A statistical referential theory of content: Using information theory to account for misrepresentation". In: *Mind and Language* 16, pp. 311–334.
- Wiener, N. (1961). *Cybernetics*. 2nd ed. New York, NY: Wiley & Sons.