What are the contents of representations in predictive processing?

Wanja Wiese (wawiese@uni-mainz.de)

(Penultimate draft. The final version is available at:
http://link.springer.com/article/10.1007%2Fs11097-016-9472-0)

Department of Philosophy
Johannes Gutenberg
University
Mainz, Germany

Abstract

Paweł Gładziejewski has recently argued that the framework of predictive processing (PP) postulates genuine representations. His focus is on establishing that certain structures posited by PP actually play a representational *role*. The goal of this paper is to promote this discussion by exploring the *contents* of representations posited by PP. Gładziejewski already points out that structural theories of representational content can successfully be applied to PP. Here, I propose to make the treatment slightly more rigorous by invoking Frances Egan's distinction between mathematical and cognitive contents. Applying this distinction to representational contents in PP, I first show that cognitive contents in PP are (partly) determined by mathematical contents, at least in the sense that computational descriptions in PP put constraints on ascriptions of cognitive contents. After that, I explore to what extent these constraints are specific (i.e., whether PP puts unique constraints on ascriptions of cognitive contents). I argue that the general mathematical contents posited by PP do not constrain ascriptions of cognitive content in a specific way (because they are not relevantly different from mathematical contents entailed by, for instance, emulators in Rick Grush's emulation theory). However, there are at least three aspects of PP that constrain ascriptions of cognitive contents in more specific ways: (i) formal PP models posit specific mathematical contents that define more specific constraints; (ii) PP entails claims about how computational mechanisms underpin cognitive phenomena (e.g. attention); (iii) the processing hierarchy posited by PP goes along with more specific constraints.

Keywords: Bayesian inference, Cognitive contents, Mathematical contents, Predictive coding, Predictive processing, Structural representation.

Acknowledgments

I am grateful to Michael Madary and Thomas Metzinger for a number of very helpful comments on drafts of this paper.

What are the contents of representations in predictive processing?

Wanja Wiese

Introduction

Predictive processing[1] (PP, cf. Clark 2016; Hohwy 2013b) is a framework in cognitive and theoretical neuroscience that is becoming increasingly influential. As Paweł Gładziejewski (2015) has recently argued, PP posits structures that fulfill a robust representational role. Here, I build on Gładziejewski's work to explore the question: What is the *content* of representations posited by PP? My background assumption in this paper is that Gładziejewski's treatment is on the right track and that PP posits structures that play a genuine representational role, thereby meeting what William Ramsey calls the *job description challenge* (cf. Ramsey 2007). In order to explore the *contents* of these representations, I shall first clarify the core concepts of PP and its relation to Bayesian inference (section 1). After that, I will consider Frances Egan's helpful distinction between mathematical and cognitive contents (section 2). This will allow us to formulate the core question of this paper more succinctly, by dividing it into two sub-questions: (i) What are the mathematical contents of representations posited by PP? (ii) What are the cognitive content of representations posited by PP? The first question can be answered in a relatively general way. I suggest to approach the second question by focusing on the relation between mathematical and cognitive contents in PP, and by considering what constraints the core aspects of PP put on ascriptions of cognitive contents (sections 3 - 5).

1 Basic tenets of predictive processing

The core idea of predictive processing (PP; see Clark 2013b, 2016; Hohwy 2013b) is that the brain does not process sensory signals in a purely bottom-up fashion, but that it also forms top-down predictions of sensory signals, based on estimates derived from a hierarchical model. This model tracks statistical regularities at different levels of spatial and temporal grain (cf. Kiebel, Daunizeau, Friston, & Sporns 2008). Top-down predictions are compared to bottom-up (sensory) signals, a prediction error is computed, and the computed error is used to update the model (and generate better predictions). Crucially, this cycle, consisting of computing estimates, deriving predictions, computing prediction errors, updating estimates etc., is not just performed at the level of sensory inputs, but at all levels of a processing hierarchy. See figure 1 for an illustration. This figure captures the essential components of PP accounts (cf. Clark 2015):

1. estimates (first-order statistics),
2. predictions,

[1] I am using the term "predictive processing" in the sense of Clark (2013b, p. 202/fn. 5). Predictive processing entails hierarchical, approximately Bayesian inference using prediction error minimization (cf. also section 1).

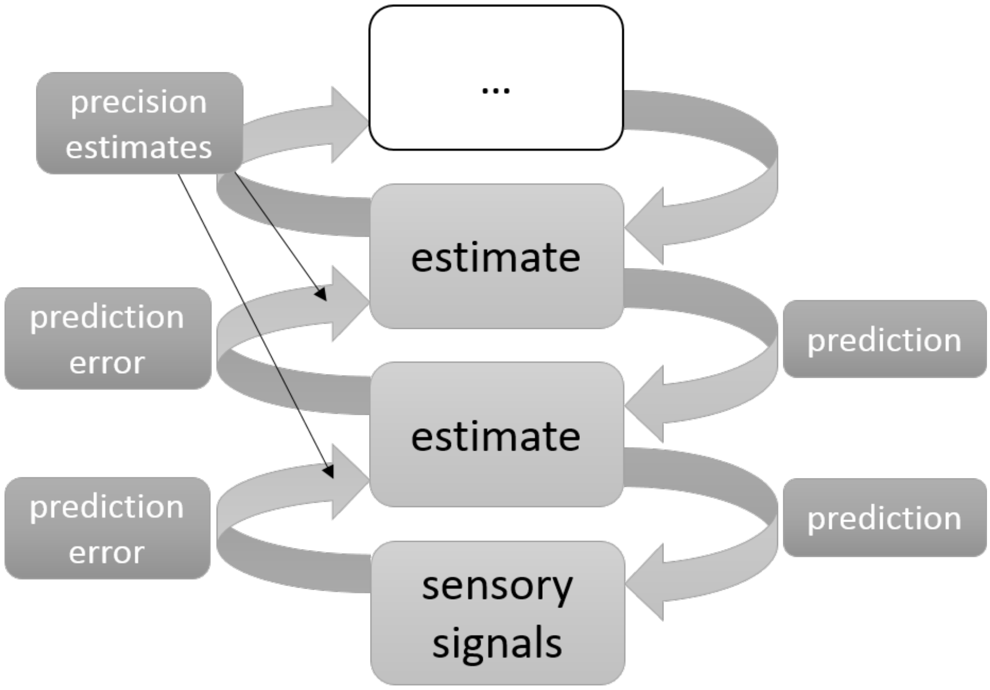

Figure 1: Core components of PP accounts: estimates (first-order statistics; generated at all levels of a hierarchy), predictions (derived from these estimates), prediction errors (resulting from a comparison between predictions and signals at the level below), and precision estimates (second-order statistics, which modulate error signals).

3. prediction errors, and

4. precision estimates (second-order statistics).

Another important feature is the above mentioned hierarchical organization. However, when it comes to the question: "What are the contents of representations in PP?", the four core components are the ones that are primarily relevant. Before discussing this question, I shall clarify how predictions are computed and how computations in PP are related to Bayesian inference, as this will become relevant below.

How are predictions computed? The basic idea is that sensory signals can be described as a function of hidden causes (which are external to the organism). So, given an estimate of hidden causes, and an estimate of the functional relation between hidden causes and sensory signals, we can compute a prediction. The general assumption is that sensory signals can be written as a deterministic function of hidden causes, plus a noise term (which captures all *unpredictable* aspects of the functional relation). This gives us the following equation (which is oversimplified, to convey only the essential idea; for a more sophisticated version, see, e.g., Friston 2008, p. 2):

$$\text{sensory signals} = \text{deterministic function of hidden causes} + \text{noise}$$
$$s = g(c) + \omega.$$

Now $g(c)$ can be used as a *prediction* of s, and the precision (inverse covariance) of ω specifies how reliable such predictions will be. This is important, because when the prediction error $|s - g(c)|$ (or the squared prediction error) is computed, the precision can give us a hint as to what extent the error was to be expected, and how strongly it should influence updates of our estimate of c. (Of course, we can in addition also update our estimate of g to improve the model.)

Estimators of s, c and g entail a generative model (which models how the world's hidden states generate the signals received by the brain). A generative model is usually given in the form of a *likelihood* (of sensory signals, giving hidden causes) and a *prior* distribution (of hidden causes, cf. Friston 2010, p. 129). In principle, these can be used to perform Bayesian inference, to infer the hidden causes of sensory signals. In practice, this typically involves complex computations that have to be approximated.

The important point to note here is that PP is, as Andy Clark points out, "broadly Bayesian" (p. 196 Clark, 2013b), but computations in PP do not necessarily involve representations of likelihoods and prior probabilities. To illustrate, I shall briefly show that not even textbook Bayesian inference requires coding the values of likelihoods and prior distributions.

1.1 Bayesian inference

Bayesian inference is a computational strategy to process information in the presence of uncertainty. For instance, if I know that a health test is not completely reliable (yielding false positives in some cases), and if I know that the disease I have been tested for is quite rare, how certain should I be that the (positive) test result is correct? Or: If I have two sources of information, each inflicted with noise, how should I combine signals from these sources to obtain an estimate that is optimal in some sense? Bayesian inference can provide (rationally justified) answers to these and similar questions.

Slightly more formally, Bayesian inference combines *prior* knowledge (or prior uncertainty) $p(h)$ with the likelihood $p(e|h)$ of the evidence (i.e., data) e (given hypothesis h). It rests on Bayes' rule, which specifies how propabilities (or probability distributions or density functions) have to be combined to yield the *posterior* probability $p(h|e)$ of a hypothesis, given the evidence. Having computed the posterior, one can then settle on a single hypothesis (like the hypothesis that maximizes either the posterior or some utility function, cf. Rescorla 2015).

Note that, even given that it is *possible* to describe perceptual phenomena in terms of Bayesian inference, this does not mean that the brain actually implements Bayesian inference. The observed behavior and reported phenomenology can also be underpinned by different computational processes,[2] or by approximations to exact Bayesian inference. PP is only

[2] Bowers and Davis (2012) draw a distinction between three types of Bayesian theorizing (extreme, methodological and theoretical), emphasizing that using Bayesian models in cognitive science does not entail a commitment to the assumption that the brain literally implements Bayesian inference (which the authors see as a weakness, as this can lead to "Bayesian just-so stories"). Cf. also Clark (2013b, p. 189).

committed to this weaker position. In particular, computational descriptions in PP models need not posit variables coding probability values, values of likelihoods, or of other (conditional) probability density functions. In fact, not even textbook Bayesian inference requires computing probabilities in the sense just mentioned.

Here is an example (taken from Kvam & Vidakovic 2007, p. 50) in which the prior and likelihood are Gaussian, and the goal is to update the estimate of the prior mean, in the light of *n* samples (which constitute the evidence). This is the formula for the update (i.e., for the a posteriori estimate; see figure 2 for an explanation):

$$\hat{h}_n = \frac{\tau^2}{\frac{\sigma^2}{n} + \tau^2}\bar{e}_n + \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n} + \tau^2}\mu. \qquad (1)$$

This formula contains parameters (τ^2,σ^2 and μ; μ is also our prior estimate of the quantity we are interested in), a term for the average of the obtained samples ($\bar{e}_n$), and one for the number of samples (n). The formula does not contain terms for probability values or for the values of a probability density function. To implement this special case of Bayesian inference, you only need variables for the parameters, the evidence, and the number of samples. Combining those in the manner specified above gives you the desired updated estimate. You do not have to represent probabilities.

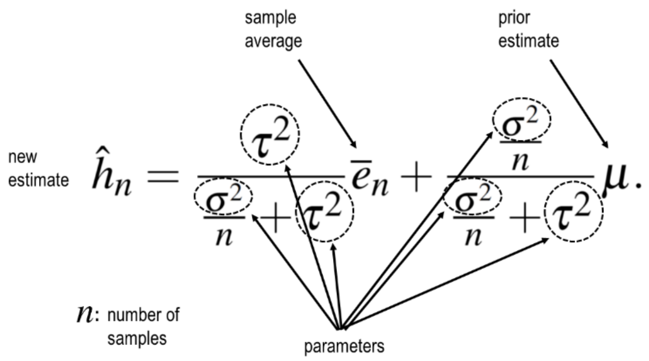

Figure 2: Equation 1 involves two parameters, τ^2 and σ^2, which code the variances of the prior and likelihood distribution, respectively (more specifically, the variance of the likelihood is given by σ^2/n in this example). The term n just codes the number of samples, μ is the prior estimate of the mean, and $\bar{e}_n$ is the sample average. There is a sense in which this involves a representation of probabilities: μ and τ^2 can be used to represent a Gaussian probability distribution. For instance, if I know the general formula of the univariate Gaussian probability density function, I can plug μ and τ^2 into the formula. However, μ and τ^2 themselves do not represent values of probabilities or densities. Neither are they used to compute such values in the update formula. In the formula, the parameters only serve to compute a weighted average of the sample average and the prior estimate of the mean. Still, this is an instance of Bayesian inference.

Note that there is still a sense in which the above formula *presupposes* a generative model (with prior and likelihood). For the formula can only be derived given the assumption that prior and likelihood have a certain form. But in order to implement the inference scheme that is based on these assumptions, you don't need variables coding probability values (or values of density functions). Let us call the problem of determining how the brain implements (an approximation to) Bayesian inference the *probability conundrum* (**ProC**). **ProC** is currently an open problem (although different *possible* solutions can be found in the literature). As we have seen, probabilistic computations (like Bayesian inference) do not require representations of probability values (or values of density functions), so postulating that representations in PP represent such mathematical objects constitutes a currently unjustified presupposition, and can lead to wrong conclusions about the contents of representations posited by PP (more on this in section 3 below). Since PP is not committed to specific implementations of (or approximations to) Bayesian inference it is, *a fortiori*, not committed to positing representations of probabilities. However, for the purpose of clarification, let me briefly indicate how Bayesian inference is approximated in *some* PP models, to show that PP is not just not committed to representations of probability values, but that there actually are computational PP models in which Bayesian inference is approximated without such representations.[3]

1.2 Approximative Bayesian inference in predictive processing

Recall from above that the problem of perception consists, according to PP, in inferring the hidden causes of sensory signals. Denote the hidden causes with c and the sensory signals with s. Given a generative model $p(s,c) = p(s|c)p(c)$, we can now in principle compute the posterior $p(c|s) = p(s|c)p(c)/p(s)$, and select the value of c that maximizes the posterior (or that maximizes a utility function). In general, however, this is computationally infeasible. But if we know something about the way in which hidden causes (c) are *functionally* related to sensory signals (in the mathematical sense), we can predict sensory signals based on an estimate of the hidden causes (cf. section 1), and by minimizing the prediction error, we can optimize our estimate (and our model).

Crucially, minimizing the prediction error (with respect to c) is in many cases equivalent to maximizing the posterior probability $p(c|s)$. Furthermore, minimization problems can typically be solved approximately using numerical methods (like gradient descent methods). An example of this general strategy can be found in Rao and Ballard (1999). Other examples of approximative Bayesian inference include the method of *Variational Bayes* (for an overview, see Murphy 2012, ch. 21).

According to Karl Friston's *free energy principle* (Friston, 2010), neural activity can be described as a computational process in which free energy is minimized. Free energy bounds the *surprisal*[4] of sensory signals from above; so minimizing free energy implicitly reduces surprisal, which means that the organism will tend to enter sensory states corresponding to situations in which it can be expected to be found, given the type of organism it is; for instance, a fish tends to be found in the water (see Friston 2010, p. 127). Minimizing free energy with respect to an internally represented probability density function, the *recognition density* (cf. ibid., p. 128), entails that the recognition density approximates the posterior density. Free energy minimization is therefore a computational strategy to approximate Bayesian inference. Furthermore, under the assumption that the generative model is Gaussian, the recognition density can be coded by its mean and covariance (see, e.g., Friston 2009, 2010; Friston & Kiebel 2009). In particular, this means that no probability values or density

[3] As an anonymous referee helpfully pointed out, there are also more general consideration regarding computational (in)tractability which provide a reason to think that explicit representations of probability distributions are not only not required, but also very unlikely to be actually employed by the brain (see, e.g. Kwisthout & van Rooij 2013).
[4] "Surprisal" denotes the negative logarithm of the probability of an event (cf. Friston 2010, p. 128).

functions have to be computed. So neither textbook Bayesian inference nor approximate Bayesian inference in actual PP models *require* representing probability values (or values of probability density functions). Before considering what this means for the *contents* of representations in PP, we need to distinguish between two types of representational content.

2 Frances Egan on mathematical and cognitive contents

According to PP, the brain tries to infer the hidden causes of sensory signals. This suggests that the brain represents external events. At the same time, computational models in PP involve random variables and other mathematical functions, which suggests that the brain computes the functions described by those models, and that neural states represent values of variables (or numbers, vectors, and matrices). To clarify how neural representations can be representations of mathematical objects, and also of objects in the world, it will be useful to apply Frances Egan's (2014) distinction between *mathematical* and *cognitive* contents. This will enable a clearer description of representational contents in PP (section 3), and also of the relation between mathematical and cognitive contents (sections 4 and 5).

Given a computational model of a system (i.e., sets of mathematical equations allowing us to describe and predict the system's behavior), we can ascribe mathematical contents to vehicles. A simple example, given by Egan, is a cash register. Such a device can be described as an adding system: given inputs $x_1, x_2, \ldots, x_n$, it computes the function $f(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \cdots + x_n$. This function not only captures the system's input-output behavior, it also provides a (coarse-grained) description of the mechanism that allows the cash register to function as an adding system. This means that numbers in the mathematical description can be mapped to state types of the system. As Egan puts it: "Whenever the system goes into the physical state specified under the mapping as $\underline{n}$, and then goes into the physical state specified under the mapping as $\underline{m}$, it is caused to go into the physical state specified under the mapping as $\underline{n+m}$." (Egan 2014, p. 116).

A computational description of a system can help us understand how a system performs a certain task. In the example of the cash register, it helps us understand how it manages to indicate the total amount a customer has to pay: it simply adds the individual prices of the obtained goods. When a customer buys an apple for 1 €, and a pear for 1 €, the register enters two states that represent 1, and is then caused to enter a state that represents 2. These are the mathematical contents carried by the representational vehicles. In general, "the inputs of a computationally characterized mechanism represent the *arguments* and the outputs the *values* of the mathematical function that canonically specifies the task executed by the mechanism." (Egan 2014, p. 123).

Crucially, we can also say that the first state represents the price of the apple, the second state represents the price of the pear, and the third state represents the total amount that is due. These contents are not mathematical contents. According to Egan, such contents can only be specified relative to an explanatory context (note that in the example, states with the same mathematical content can represent the prices of different goods). In cognitive accounts, such content ascriptions are part of an "intentional gloss", as Egan calls it; to distinguish such contents from mathematical contents, she calls them *cognitive* contents:

> Cognitive contents, on the other hand, are *not* part of the essential characterization of the device, and are not fruitfully regarded as part of the computational theory proper. They are ascribed to facilitate the explanation of the cognitive capacity in question and are sensitive to a host of pragmatic considerations, as explained above. Hence, they form what I call an *intentional gloss*, a gloss that shows, in a perspicuous way, how the computational/mathematical theory manages to explain the intentionally-described explanandum with which we began and which it is the job of the theory to explain. (Egan 2014, p. 128)

Egan introduces cognitive contents as instrumentally ascribed contents, which cannot be derived from mathematical contents and can be relative to our explanatory interests. I do not endorse instrumentalism regarding cognitive contents, and below (in section 2.2) I will argue that there are no *principled* reasons to assume that cognitive contents cannot be naturalized *in any case*. The distinction between mathematical and cognitive contents is a conceptual distinction. In itself, it does not carry strong metaphysical commitments.

2.1 Mathematical and cognitive contents in PP

In PP accounts, we can find examples of ascriptions of mathematical contents, but also of cognitive contents. For instance, in a seminal paper by Rao and Ballard (1999), we find the following passage (in which parts of a hierarchical predictive coding model are described):

> In terms of a neural network, the coefficients r_j correspond to the activities or firing rates of neurons, whereas the basis vectors U_j correspond to the synaptic weights of neurons. The function $f(x)$ is the neuronal activation function, typically a sigmoidal function such as $\tanh(x)$. The coefficients r_j can be regarded as the network's internal representation of the spatial characteristics of the image I, as interpreted using the internal model defined by the basis vectors U_j. (Rao & Ballard 1999, pp. 85 f.)

Applying Egan's terminology, we can note the following about the quoted passage:

1. It is assumed that the mathematical descriptions provided by the model can be mapped to neurobiological descriptions ("the coefficients r_j correspond to the activities or firing rates of neurons").
2. Neural states and processes can thus be regarded as representations with mathematical contents; these are the values of coefficients and vectors (and, in general, of the arguments and outputs of mathematical functions).
3. At the same time, they can also be regarded as representations with cognitive contents (e.g., "of the spatial characteristics of the image").

In general, we can say that estimates at different levels of the PP hierarchy are used as arguments of mathematical functions, which output predictions. These predictions are also used as arguments of functions (together with estimates from other levels), and the output (a prediction error) becomes an argument of a further function (together with a precision estimate) which outputs an updated estimate. (First-order) estimates, predictions, prediction errors, and (second-order) precision estimates are thus among the core *mathematical* contents of representations in PP. Other mathematical contents (like probability values) are optional. The next question then becomes what the core cognitive contents in PP are, or what we can say about the *relation between* mathematical and cognitive contents of representations posited by PP.

2.2 The distinction between cognitive and mathematical contents does not entail a commitment to instrumentalism regarding cognitive contents

Egan herself is an instrumentalist about cognitive contents: "*The structures posited by the computational theory, what we are calling the 'representational vehicles', do not have their cognitive contents essentially.* [...] And the various pragmatic considerations cited above might motivate the assignment of different cognitive contents to the structures." (Egan 2014, p. 127). Two structures that have the same computational description can thus have different cognitive contents. What about PP models? Do the same models always license the assignment of different cognitive contents? At first sight, one might be inclined to answer in the affirmative, since this seems to be part of what makes PP so elegant: The same basic computational principle (prediction error minimization), can account for a host of different (cognitive) phenomena. As Jakob Hohwy puts it:

> [T]his idea [that the brain minimizes prediction error] explains not just that we perceive but *how* we perceive: the idea applies directly to key aspects of the phenomenology of perception. Moreover, it is *only* this idea that is needed to explain these aspects of perception. (Hohwy 2013b, p. 1)

So it seems that relatively general PP models are compatible with different ascriptions of cognitive contents. However, lest ascriptions of cognitive content become arbitrary, they must at least be *constrained* by mathematical contents entailed by computational models. In the example of the cash register, for instance, we cannot say that the register indicates the amount that is due by doubling the price of the first item (although it fits the register's behavior in cases in which exactly two items, with the same price, are bought); for this contradicts the mathematical description (according to which the register computes a sum). This, however, is a very general kind of constraint, and the reason why it is not more specific is that the computational model is extremely simple. The more complex a computational model, the more the mathematical contents constrain the set of possible ascriptions of cognitive contents (which are compatible with the computational description).

In principle, a computational model could be so complex and specific as to allow only a very limited set of cognitive content ascriptions. So there is a continuum between computational models putting more or less specific constraints on ascriptions of cognitive contents. At one extreme (when a computational model is maximally complex and specific), it might even be possible to *derive* cognitive contents from mathematical contents. At the other extreme, we find computational models which are so liberal regarding assignments of cognitive contents that they will be more or less useless. The point I want to make here is not that this proves Egan wrong (i.e., that one cannot be an instrumentalist about cognitive contents), but

1. that it is *possible* to be a realist about cognitive contents, even if one accepts the conceptual distinction between mathematical and cognitive contents, and
2. that instrumentalists and realists alike are committed to the view that a model's mathematical contents must constrain the set of cognitive contents we can coherently ascribe to the system.

These two assumptions can even be accepted by all those who wish to remain neutral on the question whether cognitive contents can be naturalized, or on the question by what factors they are determined. For instance, one could also deny that cognitive contents could ever be *derived* from a model's mathematical contents, but still hold that there are other factors which uniquely determine cognitive contents, and which are not relative to our explanatory interests.

Currently, most PP models are relatively general and simple, and therefore provide only relatively weak constraints on ascriptions of cognitive contents. Hence, it is (at least currently) impossible to *derive* cognitive contents from a model's mathematical contents. So rather than yielding a cognitive explanation (i.e., an explanation positing cognitive contents) *ex novo*, computational models can at most support (independently posited) cognitive explanations. Now an interesting question is: in virtue of what do PP models support this or that cognitive explanation? Do they support an explanation in virtue of their essential components (like prediction errors, precision estimates, or the processing hierarchy)? Or do we have to add components which are not essential to predictive processing models? Although offering a definition of a model's explanatory "support" is beyond the scope of this paper, let me point out that there is an interesting relation between the degree of support a computational model lends to a cognitive explanation and between the specificity of constraints a models puts on assignments of cognitive contents. A model supports a cognitive explanation only if it is compatible with that explanation. If a model is compatible with a large number of different cognitive explanations, it provides less support (for any particular explanation) than a model which is compatible with a small number of cognitive explanations. So the more specific a model is, the more support it provides for cognitive explanations (with which it is compatible). But this is just another way of saying: the more constraints a model puts on assignments of cognitive contents, the more it supports particular cognitive explanations.[5] This observation will again become relevant in section 5.

To sum up: By accepting Egan's helpful conceptual distinction between mathematical and cognitive contents, one is not forced to accept instrumentalism regarding cognitive contents. In fact, for the purposes of this paper, I will remain neutral on the question whether one can or should assume that cognitive contents can be naturalized.

[5] I am grateful to an anonymous reviewer for pressing me to clarify this point.

All I will be assuming is that mathematical contents constrain ascriptions of cognitive contents, and the more constraints a computational model provides, the greater the support it lends to particular cognitive explanations. To further clarify the relation between mathematical and cognitive content in PP, let me next review Gładziejewski's (2015) insightful treatment of representations in predictive processing.

3 Paweł Gładziejewski on structural representations in PP

Paweł Gładziejewski's (2015) discussion pursues two main goals: the first is to show that PP posits *structural representations* (in the sense of O'Brien & Opie 2004), the second is to show that the entities described as representations in PP actually play a representational role, thereby meeting what William Ramsey (2007) calls the *job description challenge*. In passing, Gładziejewski also makes some remarks about the *contents* of representations posited by PP. However, he admits that this "treatment is obviously very sketchy" (Gładziejewski 2015, p. 14), which is fine, since he is not primarily concerned with the contents of representations in PP, but with their functional role. On the other hand, if hierarchical models in PP are structural representations, this means that contents carried by the representational vehicles are (at least partly) determined by their structure. Here, I shall assume that Gładziejewski is correct and that models in PP are in fact structural representations. As we will see, this structure is determined by *mathematical* contents. This already suggests a relevant role for mathematical contents in constraining the cognitive contents we can ascribe to representations posited by PP.

So what is the structure of models posited by PP, according to Gładziejewski? He claims that the structure of the models resembles the "causal-probabilistic"[6] structure of the world in at least three respects:

> **First**, hidden or latent variables in the model encode the probability distributions of obtaining lower-level observations. In other words, different variables correspond to different *likelihoods* of potential patterns of activity at the lower, sensory level. [...] Worldly causes are thus represented in terms of the likelihoods of producing different sensory patterns in the system. [...]
> **Second**, the hidden variables are not only related to lower-level, sensory patterns, but to each other (intra-level) as well. Their values evolve over time in mutually-interdependent ways, which are defined by the model parameters. This is how the *dynamics* of causal-probabilistic relations between worldly objects are encoded in the model. [...]
> **Third**, since our system is supposed to realize Bayesian reasoning, there is one other aspect that the model structure and the environment structure should have in common. Namely, *prior* probabilities of worldly causes should be encoded in the generative model [...]. (Gładziejewski 2015, pp. 13 f.; bold emphasis added)

So the three aspects that determine the structure of models in PP, as suggested by Gładziejewski, are: (1) likelihoods, (2) dynamic relations, and (3) prior probabilities. Let us consider these three parts of the proposal in turn.

(1) According to the passage quoted above, computational PP models contain variables that represent likelihoods. Using the notation introduced above, this would mean that worldly (hidden) causes c are represented in terms of the likelihoods $p(s|c)$ of obtaining sensory signals s. This is true in cases in which Bayesian inference

[6] N.B.: Gładziejewski does not explicitly say what he means by the "causal-probabilistic" structure. For instance, it is unclear whether he means that events in the world cause each other probabilistically (although there is one passage which seems to point in this direction, see Gładziejewski 2015, pp. 12 f.), or whether he assumes there is principled randomness of some sort etc. The most charitable reading is arguably that internal model capture the causal structure of the world; due to noise and partial lack of information (which brings about uncertainty), this model has to be probabilistic; whether the causal structure of the world is itself also probabilistic in some sense is left open; hence the term "causal-probabilistic".

is implemented using the exact computations derived from Bayes' rule (for then, the representation of the posterior always involves a representation of the likelihood). However, as we saw in the discussion above, this need not be the case. We can update our current estimate of c in the light of new evidence without computing likelihoods (and the updated estimate can, for instance, still approximate the maximizer of the posterior distribution). What Gładziejewski arguably means is that the brain represents, according to PP, *functional relations* between different variables (of course, also likelihoods describe functional relations between variables, but a relevant aspect of at least some computational schemes used in predictive processing is that they make it possible to *avoid* explicit computations of likelihoods and marginal probabilities). Recall that the most general (and oversimplified) way in which this can be formalized is in the form of an equation like this:

$$\text{sensory signals} = \text{deterministic function of hidden causes} + \text{noise}$$

$$s = g(c) + \omega.$$

The relevant part of this equation is, in this context, the function g, which represents a functional relation between s and c that can be used to compute a prediction (and then also a prediction error). PP is committed to variables containing computed values of such functions, because they are needed to compute prediction errors. Variables coding likelihoods are, strictly speaking, optional. By contrast, functional relations between variables are essential, because they specify how to compute predictions. Furthermore, these relations endow the model with a structure, and it is assumed that both parts of an organism's environment and neural processes can be described by models with the same (or a very similar) structure. This is one respect in which (part of) the brain can be assumed to structurally resemble the world, according to PP.[7]

(2) As Gładziejewski correctly points out, functional relations are also embodied in the *dynamics* of hidden variables:

> [...] hidden variables are not only related to lower-level, sensory patterns, but to each other (intra-level) as well. Their values evolve over time in mutually-interdependent ways, which are defined by the model parameters. This is how the dynamics of causal-probabilistic relations between worldly objects are encoded in the model. (Gładziejewski 2015, p. 14)

The idea is that variables at each level do not evolve independently of each other; instead, at any given time, values of some variables depend on the values of other variables. In other words: for each level in the neural hierarchy, there is a set of differential equations that can be used to describe the neural activity at that level (and these equations are connected to each other: individual variables appear in more than one equation). Crucially, the same differential equations can also be used to describe processes in the world (at least ideally and at some level of abstraction).

[7] According to the structural concept of representational content (cf. Bartels 2006), the relation of representation involves a (partial) homomorphism, i.e., a structure-preserving mapping between two structures (where the structures are defined as sets with a family of relations). The paradigmatic example is provided by cartographic maps. Here, the sets are points and the structures are given by spatial relations between the points. A more abstract example would be two dynamical systems whose evolutions can be described by the same differential equations. The structures are here given by functional relations between variables (that specify how the variables influence each other over time).

Since the functional relations between variables at any given time, and their evolution over time, are both embodied in neural dynamics and in worldly dynamics, we can (at least in principle) specify a structure-preserving mapping between world and brain (i.e., variables that can be used to describe processes in the world can be mapped, in a structure-preserving manner, to variables that can be used to describe processes in the brain). PP models are thus *structural representations* (in the sense of Bartels 2006; O'Brien 2015).

(3) Regarding representations of priors, Gładziejewski suggests the following:

> [...] prior probabilities of worldly causes should be encoded in the generative model [...]. If it [i.e., the cognitive system] is to minimize prediction error by realizing Bayesian reasoning, its generative model should be sensitive to the probability of the system's stumbling upon a given type of object, regardless of the current context or current incoming sensory signal. (Gładziejewski 2015, p.14)

This suggestion seems to entail that there are some dedicated variables in the model that encode context-invariant knowledge about the causal structure of the world. Of course, it is plausible to assume that such knowledge is stored *somehow* in the brain, but how best to describe this in terms of PP is as of yet arguably an unsettled issue (for instance, one could also argue that part of such context-invariant knowledge is stored in representations of the functional relations between variables).

A weaker reading of Gładziejewski's suggestion would be that prior probabilities are "attached to a proposition (to a hypothesis, or to an event), before being presented with evidence" (thanks to an anonymous referee for suggesting this reading). The idea would be that prior probabilities are not really context-independent (or even innate), but they are still probabilities, so there have to be representations of probability values or of probability density functions. As we saw above, PP is not committed to representations coding probabilities. So adding prior probabilities to the picture presupposes an unjustified answer to the probability conundrum (**ProC**).

If we are careful to avoid premature assumptions regarding solutions to **ProC**, it therefore seems that functional relations between variables are most relevant to understanding how PP models represent. In fact, the most charitable reading of Gładziejewski would be that functional relations between variables at different levels resemble the causal structure of the world (cf. Gładziejewski 2015, p. 13), and that this is how representations in PP get their contents. In the following section, I will argue that this is precisely the sense in which the contents of (articulated) emulators in Rick Grush's emulation theory are determined. The relevance of this point is the following: if the constraints on cognitive contents provided by mathematical contents in PP are the same constraints that are provided by computational descriptions of emulators (in the sense of Grush's theory), this would suggest that the core aspects of PP have no *specific* bearing on cognitive contents (only via structures that are also posited by computational descriptions in other frameworks). This might then, in a further step, also suggest that personal-level descriptions (for instance, phenomenological descriptions) are independent of subpersonal-level computational descriptions. To some people, this may not be surprising, but it does at least challenge Hohwy's claim (already quoted above) about the putative relevance of computational descriptions for phenomenological descriptions: "[T]his idea [that the brain minimizes prediction error] explains not just that we perceive but *how* we perceive: the idea applies directly to key aspects of the phenomenology of perception." (Hohwy 2013b, p. 1).

4 Structural representations in Rick Grush's emulation theory

In this section, I will briefly review Rick Grush's emulation theory of representation, and will point out the formal resemblance between emulators and models in predictive processing. This will show that the representations PP is committed to have the same general type of computational description as representations in other frameworks, which threatens claims to the effect that computational descriptions in PP have any unique bearing on phenomenological descriptions, or on ascriptions of cognitive contents. The aim of this section is not to show that PP is equivalent to the emulation theory. Instead, the claim I will argue for is the following: Showing that representations in PP are structural representations can only be the first step in the project of clarifying the *unique* contribution made by PP models. Since other frameworks (like Grush's) also posit structural representations, and since this posit is based on computational models featuring the same types of mathematical content that form the core components of PP models, more work has to be done in order to show that PP puts unique constraints on assignments of cognitive contents. (I will provide some suggestions regarding this issue in section 5).

The basic idea of Grush's theory is that the brain uses emulators (forward models) to facilitate cognition and control. An emulator is a system that has the same (or similar) input-output behavior as another system. For instance, a neural system could be an emulator of part of the body, like an arm. If this emulator receives an efference copy of a motor signal, it will compute an approximation to the sensory feedback that will be received after the motor signal has been executed by the arm. Such an emulator can be used online (to provide fast mock feedback), but also offline (e.g., to simulate possible movements). To what extent are such emulators representations? Here is how Grush puts it:

> For each physical parameter that plays a role in the evolution of the system, assign a processing unit, such as a neuron (or a pool of neurons), whose output at any given instant is proportional to (or perhaps a monotonic function of) the value of the physical parameter which it represents. Interconnect the units in such a way as to mirror the relationships between the physical parameters [...]. Thus, when some of the parameters are specified via efferent copy information, the entire ensemble is set in motion, the units' activities and interactions, under the physical interpretations used above, exactly mirroring, in real time, the real physical parameters governing the real arm in motion [...] – an intricate interpretive dance between perhaps thousands of neural partners, choreographed by experience, and stepping in time with the dynamic evolution of the external system. (Grush 1995, p. 60)

When a neural system constitutes an (articulated) emulator of another system, parameters of neural activity can thus be mapped to parameters of the activity of some other system (like an arm), and these parameters evolve in the same way over time (or at least similarly). In other words, the "dynamic of the emulator is formally equivalent to the dynamic of the target system" (Grush 1997, p. 16). This means that the evolution of the respective variables can be described by the same set of differential equations – and this is exactly what we found above with respect to models in PP.

One could object: This merely shows that Grush's emulators mimic the dynamics of target systems in the sense in which hidden variables mimic the dynamics of worldly causes. But it does not show that functional relation between variables at different levels play a role. And what about prediction errors? Aren't they a special ingredient of PP models?

First, consider how prediction errors *do* play a role in Grush's theory. The basic idea is that emulators allow the brain to track the behavior of a target system (like an arm), without having to wait until sensory feedback informs the brain how the target system has changed (after executing a motor signal). However, once the actual feedback is available, it is integrated with the previously computed estimate of the target system's state (using a Kalman filter, cf. Grush 2004, p. 381). In order to do this, a mock sensory feedback is computed (this corresponds to the *prediction* $g(c)$), and this is compared to the actually received sensory signal, yielding a *prediction error*. Secondly, the computation of mock sensory feedback also rests on a *functional relation* between hidden causes and sensory signals: the sensory signal is conceived as a measurement (a deterministic function) of the current state plus (unpredictable) sensor noise (cf. ibid., pp. 381 f.). Hence, functional relations and prediction errors are also an essential part of Grush's theory.

Note that this generalizes to many schemes using Kalman filters. As indicated above, a Kalman filter (Kalman, 1960) combines an estimate of state variables with noisy observations of those variables. In order to compute an optimal estimate, given the observations, one has to model how the individual variables are functionally related over time, and how the values of state variables are related to observations (so these models embody structural information about the representational targets, cf. above). There are different ways to implement this general scheme, but at least a common type of implementation involves computations of prediction errors (for a critical discussion of such implementations, see Eliasmith & Anderson 2003, §9.4.2). Hence, *all* approaches using a common implementation of Kalman filtering involve structural representations of exactly the kind found in Grush's emulation theory and in predictive processing, *and* they also involve representations of prediction errors.

An aspect that is not stressed in Grush's theory is the use of *hierarchical* models. Grush's theory is compatible with a hierarchy of emulators, but (to the best of my knowledge) Grush does not exploit this possibility in his theory. Predictive processing, on the other hand, puts hierarchical models center stage, which constitutes an important extension of the basic idea. This also means that there is an extended sense in which models use structural information (in the form of functional relations between variables at different levels).

Another difference is the way perception and action are conceived of. In predictive processing, the duality between optimal control theory and Bayesian inference (cf. Todorov 2009) is exploited (at least in models based on the free-energy principle), making it possible to model motor control as an inference process that does not involve motor commands (cf. Adams, Shipp, & Friston 2013). This also comes with a circular causality brought about by many action-perception cycles (cf. Friston 2013, p. 8). Furthermore, PP highlights the roles played by precision estimate (Clark 2013a). However, it these differences mainly concern the *use* representations are put to in PP, not their mathematical *contents*. Hence, as far as mathematical *contents* are concerned, it seems that representations in PP are not substantially different from representations in other frameworks.[8]

[8] This is not to deny that they are importantly different from traditional, symbolic representations figuring in classical theories of cognition.

5 What specific constraints does PP put on ascriptions of cognitive contents?

The above may seem to suggest a negative answer to the question posed in this section's title: mathematical contents in PP are not significantly different from mathematical contents in other frameworks, so they do not put any specific constraints on ascriptions of cognitive contents (i.e., no constraints that would be unique to PP). However, there may be other ways in in which PP constrains ascriptions of cognitive contents. So in order to resist a negative answer to the question, I shall briefly explore the following (non-exclusive) strategies:

1. If we want to make the unique contribution of computational descriptions in PP explicit, we have to focus on particular models. Although PP is, in general, only committed to the core mathematical contents mentioned in section 1 (i.e., estimates, predictions, prediction errors, and precision estimates), particular models will feature further types of mathematical contents (perhaps even probability values). These further ingredients may indeed make a more specific contribution. Other models may posit more specific mathematical contents (e.g., specific types of estimates) that provide additional constraints.

2. As already indicated at the end of the previous section, PP not only entails claims about computational mechanisms, but also about how they underpin perception, attention, cognition, and action. These functional descriptions already specify some cognitive contents. In general, they put more specific constraints on ascriptions of cognitive contents.

3. Computations described by PP can be regarded as approximations to *hierarchical* Bayesian inference (see section 1.2). The hierarchy adds further constraints on cognitive contents and phenomenological descriptions, because a particular precision estimate will only fulfill its role *relative to* other precision estimates (at other levels of the hierarchy). For instance, high-level estimates may sometimes have a large influence on lower-level estimates, if their associated precision estimates are high, *relative to* lower-level precision estimates (see, for instance, Hohwy 2013b, ch. 6).

Let me now elaborate a little upon these three points. The upshot will be that focusing only on the structural information provided by functional relations between variables at different levels provides an overly narrow view on the (cognitive) contents of representations in PP.

(1) PP is committed to postulating computations of estimates. A statistical estimate is a function of data obtained by sampling a certain domain. The result can be a single value, a vector, a matrix etc. In particular, it *can* be a probability vector, containing estimated probability values of a certain, say, finite set of events. *If* such probability vectors are computed in a particular PP model, this *might* suggest that different hypotheses are entertained at the same time, together with their associated probabilities. This could be used to support phenomenological descriptions according to which we can, at least sometimes, consciously experience two or more (incompatible) alternatives at the same time. Consider the following example by Henrik Ehrsson (in the context of the *three-arm illusion*):

> [T]he brain is estimating the most probable location of the arm allowing for biphasic probability distributions (Ma, Beck, Latham, & Pouget 2006). In this framework, the result of the multisensory integration process would be that there are two equally probable locations of the right arm. The perceptual consequence of this is the experience of having two right arms. (Ehrsson 2009, p. 312; citation style adapted)

The idea is that the brain is constantly comparing its estimate of the location of the right arm to perceptual input. In experimental setups (which create rather unusual perceptual situations), the brain may be unable to settle on a single location estimate. This may then either lead to the impression that the right arm is not located in a determinate place (but is in a "superposition" of two locations), or it may lead to a revision of the hypothesis that the body has a single arm (as Ehrsson suggests in the quoted passage). So as soon as a particular solution to the probability conundrum (**ProC**) has been found (e.g., a solution according to which the brain does implement probabilistic computations by representing probability values), this will make PP models more specific, and hence such models will provide unique constraints on (and lend specific support to) assignments of cognitive contents. Since **ProC** is still an unsolved problem, let me also briefly explore the consequences of an alternative solution (according to which the brain does not represent probability values or density functions). There are PP models in which the main estimates that are computed are sample means and covariances (as mentioned above, this option is embraced in many models by Karl Friston and colleagues; see e.g. Friston 2009, 2010; Friston & Kiebel 2009). Computing a sample mean (average value) entails that a lot of information contained in a sample is disregarded. For instance, if we sample lightness values in a certain region of the visual field, and then compute the average of these values, we still have information about the light intensity in that region, but we lack information about individual lightness values (the average is indeterminate with respect to these). Following a similar idea, Michael Madary has recently put forward the idea that the probabilistic format of representations in PP can account for the *indeterminacy* of visual perception:

> The indeterminacy of the visual periphery can be interpreted as a probabilistic representational format. Similarly, the missing details as revealed in inattentional and change blindness experiments reveal that our generative models are more successful with the gist of a visual scene and offer only vague estimates about the details. (Madary 2012, p. 1)

So representations of sample means may be an example of representations having a probabilistic format in Madary's sense, and their mathematical content may have a direct bearing on phenomenological descriptions of our (visual) perception. In fact, referring to mathematical contents could even be regarded as a way of *clarifying* part of what we mean when we say that the visual periphery is indeterminate.[9] In the passage quoted above, Madary also mentions perceptual gist, which could similarly be regarded as a cognitive content associated with a representation that has a sample average as its content. Yet another example would be the perception of ensemble properties (cf. Haberman & Whitney 2009): When subjects are briefly exposed to an array of, for instance, faces displaying various expressions, subjects are able to report the average facial expression, without being sensitive to the individual facial expressions. So particular PP models (in which it is made explicit what the estimators are)

[9] The underlying idea is similar to the basic idea of eliminative/revisionist materialism. As Patricia Churchland has pointed out, following a reductionist strategy regarding consciousness and neurobiology is compatible with the idea that "hypotheses at various levels can *co-evolve* as they correct and inform one another." Churchland (1994, p. 25).

can at least support more specific ascriptions of cognitive contents (by providing more constraints on such ascriptions).

(2) Computational descriptions in PP models not only specify mathematical contents, they also specify at least some cognitive contents. For instance, according to the theory of active inference developed by Friston and colleagues, action is not brought about by motor commands, but by predictions of the (perceptual) changes that will be brought about by the respective actions (cf. Adams et al. 2013). Motor commands and perceptual changes are not mathematical contents, they are cognitive contents. To the extent that this basic idea of active inference forms an essential part of PP (which I am inclined to endorse, see also Clark 2015, pp. 8 f.), PP thereby already puts specific constraints on ascriptions of cognitive contents (or even posits representations with specific types of cognitive content).

Apart from that, optimizing precision estimates is typically regarded (by proponents of PP) as the computational mechanism underpinning the allocation of attention (cf. Feldman & Friston 2010; Hohwy 2012). This also entails a cognitive interpretation of mathematical contents.

So this second point emphasizes that PP is not just a label for a class of computational models, but also contains assumptions about cognitive contents. This puts additional constraints on cognitive explanations (but constraints which are not just entailed by positing representations of certain mathematical contents).

(3) I noted above that computations in PP approximate hierarchical Bayesian inference. Hierarchical Bayesian inference puts constraints on ascriptions of cognitive contents. For instance, it favors accounts according to which phenomena of interest can be explained by appeal to the (precision-mediated) interplay between estimates at different levels. Let me just cite one particularly interesting example: Jakob Hohwy's work on mental disorders and their symptoms. Here is how Hohwy puts the general idea of his PP account of delusions (for more details, see also Hohwy 2013a):

> As I have noted, expected precisions and the associated confidence estimates play a key role for the very general balance between bottom up sensory evidence and top down expectations: when the prediction error is expected to be noisy, then it is dampened down and instead the system relies on top-down priors. This tells us that chronic problems with expected precisions can generate chronic problems with getting that balance right. [...] An individual who constantly expects the sensory signal to be noisier than it really is will then tend to be caught in his or her own idiosyncratic interpretations of the input and will find it hard to rectify these interpretations. This gives rise to the idea that psychosis and delusions stem from such faulty uncertainty expectations. (Hohwy 2013b, p. 158)

Hohwy emphasizes the balance between precision estimates at different levels of the hierarchy. This balance is not static, but always context-dependent. In a noisy environment, it can be useful not to rely too heavily on certain perceptual inputs. For instance, in a dark room, vision is not as useful as in bright daylight, so it will make sense to attenuate precision estimates for visual signals. The result is that processing in the hierarchy will be influenced less by vision, and more by top-down signals (or signals pertaining to other sense modalities). However, in spite of the fact that there is no fixed optimal balance between precision estimates, there can still be an optimal balance, relative to each particular (type of) situation.

Furthermore, certain *types* of deviating from the optimal balance in a given situation can lead to certain *types* of psychotic symptoms, like delusions. For instance, a tendency to underestimate perceptual (relatively low-level) precision estimates may lead to delusional beliefs (cf. also Friston, Stephan, Montague, & Dolan 2014, p. 153). As Hohwy points out, this general "account of delusions is not tied to specific representational contents" (p. 160 Hohwy 2013b)—in other words, it does not specify how delusional beliefs with particular contents are generated—but this can actually be seen as an advantage, since it "can go some way to explain the very heterogeneous nature of psychotic symptoms where delusions are elaborated in many different ways." (ibid.). So even general descriptions about mathematical contents (like "the balance between low-level and high-level precision estimates is abnormal or suboptimal") put constraints on cognitive contents, because if a system can be modeled using mathematical contents that deviate in a certain way, it can be expected to behave in a way that favors certain ascriptions of cognitive contents (like delusions).

At present, PP accounts of mental disorders may be relatively general, but as computational models become more fine-grained, the relation between mathematical and cognitive contents (ascribed on the basis of such models) will become tighter. In fact, this is the goal of the field of *computational psychiatry*: "Computational psychiatry uses formal models of brain function to characterise the mechanisms of psychopathology, usually in a way that can be described in computational or mathematical terms." (Friston et al. 2014, p. 148).

So the hierarchical nature of processing entailed by PP puts interesting constraints on ascriptions of cognitive contents. Hierarchical (Bayesian) inference is, of course, not unique to PP (not every hierarchical model is a PP model). But what is emphasized in PP is the special role played by precision estimates (cf. Clark 2013a).

What this hopefully shows is that PP does put specific constraints on ascriptions of cognitive contents (and some of them can be derived from mathematical contents posited by PP). A relevant task for future work is to bring such analyses closer to the structural resemblance discussed in section 3 above. Since functional relations between estimates play a role in generating prediction errors, the information contained therein is relevant for computations in PP. What we can already add now is that relations between precision estimates also play a role; because it is always *relative* precision that is relevant to certain (mal)functionings, as seen in certain mental disorders, for instance. So part of the structure that determines (or at least constrains) cognitive contents is constituted by relations between precision estimates at different levels (for they govern how much influence estimates at some level have on updates at other levels). Other constraints on ascriptions of cognitive contents come from PP's commitment to certain views on action generation and the computational underpinnings of attention.

6 Conclusion

This paper has been concerned with the contents of representations posited by predictive processing (PP). I first identified the core types of representational content in PP: estimates, predictions, prediction errors, and precision estimates. These contents are mathematical contents in the sense of Egan (2014), as opposed to cognitive contents. I next explored to what extent these mathematical contents constrain ascriptions of cognitive contents. Here, Gładziejewski's (2015) insightful treatment of the functional role fulfilled by representations in PP proved helpful. By building on, and clarifying, his remarks on the contents of these representations, I showed that their contents are partly determined by their structure, which is given by functional (mathematical) relations between estimators at different levels. Since predictions are computed on the basis of such functional relations, representations in PP are structural representations (as Gładziejewski rightly points out), and there is a robust sense in which mathematical contents in PP constrain ascriptions of cognitive contents (note that structural theories of representational content are intended as general theories of representational content, which includes cognitive content). One could point out that this relation between mathematical and cognitive contents is not unique to PP, as I showed by referring to Grush's (2004) emulation theory of representation. However, taking the processing hierarchy and the special role(s) assigned to precision estimates in PP into account, I argued that PP does support relative specific ascriptions of cognitive contents. Even more specific claims about cognitive contents in PP can be derived from particular PP accounts. But the focus of this paper has been on the core representational posits common to *all* PP accounts (just as Gładziejewski argues that *all* PP accounts posit genuine representations). Teasing the general representational commitments of PP apart from those content ascriptions that are only entailed by particular PP accounts (like probability values; recall **ProC**), will hopefully help to advance the debate on representationalism and PP with increased conceptual rigor.

References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: active inference in the motor system. *Brain Structure and Function*, *218*(3), 611–643. doi: 10.1007/s00429-012-0475-5

Bartels, A. (2006). Defending the structural concept of representation. *Theoria*, *55*, 7–19.

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychol Bull*, *138*(3), 389-414. doi: 10.1037/a0026450

Churchland, P. S. (1994). Can neurobiology teach us anything about consciousness? *Proceedings and Addresses of the American Philosophical Association*, *67*(4), 23–40. doi: 10.2307/3130741

Clark, A. (2013a). The many faces of precision (replies to commentaries on "whatever next? neural prediction, situated agents, and the future of cognitive science"). *Frontiers in Psychology*, *4*, 270. doi: 10.3389/fpsyg.2013.00270

Clark, A. (2013b). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. doi: 10.1017/S0140525X12000477

Clark, A. (2015). Radical predictive processing. *The Southern Journal of Philosophy*, *53*, 3-27. doi: 10.1111/sjp.12120

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

Egan, F. (2014). How to think about mental content. *Philosophical Studies*, *170*(1), 115-135. doi: 10.1007/s11098-013-0172-0

Ehrsson, H. H. (2009). How many arms make a pair? Perceptual illusion of having an additional limb. *Perception*, *38*(2), 310–312. doi: 10.1068/p6304

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: computation, representation, and dynamics*. Cambridge, MA and London, UK: The MIT Press.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*. doi: 10.3389/fnhum.2010.00215

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, *4*(11), e1000211. doi: 10.1371/journal.pcbi.1000211

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. doi: 10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. doi: 10.1038/nrn2787

Friston, K. (2013). Life as we know it. *Journal of The Royal Society Interface*, *10*(86). doi: 10.1098/rsif.2013.0475

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221. doi: 10.1098/rstb.2008.0300

Friston, K., Stephan, K. E., Montague, R., & Dolan, R. J. (2014). Computational psychiatry: the brain as a phantastic organ. *The Lancet Psychiatry*, *1*(2), 148–158. doi: 10.1016/S2215-0366(14)70275-5

Gładziejewski, P. (2015). Predictive coding and representationalism. *Synthese*, 1–24. doi: 10.1007/s11229-015-0762-9

Grush, R. (1995). *Emulation and cognition* (Unpublished doctoral dissertation). University of California, San Diego.

Grush, R. (1997). The architecture of representation. *Philosophical Psychology*, *10*(1), 5–23. doi: 10.1080/09515089708573201

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*, *27*(3), 377–396.

Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–734. doi: 10.1037/a0013899

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, *3*. doi: 10.3389/fpsyg.2012.00096

Hohwy, J. (2013a). Delusions, illusions and inference under uncertainty. *Mind & Language*, *28*(1), 57–71. doi: 10.1111/mila.12008

Hohwy, J. (2013b). *The predictive mind*. Oxford: Oxford University Press.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, *82*(1), 35-45. (10.1115/1.3662552) doi: 10.1115/1.3662552

Kiebel, S. J., Daunizeau, J., Friston, K. J., & Sporns, O. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, *4*(11), e1000209. doi: 10.1371/journal.pcbi.1000209

Kvam, P., & Vidakovic, B. (2007). *Nonparametric statistics with applications to science and engineering*. Hoboken, NJ: John Wiley & Sons.

Kwisthout, J., & van Rooij, I. (2013). Bridging the gap between theory and practice of approximate Bayesian inference. *Cognitive Systems Research*, *24*, 2-8. doi: http://dx.doi.org/10.1016/j.cogsys.2012.12.008

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432-1438. doi: 10.1038/nn1790

Madary, M. (2012). How would the world look if it looked as if it were encoded as an intertwined set of probability density distributions? *Frontiers in Psychology*, *3*(419). doi: 10.3389/fpsyg.2012.00419

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective.* Cambridge, MA, and London, UK: The MIT Press.

O'Brien, G. (2015). How does mind matter? In T. K. Metzinger & J. M. Windt (Eds.), *Open mind* (chap. 28(T)). Frankfurt am Main: MIND Group. doi: 10.15502/9783958570146

O'Brien, G., & Opie, J. (2004). Notes toward a structuralist theory of mental representation. In H. Clapin, P. Staines, & P. Slezak (Eds.), *Representation in mind: New approaches to mental representation* (pp. 1–20). Amsterdam (a.o.): Elsevier.

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

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. doi: 10.1038/4580

Rescorla, M. (2015). Bayesian perceptual psychology. In M. Matthen (Ed.), *The oxford handbook of philosophy of perception* (pp. 694–716). Oxford: Oxford University Press.

Todorov, E. (2009). Parallels between sensory and motor information processing. In M. S. Gazzaniga

(Ed.), *The cognitive neurosciences* (4th ed., pp. 613–623). Cambridge, MA / London, UK: The MIT Press.