Shannon + Friston = Content:

Intentionality in predictive signaling systems[1]

Carrie Figdor

(To appear in *Synthese*. Please contact me at carrie-figdor@uiowa.edu if you wish to circulate or quote from this final draft.)

Abstract

What is the content of a mental state? This question poses the problem of intentionality: to explain how mental states can be about other things, where being about them is understood as representing them. A framework that integrates predictive coding and signaling systems theories of cognitive processing offers a new perspective on intentionality. On this view, at least some mental states are evaluations, which differ in function, operation, and normativity from representations. A complete naturalistic theory of intentionality must account for both types of intentional state.

Introduction

What is the content of a mental state? This question expresses the problem of intentionality, or that of explaining the fact that mental states are about other things — for example, that a thought is about rain, or a visual percept is of a face. Aboutness in turn is

[1] Acknowledgements: I am very grateful to the following for comments and questions that were invaluable for helping me develop this paper: Alistair Isaac, Mark Sprevak, and Robert Rupert (early stage); Beate Krickel, Stephen Mann, and the entire lively audience at the conference "Mental Representations in a Mechanical World: The state of the debate in philosophy of mind and philosophy of science" (Nov. 28-29, 2019), at Ruhr-Universitat Bochum, and to Albert Newen for his support for my research visit at Bochum (middle stage); and three anonymous referees for this journal (final stage).

identified with representation: the thought represents raining, the percept represents a face. Following this philosophical tradition, the corresponding problem in cognitive science is to explain mental representation within some type of physicalist framework. This paper argues that contemporary theories of cognitive processing motivate reconsideration of this received understanding of intentionality and thus of the problem to be solved. They show that at least some intentional mental states are evaluations, not representations, and that identifying intentionality with representation leaves out an essential type of mental content. A complete naturalistic theory of intentionality must explain both types.

To support this conclusion, I draw on aspects of two leading research programmes in cognitive science: signaling systems theory, based in Shannon's (1948) mathematical theory of communication, and predictive coding theory, based in Friston's (2005, 2010) account of how brains process information. To date, these programmes have been conceptually elaborated and critically discussed largely in isolation from one another, even though they offer complementary perspectives on the processing of intentional states in adaptive biological systems. To a first approximation, signaling systems theory provides an account of the basic structure of efficient signaling systems, while predictive coding is an account of how signaling efficiency is realized in specific kinds of these systems. Their explicit conceptual integration makes vivid the dual nature of aboutness in contemporary cognitive science and its departure from philosophical tradition. This broader issue has not been raised within either programme independently considered.

I will proceed by motivating the conceptual integration of the two research programmes and elaborating the consequences of this integration for understanding the nature of intentionality. Details matter for setting up the general problem, but it arises vividly once the pieces are in place. I begin with a clash. Skyrms (2010a, 2010b) offers a prominent account of the emergence of signaling systems and signal content, and Godfrey-Smith (2012) counters with an intuitively compelling content-related objection. This objection motivates the integrated predictive signaling systems framework. To a first approximation, prediction error signals in these systems are externally-constrained internal evaluations that the system can use to adjust to various contingencies. These signals have their own purpose, conditions of operation, and type of normativity; they are not plausibly understood as representations, as Shea (2012) proposes. Godfrey-Smith's objection also raises the question of why mental evaluations are so counterintuitive. I trace this response to the conception of intentionality enshrined in philosophy of mind by way of standard truth-conditional semantics and its mental correlate, the propositional attitudes. I also consider how the evaluationist view relates to anti-representationalism and to Bayesian descriptions of cognitive processing, and conclude by considering a few objections.

For space reasons I set aside some basic issues. Most importantly, I do not provide empirical (including neuroscientific) evidence for predictive coding, distinguish predictive coding algorithms, defend the predictive coding approach from objections, or distinguish between perceptual and cognitive processing (e.g., Spratling 2016, 2017). If the brain is *not at all* a predictive coding machine, that would not falsify the signaling systems perspective on cognitive processing, but one would have to give up a predictive coding explanation of what senders with brains do to efficiently encode input. (This point will become clear below.) Alternatively, predictive coding might be restricted to specific processes.

For some, the fact that predictive coding is currently so open to refutation makes my approach unacceptably speculative. But the integration of a new theory with an established one is considered a key theoretical virtue of the former. These two theories already share mathematical background, which informs empirical cross-talk (e.g. Sengupta, Semmler, and Friston 2013) and is noted in philosophical work (e.g., Skyrms 2010a, Clark 2013, 2016). I augment this with conceptual integration, which is most vivid in their overlap in the case of minded biological agents (Clark 2016; Shea et al. 2017; Shea 2014; Godfrey-Smith 2014). In addition, even if we treat predictive coding as purely hypothetical, we stand to gain insight into how intuitions about intentionality can harmfully constrain cognitive science.

I also do not defend the basic commitment to the possibility of a naturalistic account of intentionality. It is common ground to distinguish natural and non-natural meaning (Grice 1957). Naturalization projects treat non-natural meaning as a special case of natural meaning in which the possibility of misrepresentation arises due to learning, evolved function, or some other naturalistically acceptable feature. Natural meaning, meanwhile, is often analyzed at least in part in terms of some sort of world-organism or informational relation, following Dretske (1981, 1983). I focus on the contribution of this informational aspect to internal processing. This does not rule out non-information-based theories of representational content, such as Millikan's (1984, 1989) consumer-based teleosemantics, as representational content and its explanation are not at issue. Similarly, debate between representationalism and anti-representationalism is touched on only to the extent that it affects the question of whether intentionality just is representation. In this discussion, representationalism is the view that mental states are internal states that represent other items: to be an inner state with intentionality is to be a mental representation (e.g., Jacob 2019).[2] I assume a generic commitment to representational states for the sake of argument, as it is a widely-held position from which I can highlight what cognitive science is telling us about intentionality. If my position is sound, anti-representationalists – in its broadest sense, those who reject representations so conceived – must adopt a view on evaluations, however they articulate a representation/evaluation distinction in their preferred terms.[3]

Finally, I use the terms "semantics", "meaning", and "content" interchangeably except when specifically distinguishing linguistic meaning and mental content; and I use "cognitive" and "mental" interchangeably. This usage falls well within the range of ways in which these terms are commonly used in the relevant literatures. The term "information" is multiply ambiguous for independent reasons, though debates about its meanings do not play a role here. As will become clear, however, I reject the idea that Shannon's theory is

---

[2] The view of consciousness or phenomenal experience that explains it in intentional terms is also called representationalism, thus presupposing the identification of intentionality with representation. This form of representationalism makes a brief appearance in fn. 20.

[3] These terms can vary significantly. Anti-representationalism predates the predictive processing framework (e.g. Stich 1983; Dennett 1987) or is independent of it (e.g., Ramsey 2007). Some anti-representationalists provide accounts of intentionality in terms of embodiment, in different ways (e.g. Hutto and Myin 2017; Chemero 2009; Bruineberg and Rietveld 2014 integrate enactivism with Friston's work, focusing on the environmental structures to which the organism responds). Since I am not arguing for or against any form of anti-representationalism or representationalism, I set aside these nuances.

non-semantic (Piccinini and Scarantino 2011); to the contrary, my discussion contributes to recent advances in philosophy of mind and naturalized semantics that draw fresh insights from Shannon's theory (e.g., Usher 2001, Lombardi 2005, Lean 2014, Scarantino 2015, Lombardi et al. 2016, Isaac 2019, Sprevak 2019).

.

1. Content in signaling systems

The blending of naturalized semantics and Shannon's information theory is most familiar from Dretske (1981, 1983), who was also an early proponent of a probabilistic theory of content. As a group, such theories eschew a causal relation as the fundamental naturalistic relation that fixes informational content in favor of probabilistic regularities. In Dretske's account, for a signal *s* to carry the information about the state of affairs that *r* is F, the probability of *r*'s being F given *s* (and background knowledge *k*) had to equal 1. The signal's occurrence guaranteed that the state occurred, making it possible to know the state occurred by observing the signal. Contemporary probabilistic theories adopt the more forgiving position that a signal carries information about an event, object, or state of affairs if they are probabilistically related, without guaranteeing that the state occurred.

While Dretske relied on Shannon for prompting ideas relevant to his epistemic concerns, Skyrms' view of content hews more closely to Shannon's original framework in ways that will prove critical.[4] Skyrms analyzes signal content in terms of probability

---

[4]Sayre (1983:78-9) argues that Dretske rejected every important aspect of Shannon's theory except for the use of a quantitative measure in an account of content. Dretske (1983:82-3) admits as much, but insists his interest was in "the ideas clothed in mathematical dress".

distributions defined over sets of items. This general approach is common to probabilistic theories, although unlike other probabilistic theorists Skyrms identifies content with the set of changes in the probabilities of states given the signal compared to the prior probabilities of the states.[5] The informational content of a signal depends on this comparison. Thus, assume the world provides us with a set of individuated states that a signal could be used to indicate. The signal's content consists in *how* it "moves" the conditional probabilities of each member in the set of world states given the signal, while the quantity of information it provides is measured by *how much* it "moves" these probabilities (Skyrms 2010a: 34). It is this identification of content with (a set of) probability changes that induces both the integration of signaling systems theory with predictive coding and the break with philosophical tradition regarding intentionality. I'll describe the theory in sufficient detail now in order to make these points later on.

Skyrms (2010b: 156) illustrates this with a simple case. Start with two equiprobable states of the world S1 and S2, and two signals A and B that are equally probable of being

---

Unfortunately, Dretske bowdlerized Shannon's ideas to fit philosophical assumptions about intentionality, rather than adjusting the latter to Shannon's unadulterated ideas.

[5] Scarantino's (2015) Probabilistic Difference Maker Theory shares this feature, but with an important difference discussed in the text; also, on his view content is fixed by a 3-place relation where all but one of the world states in the set play a role as background data. Other probabilistic accounts identify the content of a representation with one state of affairs in the set, differing in how that condition is identified (e.g. Shea 2007; Eliasmith 2005; Stegmann 2015 for a critical review).

selected from a set of possible signals. Signal A's being sent "moves" the probability of S1 to .9 and S2 to .1, while Signal B's being sent "moves" the probability of S1 to .1 and S2 to .9. For example, given smoke (signal A), the probability of fire (S1) increases and the probability of rain (S2) decreases. In this toy example, signals A and B are identical in terms of how much information they contain (which depends on *how much* the probabilities have moved), but not in what informational content they contain (which depends on *how* the probabilities have been affected). The contents differ because the probabilities are affected differently.

This content is mathematically described by a vector whose slots are occupied by the logs of values that signify each of the increases or decreases in conditional probabilities of states given a signal, compared to the states' prior values.[6] Propositional content is the

[6] Worked out, the content of signal A is represented by the vector of the logs of these ratios:

$$\nu(A) = \langle \log\left(\frac{P(S1|A)}{P(S1)}\right), \log\left(\frac{P(S2|A)}{P(S2)}\right)\rangle = \langle \log\frac{(.9)}{(.5)}, \log\frac{(.1)}{(.5)}\rangle = \langle \log(1.8), \log(.2)\rangle$$

and likewise, mutatis mutandis, for the content of signal B. If, with Skyrms (and following Shannon), we choose log base 2 and round to 2 decimal points, ν (A) = <.85, -2.32>. What makes a vector semantic is its intended interpretation as a model of an actual signaling system type. The example presupposes but omits reference to a physical system that observes world-states and sends signals, even though world-states and signals are equally events as far as the mathematical formalism is concerned. Each vector slot represents a distinct possible change for that system (Skyrms 2010a: 35 fn. 2). World states are "whatever the sender can discriminate", and the evolution of categories is "driven by pragmatics – available acts and payoffs" (Skyrms 2010a: 107-109, 139; Harms 2010). Like special case in which the probability of at least one state in the set decreases to zero given the signal; this change is represented by ~∞ in that vector slot (a zero represents no change). Rather than thinking of propositional meaning in the usual terms of the set of possible worlds where a proposition is true, we think of it in terms of the set of possible worlds that retain at least some probability following a change in which at least one world is ruled out.

While Skyrms' framework is far from traditional even at first blush, in some ways it is a conservative extension of standard formal semantics. Skyrms himself considers his account of propositional content to be a generalization of standard theories. We can also intuitively understand the idea of a vector with multiple slots. For example, if meaning is use, and use is or involves what one can infer from a signal, it make sense to analyze meaning (or content) in terms of multiple inferences. A sentence such as "Smoke means fire" may not mention all the states whose probabilities are affected by a smoke signal, but in principle the unmentioned probabilities may be considered pragmatically omitted elements of its content. Finally, if a signal sender's job – especially in perceptual processing – is conceived as that of inferring the distal cause of its input given multiple possible causes, a theory like Skyrms' can dissolve the problem by embracing all the possible causes as part of the signal's content.

2. Signaling systems and predictive coding

---

other naturalists, Skyrms elaborates his view with nonconceptual signals – which lack the syntax of natural language – but also indicates how to extend the theory to complex signals.

But why should we go along with Skyrms' idea of identifying signal content by means of *changes* in probabilities, let alone multiple changes? Godfrey-Smith (2012) raises precisely this objection. Surely an organism wants to know how the world probably stands now given the signal, not how much it changed from before:

> A natural reaction to Skyrms' view is that it seems wrong to say that the content of a signal is a matter of how a signal has changed the possibilities of states. Instead, a signal's content is a matter of how the world is, or probably is. … If a receiver can learn from a signal how the world probably is, it is of no use to also learn how the probabilities have changed. And if a receiver does not learn from a signal how the world probably is, the signal cannot be used to guide action. (2012: 1291).

If so, Skyrms' account does not provide a semantics for signals that receivers want. Instead, he suggests, the vector slots should list the posterior probabilities of the states given the signal. Scarantino (2015:429 and fn. 15) agrees. While he includes changes in probabilities of states in the signal's content, he also stipulates the inclusion of the posterior probabilities in a complete description of its content. Shea et al. (2017) also agree. They adopt this view for informational content, leading them to stipulate that when a signal doesn't affect any of the probabilities it has no content.

This objection is intuitively compelling: what can it mean for the aboutness of a signal to be explained in terms of a *change in* probabilities? It makes intuitive sense to hold that a signal is about a state of affairs with which it is probabilistically related; but it is puzzling to understand what a signal is about if it is only related to a relation between probabilities, not to states of affairs. As Scarantino (2015: 429) puts the objection, "[i]nformative signals do not just tell us how much probabilities have changed; they also tell us what are the states of affairs that had their probabilities changed." In section 4 I offer a plausible explanation of our intuitive puzzlement. Here, I will show that the puzzlement is misguided in a way that overlooks a fundamental aspect of Shannon communication.

Consider Shannon's original (1948: 380) Fig. 1 diagram of a communication channel, drawn with linguistic signals between people in mind:

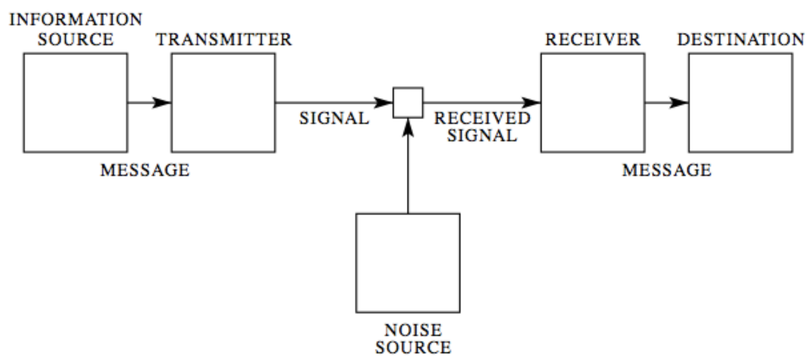

Fig. 1 — Schematic diagram of a general communication system.

I'll use Shannon's terms (always capitalized) when ambiguity threatens, but to cohere with contemporary usage I'll also call the Message the input, the Transmitter the sender (or producer), and the Information Source the world or some aspect of it.[7] Thus, the world

[7]Ambiguity can arise because "signal" and "message" are often used as synonyms, and "sender" and "receiver" are used to pick out distinct individuals or distinct subprocesses within one individual (and can go in both directions: Lean 2014). Where Shannon has two "signals" – the Message, with the Information Source and the Transmitter as the relata, and the Signal, with the Transmitter and the Receiver as the relata – some philosophical presentations simplify his schema to show only one (Scarantino 2015: 424, fn. 6, Cao 2012: 50, Lombardi 2005: 24). Although Shannon's distinction does not always matter in a given

provides sensory or linguistic input. The sender operates on the input to yield the signal she sends. The receiver decodes the received signal for use, *modulo* transmission errors and noise in the communication channel. Thus, a sender might get English input from a speaker, encode this in Morse, and transmit the encoded signal through a telegraph wire to a receiver, who decodes it back into English with as little error as achievable.

From Skyrms' perspective, Fig. 1 represents an existing signaling system in some kind of equilibrium. His primary theoretical interest lies in explaining how such systems come to exist from scratch: following Lewis (1969), they arise from the game-theoretic coordination of signing and acting behaviors between entities that achieve some kind of equilibrium of behaviors and payoffs. In the course of this bootstrapping, such agents exploit the natural probabilistic relationships required for signals to have natural and non-natural meaning.

Shannon's interest, however, was in reliable signaling in the face of noise, given a signaling system. His solution depended on efficient coding, which he analyzed in part in terms of reducing signal redundancy to within acceptable limits.[8] From his perspective, the

---

discussion, this omission may suggest that the encoding step involves nothing of philosophical interest. Others use the full version (Godfrey-Smith 2013: 43, Lombardi et al. 2016: 1985, and, in essence, Martinez 2018: 4 in pre-print).

[8] Shannon's entropy formula captures the average uncertainty of a signal in the light of the probability distribution defined over the members of the set from which the signal might be chosen. Slightly more technically, a signal (e.g. "Q") is a value of a random variable X selected from a set of possible values (e.g. "A", "B", …, "Z"); the entropy of X is defined

significant aspect of communication "is that the actual message is one *selected from a set* of possible messages" (1948: 1, his italics), since this statistical structure could be exploited to achieve data compression and thereby help eliminate noise as a communication problem.[9] Natural languages contain a great deal of redundancy. For example, the English

---

over all the possible values that X can take weighted by each value's frequency. This average uncertainty (entropy) is the sum of the weighted average of the log probabilities of each signal in a set of possible signals. It is greater if there are more possible signals to select from and if their individual probabilities of selection are closer to being equiprobable. His mutual information formula, defined for joint probability distributions, expresses the reduction in average uncertainty (entropy) of X given another variable Y whose value is known – for example, after "Q" is chosen, the entropy of X (ranging over the English alphabet) is greatly reduced because "U" is now statistically highly likely to be chosen. The mutual information is zero if the variables are independent, but naturalization projects rely on probabilistically related sets, in particular world states and signals. In addition, Shannon's goal of reliable communication is achieved in his theory by compression (to eliminate redundancy) and selective insertion of redundancy (to manage noise). I focus on encoding for compression. (I thank an anonymous reviewer for drawing my attention to this distinction.) A predictive coding system also uses precision weighting to manage noise (see fn. 11).

[9] Soni and Goodman (2017) recount how prior efforts to eliminate noise focused on trying to improve the transmission channels, such as by making undersea cables as strong and insulated as possible. Obviously, physical media matter for reliability, whether this is

alphabet has 26 letters to choose from, with widely different probabilities of being selected. This redundancy makes crossword puzzles solvable and interesting (Shannon and Weaver 1949; Weaver 1949) and lies behind the values assigned to letter tiles in Scrabble. Statistical dependencies mean it makes almost no difference if a signal of "QUEEN" is corrupted into "Q...EEN"; in terms of entropy, sending “Q” reduces your recipient’s uncertainty a lot about the next letter. It follows that Shannon's Transmitter was never conceived as a passive conduit for Messages, even if the channel through which the produced Signals were sent was ideally exactly that. Shannon only took into explicit consideration alterations of Message format, but other types remain possible.[10]

Skyrms’ semantic vectors replicate the structure of Shannon’s solution; as Isaac (2019: 111) puts it, these vectors capture “the basic semantic object implicit in standard information theory”. However, the simplicity of a toy example obscures the importance of statistical structure in Shannon’s framework. In the case of linguistic Messages, the statistical structure is ultimately the conventions or patterns of behavior among language users that create meaning, as Skyrms theorizes. That Shannon abstracted from language

undersea cable quality and insulation or neural integrity and myelin sheaths. The point is that reliability is redefined in terms of the probabilities of the signals in a set such that, given a physical channel (with a certain transmission capacity), the same message can be encoded in ways that make transmission through that channel more or less reliable.

[10] For example, he mentions (1948: 15) semantic compression in opposite extreme cases: when an English sentence is transformed into Basic English, which contains about 850 words, and James Joyce’s *Finnegans Wake*, where neologisms replace long phrases.

use to statistical structure is not usually emphasized in philosophical semantics, which tends to analyze content individualistically – in Dretske's case, individualizing Shannon's original information measures so that a signal carries information independently of its relations to other signals. In the case of world-state Messages, the statistical structure is provided by distinct probabilistic relationships in the world that we can notice, such as those between smoke and fire, or between "Q" and "U". In other terms, a Message of smoke makes fire redundant in the same way that a Message of "Q" makes "U" redundant. These observable patterns are the objective relations that naturalistic theories of content exploit. Once a signaling system exists, a Signal of "smoke" sent within that system will reduce uncertainty regarding fire because of these objective redundancies. Note that "signaling system" can refer to a system comprised of at least two agents who come to cooperate in their behavior (Skyrms' concern) or to a statistically structured set of signals from which agents can choose (Shannon's concern). These are complementary.

The efficiency aspect of Shannon communication directly links signaling systems theory to predictive coding theory in a way that makes clear how to counter Godfrey-Smith's objection to Skyrms. Efficient coding is not just of concern to wartime telegraphy. It is an imperative for biological entities that use signals to guide their adaptive behavior. Because creating and sending signals are costly operations for a biological entity, it will encode them as efficiently as possible and send them as efficiently as possible in the light of its needs and the circumstances (Clark 2016: 25-26, 28-29; 2013). Predictive coding is a leading empirical theory of how biological systems with brains satisfy this imperative. Rao and Ballard (1999) initially proposed a predictive coding theory of vision, while Friston generalized predictive coding to brain function in general "to explain perception and action and everything mental in between" (Hohwy 2013). Clark has elaborated Friston's theory along enactivist (active inference) lines; Hohwy (op.cit.) takes a more internalist approach. My proposed integration of these theories of communication and encoding is compatible with either elaboration.

In a predictive coding system, generative models (GMs) produce predictions about future input, the predictions are compared to actual sensory input to yield prediction errors if any, and signals encoding these prediction errors (PE signals) are used to update the GMs and guide action. [11] There is a hierarchy of such processing; except where noted, I will take this aspect for granted. Ultimately the system seeks to minimize long-term average prediction error through iterations of this processing over time. Even without predictive coding, self-generated aspects of input must be suppressed, but predictive coding entails that all of it may be. By minimizing prediction error, the agent minimizes surprise, defined as "the difference between the organism's predictions about its sensory inputs (embodied in its models of the world) and the sensations it actually encounters" (Friston and Stephan 2007; Friston et al. 2012). Conceptually, Shannon's reducing uncertainty is Friston's reducing surprise (or "surprisal") in the case of adaptive systems that exploit probabilistic

[11] I suppress many important details to focus on the theory's basic functional distinctions between the generative model and predictions, and the prediction error, and assume its basic principle that the organism aims to minimize long-term average prediction error. For example, Friston and Frith (2015) distinguish between PE signals used to update the GM and those that are used to update action; these correspond to sensory and proprioceptive predictions. I discuss precision weighting of the PE signal in section 3.

relationships to generate expectations. At each level of the processing hierarchy, the system efficiently minimizes the probability it will confront input it does not expect and reduces the energy it expends on creating and sending signals.

In the predictive coding literature, discussion of the uncertainty or surprise that is reduced focuses on that of the generative model, conceptualized as a complex internal representation that is the Receiver of error signals. (For my purposes, generative models and predictions can be treated as a package.) For example, in approximate Bayesian descriptions of this process, the probability of a hypothesis is raised or lowered in response to new evidence, where the generative model is comprised of these hypotheses and the prediction error signal is the evidence. But the heart of predictive coding in adaptive systems involves the uncertainty of the Transmitter or sender, which produces the prediction error signals that drive the system. In terms of Shannon's Fig. 1, predictive coding is a theory of what Transmitters with brains do when encoding Messages.

Now reconsider Godfrey-Smith's objection to Skyrms. In the simplest case of signaling, Skyrms' semantic vectors capture changes in probabilities of world-states, given a signal, compared to their prior probabilities. In our world, the probability of fire given smoke increases compared to the prior probability of fire; in a world of constant fire, there is no such increase given smoke. However, while signaling systems in general need not have a "mental life", those that do are a special case in which information transfer is driven by adaptive dynamics (Skyrms 2010a: 43-4). Adaptive signaling systems with mental lives have intentional states with conceptual or nonconceptual content. For such systems, changes in objective and subjective probabilities are involved; the formalism is compatible with both (Skyrms 2010a: 43-44; Isaac 2019: 109). In particular, these systems have subjective probabilities (expectations) over the world-states. Such systems strive to bring their expectations in line with the objective probabilities in the world, *modulo* adaptive biases. If they are predictive signaling systems, they do this by comparing what they predict they will detect at T + 1 given what they detected at T, and what they detect at T + 1; if this comparison yields a difference, it is captured in a PE signal. In these systems, the customary omission of mention of the presupposed realizing system (fn. 5) also omits a key intermediary in their world-state-to-representation processing. In effect, informativeness of the world is encoded in one type of intentional state, representation of the world in another.

Godfrey-Smith's challenge is that Receivers of signals don't care about changes in the probabilities; what they need to know are the posterior probabilities—where they end up. But in a predictive signaling system, these changes are *exactly* what Receivers want. They want to be informed only of what they did not expect. A Transmitter that anticipates future input is in a stronger position to maximize efficient signaling. It will encode how and by how much its subjective probabilities differ from the actual Message, and send only the upshot to the Receiver. It can ignore input that is not news, and only create and send signals that contain only news. In a world of constant fire, such a Transmitter wastes no energy responding to smoke by producing and sending unneeded signals of smoke to a Receiver who already knows there's fire. The system's GMs and predictions may come to be appropriately tuned to the posterior probabilities, but there is no reason to include them in the content of the Transmitter's signal or for the Transmitter to send a redundant signal.

In brief, a Shannon Transmitter gets Messages and produces efficiently encoded Signals; a Friston Transmitter is a specific kind of Shannon Transmitter, found in at least some biological systems with brains, that transforms Messages into Signals that encode just the differences between what this system expects and the Messages it gets; and Skyrms' semantic vectors encapsulate these Signals' contents. With respect to the content of prediction error signals in adaptive predictive signaling systems with mental lives, Shannon, Friston, and Skyrms are all on the same page. Godfrey-Smith's objection fails for these signals.

3. Predictive signaling systems and mental evaluations

Skyrms' semantic vectors encode differences in probabilities of the members of a set (e.g., choice of A or … choice of Z) given an event (e.g., choice of Q) from their prior probabilities. In the special case of predictive coding systems with brains, these vectors capture the content of prediction error signals. But as Godfrey-Smith's objection shows, this way of thinking about content is very puzzling. How can a difference be *about* anything?

It is common ground that generative models, predictions, and prediction error signals are intentional mental states with distinct functional roles in cognitive processing. But there is a further difference. As Clark (2016: 46; 2013: 187) describes them, prediction error units "bear fine-grained information about very specific failures of match" between traditional representations and the world; they are "representations of error" that pass along "unexplained input". But what exactly is a failure-to-match content? The signals encode *a difference*, but to encode a difference is not to misrepresent, represent falsely, or represent an error. These descriptions imply either that the prediction error signal at least sometimes represents truly (which it doesn't) or else that it is always inevitably *wrong* (which makes no sense). It is this kind of contentful state that needs illumination. Representation-talk may be appropriate for the GM and predictions, but it only obscures this explanandum.

The integrated predictive signaling system model suggests that, at any level in the processing hierarchy, the content of the PE signal is an internally accessible evaluation of the system's (or Receiver's) states and actions in relation to its expected environmental contingencies. It is an evaluation, not a representation. It is used to calibrate its expectations, goals, and actions by informing the organism of "its state of readiness for goal-directed, adaptive activity" (Mackay 1969: 23) – that is, to how well adjusted it is to how the world appears to be unfolding given how it expects it to unfold and its goals. Evaluations – PE signals – are generated at any level of the processing hierarchy where there is a difference, not just the organism level that concerns Mackay. They play the functional role they do by providing the generative models with internally accessible evaluations of their predictions. A GM uses the PE signal to adjust some aspect of itself or act. That the GM was not accurate or true in this respect is implicit in its response, given its goal of minimizing long-term average prediction error. Describing the PE signal as "indicating a degree of misrepresentation in the generative model" is a gloss on the processing that the predictive coding account deconstructs into distinct intentional states and operations.[12]

[12] As [acknowledgement] (in personal communication) and two anonymous reviewers note, this nudging will be modulated by precision weighting of the PE signal to account for noise or other disturbances. Precision error weighting – conceptually, how reliable the error signal is or how much confidence should be placed in it (Friston and Frith 2015) – helps

There are a number of reasons why evaluations and representations are distinct kinds of intentional state. First, the familiar representational concepts of mind-to-world or world-to-mind fit, mainly used for beliefs and desires respectively (e.g. Anscombe 1963), don't fit the PE signal. Neither beliefs nor desires encode failures-to-match; occasionally, they simply fail to match, as they ought. In contrast, the PE signal doesn't try, and occasionally fail, to match; it ought only to encode a failure-to-match, a discrepancy between the content of an item that ought to match and the state of affairs it ought to match. There is no need to commit to discrepancies (or differences) as a new type of intentional object that it might match, which might seem to justify talk of its accuracy. Such a move towards the ontological promiscuity of Meinong should worry naturalists, given that intentional inexistence is a problem for naturalizing the mind.[13] The mere availability of a

---

explain how the GM responds to the PE signal, but does not affect the latter's status *qua* evaluation. For example, in poor visual conditions the PE signal may be weighted as unreliable; since it is more likely to misevaluate in such conditions, the GM in effect treats the evaluation with a grain of salt. A generative model might also systematically assign more weight to optimistic misevaluations (those that downplay or understate the actual error size), manifesting a kind of Dunning-Kruger effect; see Prosser et al. (2018) on precision weighting and psychopathy. I discuss misevaluation below.

[13] Intentional inexistence is Brentano's (1874/2014) label for the fact that we can think about things that don't exist, such as Santa Claus or unicorns. In response, Meinong (1904) argued for ontological commitment to objects that do not exist – a view that has long been anathema to naturalists.

representational gloss of the PE signal's content is insufficient motivation for exacerbating that problem. If that's what it takes to preserve the identity of intentionality with representation, it is a very high price indeed.

Second, we know false theories can produce true predictions. So even if we assume that the GM's contents (at least, its beliefs) ought to match the world, the PE signal's content ought only to nudge the GM towards making better predictions. It does this by presenting the GM with its evaluative content. It says (roughly): "Discrepancy of this size in this direction."[14] This evaluation is distinct from other normative assessments of other elements in the system – a bad prediction, an inaccurate or false GM. For example, upon receipt of a PE signal, the GM may generate a representation that says (roughly): "This situation is bad for me". I discuss this possibility further below.

Third, if prediction error signals were representations, presumably Dretske's (1983: 57) Xerox principle regarding content transitivity would hold. Since the PE signal is produced in response to sensory input, its content, if representational, ought to match the world and at least part of the GM's content ought to match that of the error signals it

[14] Linguistic descriptions of nonconceptual and subpersonal contents are usually charitably understood as approximations or glosses of the information that a given theory says such states contain (as Martinez and Klein 2016: 284 pointedly note). I assume a similar attitude towards the various content descriptions used by me or others throughout this paper.

receives.[15] This contradicts the predictive processing framework's functional distinction between these intentional states.

Prediction error signals also have their own type of normativity. The possibility of error, which is required for intentionality, is usually understood as the possibility of misrepresentation. On the evaluationist account, the appropriate analogue is misevaluation – an error in evaluating, or (using the usual label) an error in the error. A PE signal is a misevaluation if the difference it encodes is not the actual difference. This can occur if it encodes a difference when there is none – the evaluationist analogue of a hallucination. The other, presumably more usual, misevaluation is an encoded difference that differs from the actual difference – the evaluationist analogue of (e.g.) a COW representation triggered by a horse.[16] However, misevaluation (like evaluation) is a vector concept in that it differs from the actual difference in a direction. If the PE signal encodes the actual difference between prediction and input, it is a Goldilocks error signal. In a misevaluation, the

[15] The Xerox principle (Dretske 1983: 57) is: "If C carries the information that B, and B's occurrence carries the information that A, then C carries the information that A. You don't *lose* information about the original (A) by perfectly reproduced copies (B of A and C of B). Without the transitivity this principle describes, the flow of information would be impossible." Predictive processing offers a different account of this flow.

[16] A problem also occurs when there is no prediction error signal but there should have been one: the system fails to get the evaluation it should have gotten, and it proceeds as if it needs no adjustment. This may be treated as a degenerate case of misevaluation, but is better treated as a different kind of problem – for example, the sender may be damaged.

encoded difference is larger or smaller than the actual difference. A misevaluation is pessimistic if it encodes a larger error that it should have: it exaggerates the error. A misevaluation is optimistic if it encodes a smaller error than it should have: it downplays the error. (I avoid the term "minimizes" since the system as a whole aims to minimize prediction error over the long term – more precisely, it aims to minimize *actual* prediction error.) In both cases, the misevaluation may induce suboptimal adjustment on the part of the Receiver, again in one of two directions. For example, the organism may overcorrect or undercorrect an ongoing reaching movement that is missing its target, depending on the direction of the misevaluation.

Optimism and pessimism, like truth and falsity or goodness and badness, are normative concepts that can be used to qualify more than one type of item – e.g., an optimistic person, belief, prior probability assigned to a hypothesis, or, in this case, a difference. In all these cases, the assessments express a downplaying or exaggerating feature of the item they qualify.[17] Degrees and averaging also make intuitive sense for this

[17] A reviewer suggests that optimism or pessimism depend on the content of the prediction as well as the size of the error: for example, if I predict seeing a lion but in fact encounter an impala, "this may produce a large prediction error, albeit the error intuitively counts as optimistic. A low prediction error relative to the same hypothesis is pessimistic". These uses of the concepts of optimism and pessimism appear to qualify the GM (e.g., its expectations). For PE signals, actual error size and misevaluation are distinct: for example, a very large error term may be exactly right. Suppose N is the actual large discrepancy between the prediction (Lion) and the actual input (Impala). If my PE signal encodes N, it

type of normativity. In terms of Skyrms' semantic vectors, the difference in each vector slot may be optimistic or pessimistic, each misevaluation can differ in its degree of optimism or pessimism, and the vector can be on average optimistic or pessimistic. These features are ill-suited to mental items whose content is usually normatively assessed in terms of truth. One can assign a probability to the truth of a proposition *p* or a belief that *p*, but its being true is not (classically) a matter of degree.[18]

---

is a Goldilocks evaluation. If it encodes a smaller discrepancy than it should (e.g., as if what was in front of me is a juvenile lion), it is an optimistic misevaluation: it tells me there's a discrepancy, but it downplays the difference. If my PE signal encodes a larger error than it should (e.g., as if a rhinoceros was in front of me) it is a pessimistic misevaluation: in this case, it exaggerates the discrepancy.

[18] This issue of degrees may not be a problem for accounts of the GM's representational vehicles in terms of structural similarity (e.g., Gladziejewski and Milkowski 2017). Maplike structures are normatively assessed in terms of accuracy, which comes in degrees. The contents of such representations would still be the worldly states they represent, even if they are not assessed for truth. Also, Kiefer and Hohwy (2018: 2404) suggest that the Kullback-Leibler (KL) divergence – a mathematical measure common to Shannon communication and predictive coding – can be used as a measure of misrepresentation. The KL divergence is a method of averaging over all the log ratios in probability vectors that lets one mathematically compare one probability distribution to a reference probability distribution (see also Skyrms 2010a: 36). Optimally, one minimizes the KL divergence from the reference distribution. They suggest taking the KL divergence between the GM's

Further illumination of the evaluationist proposal comes from contrasting it with Shea’s extension of Millikan's concept of pushmi-pullyu representations to interpret PE content (Millikan 1995; Shea 2012). Pushmi-pullyu representations are representations that “face in two directions at once” (Millikan 1995: 186). They have two types of representational content, one indicative or descriptive and one imperative or directive. Borrowing her familiar bee-dance example, the dance is a pushmi-pullyu representation with indicative content "There is nectar at location X" and imperative content "Fly to location X". In terms of propositional attitudes, in which mental states are analyzed as relations to propositions, a pushmi-pullyu representation involves a primitive combination of attitudes towards a nonconceptual content; sophisticated systems have distinct attitudes (e.g., belief and desire) towards a nonconceptual content or a proposition (Millikan 1995: 196).

Shea proposes that PE signals are pushmi-pullyu meta-representations. From this perspective, the sender (or producer or Transmitter) produces a signal whose content combines a truth condition – the worldly state of affairs that explains the successful

---

posterior distribution and the causal structure of the world as an internal proxy for an objective notion of mismatch between the GM and the world. The KL divergence may also work as an average measure of misevaluation of PE signals, since in predictive signaling systems aiming for KL optimality is minimizing prediction error over the long term. This could provide a measure of whether a system’s (or the Transmitter’s) PE signals are on average Goldilocks signals (or not); thus, a system that systematically generates optimistic misevaluations (long-term, on average) might itself be judged optimistic.

performance of the consumer system's function – with an imperative satisfaction condition that commands what to do in response. On this view the (nonconceptual) content of PE signals is a correctness condition that concerns the content of another representation – for example, "The current visual representation of the location of the light is likely to be false":

> The indicative content is that the content of another representation ("the agent's (first-order) representation of the reward that will be delivered on average for performing a given action") differs from the current feedback, and by how much. The imperative content instructs that it be revised upwards or downwards proportionately. (2012: 4)

However, this is not the most perspicuous interpretation of these signals. The signal informs the generative model by how much its predictions are off and in which direction. The generative model can respond by updating its predictions in the appropriate direction or by initiating an appropriate action or change in course of an ongoing action. The signal does not impel revision of a reward prediction, and a small error in either direction may not be significant enough to be passed up the processing hierarchy, depending on how it is precision-weighted. So the imperative element as described is either absent or optional. It's not clear how to understand an imperative that is optional.

The indicative element is even less apt. The binary (true/false) normativity typically associated with representation, hence meta-representation, is inappropriate. On the meta-representational view, a PE signal's indicative component represents the falsity of the first-order representation. But the current visual representation of reward can be in the ballpark or wildly inaccurate, in either of two directions. The system may be much closer to its goal state than it was a moment ago or just a bit closer, regardless of whether that goal is accuracy. On the meta-representational view, all these departures count equally as false even though they have significantly different implications for how the GM will respond. The evaluationist view, in contrast, naturally captures these important processing nuances.

Nor does the PE signal operate the way a representation should. No prediction error signal will represent or misrepresent a true first-order representation, because prediction error signals do not encode non-errors. No PE signal encodes "The current visual representation of the location of the light is likely to be true" when the visual representation is true or when it is false. It follows that half the cases in which the signal might represent the first-order representation, including half of those in which it might misrepresent it, are ruled out. What would motivate an explanation of the signal's content in which it does only half the work it should do given that explanation? A sender of representations has no reason to stop sending representations exactly when they would satisfy the aim of representation. The meta-representational view implies that redundant PE signaling should be widespread and normal, not minimal and pathological.[19]

Finally, what is missing from pushmi-pullyu representations is the evaluative element that mediates the indicative and imperative elements. How an organism – including a honeybee – should respond depends on how it evaluates the state of affairs it confronts. The predictive processing framework deconstructs this processing and posits an

---

[19] Millikan (1995: 190) moots an adverbial account of the attitudes, rather than the standard relational account, in which representational content depends essentially on the functional role the representation plays in the system. However, an adverbial structure for propositional attitudes would not turn them into evaluations.

evaluative element in it. In the case of the honeybee's dance, there are two prediction error evaluations going on. The *observing* bees will compare the incoming input of the dance with their predictions about how the dancer will move given how it has been moving, and generate a PE signal if needed (and some may generate misevaluations). This signal may be used to yield true or more accurate GM representations of where the nectar is and which imperative it should obey. The *dancing* bee will compare its predictions of the visual input it should have if it is dancing as it should given its communicative goal, and will adjust its movements should there be a discrepancy (and its PE signals also may misevaluate). This signal may be used to ensure the bee's dance truly or accurately indicates the nectar location. The pushmi-pullyu account may be correct for the dance's content as input to the observing bees and for the dancer's generative model, but it doesn't explain the role of PE signals in these processes. Pushmi-pullyu representations themselves are deconstructed into beliefs and desires in some sophisticated systems. The predictive signaling system account puts these representations in the GM, while the PE signal is the evaluation. [20]

[20] This evaluationist view should not be confused with Bain's (2017) evaluationism regarding the content of phenomenal experience, in which pain experience represents that a body part is damaged *and* that this condition is bad for you. This is a conservative extension of representationalist explanations of experience (e.g., Cutter and Tye 2011; Aydede 2019). Bain's evaluation component is in effect a judgment that *p;* in the predictive coding framework, it is an element of the GM generated in response to tissue damage and is distinct from the PE error signal (see Weise 2016 for a predictive processing account of

Although I do not agree with his interpretation, Shea is a welcome exception in the predictive processing literature, as the nature of prediction error signal content has been largely neglected in favor of that of the generative models. As the identification of intentionality with representation would predict, these models are presumed to be representational and often, if not always, explicitly propositional (Kiefer and Hohwy 2018, 2019; Gladziejewski 2016; Gladziejewski and Milkowski 2017; Williams 2017). Even to Clark, generative models are representation units that encode the system's best hypotheses at its preferred level of description. Generative models are also a new battleground for debate over whether explicit representations are needed in cognitive science, rather than being stored dispositionally or implicitly in processing biases or connection weights (Hutto 2018; Orlandi 2014; Colombo and Series 2012).

I need not commit myself to either side of this debate. Representationalists and anti-representationalists about generative models may be right in distinct circumstances, given the high cost of computing probabilistic representations (Gershman and Daw 2012). But to argue that cognition should not be explained in terms of *representation* does not entail that it should not be defined in terms of *intentional states*. Adaptive systems can presumably exploit probabilistic relationships without creating models or generating predictions that are represented by the system. But they cannot do without subjectively accessible, subjectively relevant, time-sensitive feedback. This internal evaluation may be explicitly encoded if anything is. Perhaps an apt analogy is to a physical system that keeps explicit

pain). Lewis/Skyrms signaling theory has been invoked in relation to pain experience by Martinez and Klein (2016), but they do not invoke Skyrms' theory of content.

track of acceleration, and only tracks speed and direction implicitly. Similarly, prediction error signals may need to be explicitly encoded if anything is, but this is compatible with a dispositional interpretation of the generative models and predictions.

What I am not neutral about is the use of Bayesian language for predictive coding and predictive processing – for example, describing the system as inferring which of the hypotheses in the generative model is most likely, given the input, by passing predictions and prediction errors up and down the processing hierarchy (e.g., Kiefer and Hohwy 2018: 2401). Many take pains to distinguish predictive coding algorithms between subsystems from Bayesian predictive processing by the containing system, or differences between Bayesian reasoning and predictive processing (Aitchison and Lengyel 2017; Orlandi 2018; Rescorla 2017). My complaint is not sophisticated: Bayesian descriptions are simply too impoverished to distinguish signaling from inferring. In adaptive systems, the first involves an efficiently encoded entity being sent from one place to another with the goal of content-preservation through change of place. The latter involves creating a new contentful entity on the basis of at least one other with the goal of truth-preservation through (usually) change of content. These distinct functions involve distinct forms of processing. As a result, when a predictive signaling system is described in Bayesian terms, the description glosses over the fact that the evidence – the prediction error signal – takes a specific form that answers to the requirements of signaling, not those of inferring.

4. Evaluations and traditional philosophy of mind

Why do we identify intentionality, or the aboutness of minds, with representation? There is certainly a long philosophical history, with such exemplary figures as Aristotle and Locke, in which the contents of sensory states are conceptualized in terms of copies (or perhaps structural isomorphisms) of worldly states of affairs. But there is a very explicit embrace and articulation of this identity that comes from extending to philosophy of mind dominant analyses of meaning in philosophy of language. This history explains both the presupposed identification and why we should be skeptical of it as a valid constraint on our thinking about intentionality in cognitive science.

The idea that mental content involves some type of matching relation between mental states and the world is illustrated by the familiar terminology of world-to-mind and mind-to-world fit mentioned above. In Shannonesque terms, the corresponding idea is that the Signal sent to the Receiver by the Transmitter matches the Message in its semantics. This basic assumption grounds claims that the computational task of an organism is to explain how its internal states come to represent the external world or that the task of the sensory subsystem is to infer the external cause (or most probable distal cause) of its inputs, which the Transmitter's signal then represents. As Dretske (1983: 55) voices this received view, "[the] content of a signal is expressed by a sentence describing the condition at the source on which the signal depends in a lawful way." Setting aside his lawfulness requirement, the rest of his description is orthodox, although it doesn't work very well as a description of the content of the PE signal.

Of course, Brentano (1874/2014) made clear that intentionality couldn't be equated with representing *the world*, because the object of thought need not exist. Semantic naturalization projects tend to set aside this difficult problem of intentional inexistence without setting aside the idea that the mind's being *about* something is *representing* that thing. But aboutness may be neutrally characterized as "the relation that meaningful items bear to whatever it is that they are *on* or *of* or that they *address* or *concern*" (Yablo 2014: 1). Something can *concern* or *address* something else without representing it. A theory of intentionality or aboutness is plausibly in part a matter of representing, but representing is only contingently the whole story of intentionality.

The preoccupation with representing in thinking about intentionality is directly connected to longstanding philosophical concerns with knowledge- and truth-preserving reasoning, as Dretske (op.cit.) makes clear. In philosophy of language, this epistemic concern motivated placing truth and reference at the center of explanations of sentence and word meaning. Traditionally, propositions are the abstract entities posited to fill this explanatory role. They are what sentences are used to express, and are the primary bearers of truth and falsity: a sentence's truth condition is encapsulated in the proposition it expresses, and sentences in different languages ("Snow is white", "Der Schnee ist weiss") have the same truth conditions when they express the same proposition (McGrath and Frank 2018).[21]

Crucially, this analysis of linguistic meaning was extended without revision to mental content in the form of the analysis of mental states as propositional attitudes. Attitudes are different types of inferential or computational role for propositions in a cognitive system. As Schiffer (1981) put it, to believe that *p* is to have a sentence that expresses the proposition that *p* tokened in one's "belief box" – a picturesque way of saying that believing that *p* is entertaining a proposition that *p* in a particular way. Desiring that p

[21] The ontology of propositions is disputed, although identifying them with sets of possible worlds dominates (Skyrms adopts this view, albeit with his modifications).

is entertaining a proposition that p in a different way; judging that *p*, expecting that *p*, hoping that *p*, perceiving that *p*, and so on are different ways of entertaining propositions. Given this background theory, initial attempts to naturalize mental content involved naturalizing propositional attitudes (e.g. Field 1978) and commonsense belief/desire psychology was configured as propositional attitude psychology (Fodor 1987).[22]

In brief, according to this philosophical orthodoxy, mental states are particular contexts of manipulation of propositions. An intentional state's content is the semantics of its proposition, and the referential aboutness of propositions (or sentences expressing them) is the representational aboutness of thoughts and their component concepts. Philosophers of mind have moved beyond this orthodoxy in many different ways. Many criticized it even in its heyday, not least to avoid ontological commitment to propositions as abstract objects. Nevertheless, our basic understanding of intentionality seems not to have advanced beyond the original truth- and reference-dominated theories of meaning from which analytic philosophy of mind emerged.

Skyrms is responding to this tradition when he claims that propositional content is a special case of his vector semantics. As he puts it, in traditional views, "the information

---

[22] For example, in his influential proposal, Field (1972, 1978) analyzed "believing that *p*" in terms of a naturalistically acceptable relation (dubbed believes*) to a sentence S, and S means that *p*. Truth is explained in terms of reference (or denotation) to objects and properties, and reference in terms of a causal theory of reference. Jacob (2019) provides an excellent overview of the orthodoxy regarding intentionality; see also Pitt (2020).

content in a signal is conceived as a proposition. The information in a signal is to be expressible as "the proposition that ___" and signals "are thought of as the sorts of things that are either true or false" (2010a: 34). His alternative generalized concept of information content is "non-propositional" in a proprietary sense: it is the special case of a vector in which, given a signal, the probability of at least one state of affairs moves to zero, represented by $\sim\infty$. In the general case, no vector slots contain $\sim\infty$. The dominant ontology of propositions is also maintained, albeit while focusing on the possible worlds in which at least one proposition has been ruled out.

This view of propositional content is one of Skyrms' explicit departures from tradition. If it were his only departure, we might be tempted to express mental content still by using sentences that describe the external world, only now we have one for each vector slot. If we think of the content of these vectors in terms of the posterior probabilities of each sentence's being true after the input is received, we have Godfrey-Smith's intuitively compelling assumption of what Receivers really want. But Skyrms also argues that content consists in how a signal changes the probabilities, and this is what makes his view a Trojan horse as far as the orthodox view of intentionality is concerned. He retains the terminology and ontology of traditional propositions, but implicitly rejects their essential truth-conditional nature and thereby the orthodox identification of intentionality with representation. The integration of signaling systems theory with predictive coding simply makes this rejection explicit. Prediction error signals are non-propositional in Skyrms' explicit proprietary sense as long as the semantic vectors that encapsulate their content have some probabilities ruled out. But they are *also* non-propositional in the sense that the contents encapsulated in the vector slots are not adequately analyzed in terms of traditional propositions. His semantic vectors capture the implicit semantics of communication theory, not the explicit semantics of traditional philosophy of mind and language.

5. A few objections

Resistance to this broadening of the concept of intentionality may be called anti-evaluationism, although this has a stronger and weaker sense. The stronger sense denies that there is a distinct kind of intentionality, affirming the traditional identification of intentionality with representation. The weaker sense accepts that there is a distinct kind of intentional state. In the first case, anti-representationalists will argue as they have been. In the second case, they would draw a representation/evaluation distinction in their own preferred idiom, and can accept evaluations while rejecting representations, or else reject them both. I indicated above reasons to think mental evaluations are explicitly encoded if anything is, but I leave open whether those reasons are good enough. I am not defending representationalism or anti-representationalism. Of importance here are objections to the very idea of mental evaluations as a distinct type of intentional state.

First, one might worry that it leads to regress, on the basis that misevaluation requires evaluating an evaluation. But the possibility of misevaluation does not imply positing evaluations of evaluations. It is one thing to evaluate an intentional state, and another to posit an intentional state that is an evaluation. For example, *we* evaluate a system's representations as true or false without positing an internal state that does this evaluation from the inside. Usually, we infer to this judgment based on its behavior. The evaluationist view posits an intentional state whose content is evaluative. The system does not evaluate this internal evaluation; it (specifically, its generative model) simply responds to it. *We* can also evaluate these evaluations just as we evaluate representations: from the outside. If we observe an exaggerated correction of a reaching movement, we might infer that the PE signal was a misevaluation (a pessimistic one: the reaching hand wasn't *that* far off its target).

A second objection can be derived from Birch's (2014) claim that misrepresentation is not possible on Skyrms' view and so it fails as a theory of representation; if he is right, it may also fail as a theory of evaluation. As Birch argues, if a semantic vector rules out a state of affairs, it can never be sent when that state of affairs obtains. Yet to express false propositional content, the vector must be sent at least once when the state of affairs that it rules out obtains. Since both conditions cannot be satisfied, misrepresentation can never actually happen. So the theory lacks a coherent explanation of misrepresentation.

But Birch's objection is not a problem for evaluations. There is no prediction error signal when there is no difference between expected and actual input. Misevaluation does not require that the signal be sent at least once when there is no difference; more common types of misevaluation are described above. In essence, Birch's criticism of Skyrms is the obverse of my criticism of Shea. Presupposing that semantic vectors are representational, Birch argues that the vectors will never be sent when they would misrepresent. I argued against Shea that if prediction error signals are meta-representational, the fact that they aren't sent when there is no difference entails that they will never represent or misrepresent when a prediction is true. Both objections are resolved if Skyrms' semantic vectors encapsulate evaluative content in predictive signaling systems.

A final objection may be put as follows. Although signaling systems theory has been extended from between-individual signaling to within-individual signaling, how can predictive coding be extended from within-individual communication to between-individual communication? It seems *ad hoc* to deny this possibility given the integrated account. But if it is so extended, the account implies that ordinary linguistic communication involves uttering prediction errors with evaluative content, which seems absurd. Surely philosophical tradition gets right the idea that linguistic Signals are supposed to replicate the semantics of linguistic Messages. The evaluationist view appears to reject this feature of Shannon's theory, which has its home in linguistic communication, after all.

For concreteness, consider the Revere signaling system with two world states and two Signals. The British army's coming by land is the Message S1, and its coming by sea the Message S2, each with .5 probability. Paul Revere's collaborator, Robert Newman, can send a Signal of one lantern (A) if they come by land or two lanterns (B) if they come by sea. Newman observes S2 and sends B. Revere Receives B and acts. Signal B's occurrence changes the probability of S1 to zero and that of S2 to 1; A would have done the opposite. B's content is represented by a two-slot semantic vector in which the value in the slot for S1 (whose probability moves to zero given B) is represented by $\sim\infty$; B's content is propositional in Skyrms' proprietary sense. I've argued that in predictive signaling systems the Signal is a prediction error with evaluative content, and so is non-propositional in the standard truth-conditional sense of proposition. So is Newman a Friston Transmitter? If Signal B were linguistic ("They're coming by sea"), would his utterance be evaluative?

One point is that, intuitively, Newman and Revere are not a single mind even if they cooperate to form a signaling system. Newman can form his own expectations, rejecting those Revere provides in his putative role as generative model. If Newman fully expects S2 to occur and Revere shares that expectation, upon observing S2 Newman would not do anything, and Revere would already be ready for appropriate action. But if Newman's predictions are independent of Revere's, the Revere system would fall apart, as would any processing based on shared predictions. In short, it seems predictive coding is only applicable in signaling systems that are integrated at least to the extent that Transmitters cannot generate their own predictions and ignore or supersede those provided by the generative model. In principle then, the integrated theory might be extended to the level of individuals, but only qua components of fully integrated group minds. If Newman and Revere cooperate sufficiently to form the Revere signaling system but not sufficiently to form a single integrated mind, it would not be *ad hoc* to deny that Newman is a Friston Transmitter.[23]

Even so: suppose the Revere signaling system is sufficiently integrated such that Newman counts as a Friston Transmitter. Surely it doesn't make sense to think "They're coming by sea" expresses an evaluation of the system's readiness in relation to the world, as opposed to a representation whose content matches the world in a suitably generic sense.

[23] In Friston and Frith's (2015) elaboration of communication between two agents using the active inference framework, both agents come to synchronize their generative models and expectations through sensory exchange (two songbirds, in their example). When this integration is achieved, in some sense their agency is not distinct. In the text, this degree of integration between Newman and Revere does not occur, which is true of many stable communication systems. It is an open question the degree to which social insect or bacteria colonies or other multi-agent systems or superorganisms are communicatively integrated.

I think the evaluative view can make sense even in this hypothetical case, but can only gesture at why. Linguistic communication is more than just a series of uttered propositions. *Inter alia*, it involves evaluating interlocutors, contexts, and utterances. Participants in actual linguistic communication plausibly do evaluate the system in which they participate, even as they participate in it. In other words, the objection is really to the idea that evaluation is a *semantic* feature of utterances. This is not a mere verbal dispute over how "semantic" should be used. The claim that evaluation is *not* semantic is a theory-laden answer to the question of how to model linguistic communication. And this theory can be rejected.

Given its epistemic motivations, truth-conditional semantics shunts non-truth-related aspects of communication into pragmatics. For example, in Gricean (1975) communication, the idealized default is that speakers speak truthfully, hearers expect they will, and hearers accept the proposition expressed by the speaker's utterance as true. This idealization excludes evaluation from the semantics of what is said, and at best puts it in the attitudes of hearers and speakers towards each other and their choices of linguistic acts in response to the other, which can become the common ground – part of the world state that is the next speaker's input. Such aspects of communication at best play a pragmatic role of determining which (traditional) proposition an utterance is being used to express in a context. Additional evaluative aspects of language that are left out include prosody and other acoustic features of speech, as well as the affective elements of pejoratives and other terms that are *prima facie* evaluative. All these features of linguistic communication are of little interest from a truth-conditional perspective on meaning.

Shannon made a similar theoretical choice. He deliberately set aside many salient aspects of linguistic communication in order to develop a mathematical theory of it. He and Weaver (1949: 14) held that in communication there is the technical aspect (Level A), the semantic aspect (Level B), and the effectiveness aspect (Level C), which concerns which Message will likely cause the Receiver to respond in the desired manner. His theoretical interest was solely in Level A. But the convenience of this analysis for his purposes did not prevent him from affirming that these levels overlap and have implications for each other. The probabilistic regularities in a natural language that make uncertainty reduction and signaling efficiency at Level A possible depend on Level B, the level of non-natural meaning that arises from the coordinated behavior of language users, as Skyrms argues. Level C, meanwhile, is familiar from Austin's (1962) perlocutionary acts, as well as Gricean communication. Traditional truth-conditional semantics differs from Shannon in that it focuses on Level B rather than A. But also unlike Shannon, it hermetically distinguishes Level B from A and C, in particular ruling anything in C as a matter of pragmatics, not semantics. Our theory-laden intuitions about the semantics of Newman's utterance come from this theoretical perspective.

But it is not the only option. Some theorists already include an evaluative element in visual perception via the concept of an affordance, where assessments of value (affective processes) shape the content of visual representations (Gallagher 2018; Barrett and Bar 2009). This "shaping" may be understood in traditional terms in the same way that contexts are said to constrain the propositions that utterances express. The evaluationist alternative suggests that visual states on this view may just *be* visual evaluations, not representations.

Similarly, it is coherent to think Newman's utterances have evaluative content, given a semantic theory that answers to the non-epistemic goals of communication.

Conclusion

I've offered an account of predictive signaling systems that integrates signaling systems theory, based in Shannon's mathematical theory of communication, with Friston's predictive coding theory of brain function. Much more might be said about both predictive processing and signaling systems, but my goal has been the limited one of showing how their basic commitments can be conceptually synthesized in an illuminating way. A concern with efficiency motivates Shannon's and Friston's accounts of the flow of information. Skyrms' vectors capture how this efficiency is realized in the content of prediction error signals in predictive signaling systems with brains. I've interpreted these signals as mental evaluations, and argued that a complete naturalistic theory of intentionality must account for evaluative as well as representational intentional states. Evaluations have distinct functions, are produced and operate differently, and exhibit their own form of normativity. Evaluationism conflicts with the received view of intentionality taken from philosophy of language, but is a natural fit with contemporary cognitive science.

References

Aitchison, L. and M. Lengyel (2017). With or without you: predictive coding and Bayesian inference in the brain. *Current Opinion in Biology* 46: 219-227.

Anscombe, G.E.M. (1963). *Intention*, 2nd ed. Oxford: Basil Blackwell.

Austin, J.L. (1962). *How to Do Things With Words*. Oxford University Press.

Aydede, M. (2019). Pain. *The Stanford Encyclopedia of Philosophy* (Spring 2019 edition), E. Zalta, ed., URL = https://plato.stanford.edu/archives/spr2019/entries/pain/

Bain, D. (2017). Evaluativist Accounts of Pain's Unpleasantness. In J. Corns, ed., *The Routledge Handbook of Philosophy of Pain*. Routledge.

Barrett, L. and M. Bar (2009). See It With Feeling: Affective predictions during object perception. *Philosophical Transactions of the Royal Society B* 364 (1521): 1325-34.

Birch, J. (2014). Propositional Content in Signaling Systems. *Philosophical Studies* 171: 493-512.

Brentano, F. (1874/2014). *Psychology from an Empirical Standpoint*. Routledge.

Bruineberg, J. and E. Rietveld (2014). Self-Organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience* 8 (599): 1-14.

Chemero, A. (2009) *Radical Embodied Cognitive Science*. MIT Press.

Colombo, M. and P. Series (2012). Bayes in the brain—On Bayesian modelling in neuroscience. *The British Journal for the Philosophy of Science* 63: 697-723.

Clark, A. (2013). Whatever Next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36: 181-253.

Clark, A. (2016). *Surfing Uncertainty: Prediction, action and the embodied mind*. Oxford University Press.

Cutter, B. and M. Tye (2011). Tracking Representationalism and the Painfulness of Pain. *Philosophical Issues* 21 (1): 90-109.

Dennett, D. (1987). *The Intentional Stance*. MIT Press.

Dretske, F. (1981) *Knowledge and the flow of information*. Cambridge, MA: MIT Press.

Dretske, F. (1983). *Precis* of Knowledge and the Flow of Information. *Behavioral and Brain Sciences* 6: 55-90, 82-83.

Eliasmith, C. (2005). Neurosemantics and Categories. In H. Cohen and C. Lefevre, *Handbook of Categorization in Cognitive Science*. Elsevier. pp. 1035-1054.

Field, H. (1972). Tarski's Theory of Truth. *The Journal of Philosophy* 69: 347-375.

Field, H. (1978). Mental representation. *Erkenntnis* 13: 9-61.

Fodor, J. (1987). *Psychosemantics*. Cambridge: MIT Press.

Frege, G. (1892/1948). Sense and reference. *Philosophical Review* (1948) 57: 209-230.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B, Biological Sciences* 360 (1456): 815-836.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience* 11 (2): 127-138.

Friston, K., C. Thornton, and A. Clark (2012). Free-energy minimization and the dark-room problem. *Frontiers in Psychology* vol. 3 article 130.
DOI: 10:339/psyg.201200130.

Friston, K. and C. Frith (2015). Active inference, communication and hermeneutics. *Cortex* 68: 129-143.

Gallagher, S. (2018). Decentering the Brain: Embodied cognition and the critique of neurocentrism and narrow-minded philosophy of mind. *Constructivist Foundations* 14 (1): 8-21.

Gershman, S. and N. Daw (2012). Perception, Action, and Utility: the tangled skein. In M. Rabinovich, K. Friston, and P. Varona, eds., *Principles of Brain Dynamics* (MIT): 293-312.

Godfrey-Smith, P. (2012). Review of *Signals: Evolution, Learning, and Information*, by Brian Skyrms. *Mind* 120 (480): 1288-1297.

Godfrey-Smith, P. (2013). Signals, Icons, and Beliefs. In D. Ryder, J. Kingsbury and K. Williford, eds., *Millikan and Her Critics*. Malden and Oxford: Wiley-Blackwell. 41-58.

Godfrey-Smith, P. (2014). Sender-receiver Systems Within and Between Organisms. *Philosophy of Science* 81: 866-878.

Gladziejewski, P. (2016). Predictive coding and representationalism. *Synthese* 193: 559-582.

Gladziejewski, P. and M. Milkowski (2017). Structural representations: causally relevant and different from detectors. *Biology and Philosophy* 32: 337-355.

Grice, H.P. (1957). Meaning. *The Philosophical Review* 66 (3): 377-388.

Grice, H.P. (1975). Logic and conversation. In P. Cole and J.L. Morgan, eds., *Syntax and Semantics* vol. 3 (New York: Academic Press): 41-58.

Harms, W. (2004). Primitive content, translation, and the evolution of meaning in animal communication. In D.K. Oller and U. Griebel, eds., *Evolution of Communication Systems: A comparative approach* (MIT): 31-48.

Hohwy, J. (2014). *The Predictive Mind*. Oxford: Oxford University Press.

Hutto, D. and E. Myin (2017). *Evolving Enactivism: Basic minds meet content.* MIT Press.

Hutto, D. (2018). Getting into predictive processing's great guessing game: Bootstrap heaven or hell? *Synthese* 195: 2445-2458.

Isaac, A. (2019). The Semantics Latent in Shannon Information. *The British Journal for the Philosophy of Science* 70 (1): 103-125.

Jacob, P. (2019). Intentionality. *The Stanford Encyclopedia of Philosophy* (Winter 2019 edition), E. Zalta, ed.,
URL = https://plato.stanford.edu/archives/win2019/entries/intentionality/

Kiefer, A. and J. Hohwy (2018). Content and misrepresentation in hierarchical generative models. *Synthese* 195: 2387-2415.

Kiefer, A. and J. Hohwy (2019). Representation in the Prediction Error Minimization Framework. In S. Robins, J. Symons, and P. Calvo, eds., *Routledge Companion to the Philosophy of Psychology*, 2nd ed (vol. 2). London: p. 384-409.

Kripke, S. (1979). A puzzle about belief. In A. Margalit, ed., *Meaning and Use* (Dordrecht: Reidel): 239-283.

Lean, O. (2014). Getting the Most Out of Shannon Information. *Biology and Philosophy* 29: 395-413.

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

Lombardi, O. (2005). Dretske, Shannon's Theory, and the Interpretation of Information. *Synthese* 144 (1): 23-39.

Lombardi, O., F. Holik, and L. Vanni (2016). What is Shannon information? *Synthese* 193: 1983-2012.

Mackay, D. (1969). *Information, Mechanism, and Meaning*. Cambridge, MA and London: MIT Press.

Martinez, M. (2018). Representations are Rate-Distortion Sweet Spots. *Proceedings of the Philosophy of Science Association* (PSA2018). Pre-print.

Martinez, M. and C. Klein (2016). Pain Signals are Predominantly Imperative. *Biology and Philosophy* 31: 283-298.

Meinong, A. (1904). Uber Gegenstandtheorie (English translation: The Theory of Objects). In R. Chisholm, ed. *Realism and the Background of Phenomenology*. Glencoe: The Free Press, 1960.

Millikan, R. (1984). *Language, Thought, and other Biological Categories*. MIT Press.

Millikan, R. (1989). Biosemantics. *Journal of Philosophy* 86: 281-97.

Millikan, R. (1995). Pushmi-pullyu representations. *Philosophical Perspectives* 9: 185-200.

McGrath, M. and D. Frank (2018). Propositions. *The Stanford Encyclopedia of Philosophy* (Spring 2018 edition), E. Zalta, ed., URL = https://plato.stanford.edu/archives/spr2018/entries/propositions/

Orlandi, N. (2014). *The innocent eye: Why vision is not a cognitive process* (Oxford: Oxford University Press).

Orlandi, N. (2018). Predictive perceptual systems. *Synthese* 195: 2367-2386.

Pitt, D. (2020). Mental Representation. *The Stanford Encyclopedia of Philosophy* (Spring 2020 edition), E. Zalta, ed., URL = https://plato.stanford.edu/archives/spr2020/entries/mental-representation/

Prosser, A., K. Friston, N. Bakker and T. Parr (2018). A Bayesian Model of Psychopathy: A Model of Lacks Remorse and Self-Aggrandizing. *Computational Psychiatry* 2: 92-140.

Piccinini, G. and A. Scarantino (2011). Information processing, computation, and cognition. *Journal of Biological Physics* 37: 1-38.

Ramsey, W. (2007). *Representation Reconsidered*. Cambridge University Press.

Rao, R. and D. Ballard (1999). Predictive coding in the visual cortex. *Nature Neuroscience* 2: 79-87.

Recanati, F. (2001). What is Said. *Synthese* 128 (1/2): 75-91.

Rescorla, M. (2017). Review of Andy Clark, Surfing Uncertainty: Prediction, Action, and the Embodied Mind. *Notre Dame Philosophical Reviews*.

https://ndpr.nd.edu/news/surfing-uncertainty-prediction-action-and-the-embodied-mind/

Sayre, K. (1983). Some untoward consequences of Dretske's "causal theory" of information. *Behavioral and Brain Sciences* 6: 78-79.

Scarantino, A. (2015). Information as a Probabilistic Difference Maker. *Australasian Journal of Philosophy* 93 (3): 419-443.

Scarantino, A. and G. Piccinini (2010). Information without Truth. *Metaphilosophy* 41 (3): 313-330.

Schiffer, S. (1981). Truth and the Theory of Content. In H. Parret and J. Bouveresse, eds., *Meaning and Understanding.* de Gruyter. pp. 204-224.

Sengupta, Semmler, and Friston (2013). Information and efficiency in a nervous system – a synthesis. *PLoS Computational Biology* 9 (7): e1003157, 1-12.

Shannon, C. and W. Weaver (1949). *The Mathematical Theory of Communication.* Urbana, IL: University of Illinois Press.

Shannon, C. (1948). A Mathematical Theory of Communication. *The Bell System Mathematical Journal* 27: 379-423.

Shea, N. (2007). Consumers need information: supplementing teleosemantics with an input condition. *Philosophy and Phenomenological Research* 75 (2): 404-435.

Shea, N. (2012). Reward prediction errors are meta-representational. *Nous* 48 (2): 314-41.

Shea, N. (2014). Neural signaling of probabilistic vectors. *Philosophy of Science* 81: 902-913.

Shea, N., P. Godfrey-Smith, and R. Cao (2017). Content in Simple Signaling Systems. *British Journal for the Philosophy of Science*. DOI: 10.1093/bjps/axw036.

Skyrms, B. (2010a). *Signals: Evolution, Learning, and Information.* Oxford: Oxford University Press.

Skyrms, B. (2010b). The Flow of Information in Signaling Games. *Philosophical Studies* 147: 155-165.

Soni, J. and R. Goodman (2017). *A Mind at Play: How Claude Shannon invented the information age.* New York: Simon and Schuster.

Spratling, M. (2017). A review of predictive coding algorithms. *Brain and Cognition* 112: 92-97.

Spratling, M. (2016). Predictive coding as a model of cognition. *Cognitive Processing* 17 (3): 279-305.

Sprevak, M. (2019). Two Kinds of Information Processing in Cognition. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-019-00438-9

Stegmann, U. (2015). Prospects for Probabilistic Theories of Natural Information. *Erkenntnis* 80: 869-893.

Stich, S. (1983). *From Folk Psychology to Cognitive Science: The case against belief*. MIT Press.

Usher, M. (2001). A Statistical-Referential Theory of Content: Using information theory to account for misrepresentation. *Mind & Language* 16 (3): 311-334.

Weaver, W. (1949) Recent Contributions to the Mathematical Theory of Communication. *The Mathematical Theory of Communication:* 95-117.

Wiech, K. (2016). Deconstructing the sensation of pain: The influence of cognitive processes on pain sensation. *Science* 354 (6312): 584-587.

Williams, D. (2017). Predictive processing and the representation wars. *Minds & Machines*. DOI: 10.1007/s11023-017-9441-6.

Yablo, S. (2014). *Aboutness*. Princeton University Press.