How Radical is Predictive Processing?

Nico Orlandi and Geoff Lee[1]

Forthcoming in *Andy Clark & Critics* (Eds., Colombo, Irvine, & Stapleton) (OUP, 2018)

For years now, the work of Andy Clark has set the agenda for the philosophy of cognitive science. From his early work on connectionism, to his groundbreaking *Being There* (1998), to the most recent *Surfing Uncertainty* (2015), Clark's contributions consistently invite us to reconsider fundamental assumptions about the mind. Clark's work often balances commitment to representation with embodied and embedded approaches to cognition. This tendency is present in Clark's latest project where a Bayesian approach to perception is integrated in an action-oriented framework. This will be our focus here.

In the first part of this chapter, we will discuss the predictive processing framework (PP), explicating its relationship with hierarchical Bayesian models in theories of perception. In the second part, we examine the relationship between perception and action in the PP model. Our overarching goal is twofold. We would like, first, to get clearer on the picture of mental activity that Clark is presenting. Second, we will point out that, although the framework presented by Clark certainly has interesting novel features, some of Clark's glosses on it are misleading. In particular, we think that Clark's interpretation of predictive processing as essentially a top-down, expectation-driven process, on which perception is aptly thought of as "controlled hallucination", exaggerates the contrast with the traditional picture of perception as bottom-up and stimulus driven. Additionally, we think that, despite the rhetoric, Clark's PP model substantially preserves the traditional distinction between perception and action.

1. Hierarchical Bayes in Perception

The predictive processing framework (Clark 2013, 2015, Hohwy 2013) combines two components: (1) A hierarchical Bayesian model, and (2) a predictive coding algorithm for the updating of representations at each stage of the hierarchy. (2) is intended as a proposal for how to implement (1) (Clark 2016, pp. 27-28). As we will see in section 2, Clark thinks that it is the predictive coding aspect of the package that provides the most striking contrast with traditional views. Nonetheless, it is helpful to start by considering the Hierarchical Bayesian model in isolation. In this section, we offer clarification of the details, and some interpretative points that differ in crucial respects from Clark's.

Marr's (1982) distinction between three different levels of analysis of a perceptual information-processing system (task, algorithmic, and physical-implementation levels) is a helpful framework for understanding different versions of hierarchical Bayes. On the Marrian picture, the goal of perceptual processing is to accurately derive perceptions of distal stimulus features ("the feature set") – such as surface shape, orientation and color – given proximal input (e.g. retinal stimulation). The task-level analysis specifies this input-output function. A visual perception of surface depth, for example, may result from

[1] This chapter was a fully collaborative enterprise; the order of the authors' names is arbitrary. We thank Matteo Colombo, Liz Irvine and Susanna Siegel for comments on an earlier draft of the chapter.

a certain disparity in left and right retinal inputs. The task-level analysis can also involve an explanation of why this input-output function is appropriate[2]. For instance, in a natural environmental setting, a certain retinal input may be likely to have been produced by a surface at a certain depth, and the visual system may be adapted to reflect this fact.

The "algorithmic" level, by contrast, specifies *how* the system performs the task – in the perceptual case, how it derives percepts. The algorithmic description identifies the intermediate steps between input and output. Marr appealed, at this level, to representations of various kinds – for example, 2D representations that are intermediate between retinal input and 3D representations – as well as to computational rules governing the transitions between representations.

One of Marr's crucial insights is that the same process described in task-level terms can be implemented, or realized, by a variety of different algorithms. Furthermore, an algorithm can be implemented by different neurophysiological or other physical processes (Marr's third level of analysis). However, crucially for our purposes, the algorithmic level can itself be gradually unpacked at different levels of specificity, and so can be thought of as containing its own nested "levels". Furthermore, a partial description of the algorithm can involve "subtasks" that can themselves be given a Marrian task-level analysis. As we will see, we get from the weakest versions of Hierarchical Bayes to Predictive Processing by just such an unpacking of algorithmic sub-levels and sub-tasks.

The "hierarchical" part of Hierarchical Bayesian models is a partial specification of how the task of computing the feature set is carried out. The hierarchy is a series of quasi-modular processing stages, each devoted to computing a different element of the feature set (so the overall percept is distributed across stages). Importantly, each stage is only causally affected by (i.e. only receives information input from) the adjacent stages in the series (above and below), so its series-position is determined by these input-relations (Lee and Mumford (2003)). For example, it might be that in visual processing, 2D surface shape is computed from information about edge properties at a lower stage, and from higher-stage information about the 3D shape of the object that the surface is part of. The lowest stage may be inputted directly from the proximal stimulation, or there may be further lower stages that compute features not mentioned in the initial feature set. Also, each stage may be further sub-divided into horizontally connected units, e.g. different parts of a feature-map, although we ignore this complication here.

A model is "Bayesian" at the *task* level, if the appropriateness of the input-output function is explained in terms of probability distributions over distal and proximal variables. The function (approximately) matches the inference that a rational agent would make given knowledge of the probabilities and knowledge of the proximal input. In the context of hierarchical models, we can give such an analysis both of the overall perceptual task, and the sub-task carried out at each stage of the hierarchy. Either way, it comes in the form of conditional probability distributions linking variables at adjacent stages (giving us a "Markov random field" structure (see Blake et al. 2011 ch.1)). For example, the analysis tell us how well a given shape estimate predicts the edges encoded below (the likelihood), and how well confirmed it is by the high-level information from above (the prior), jointly entailing, by Bayes theorem, a posterior distribution over surface shape percepts.

[2] "Appropriate" could be construed in different ways, depending on what the goal of the system is understood to be (e.g. reliably veridical percepts vs percepts that are conducive to reproductive success).

Typically, it is further assumed that the goal of the perceptual system is to select a single value from this distribution – usually the feature with maximum posterior probability. On some versions, however, each stage encodes the whole posterior distribution (Friston 2009). Further, the process is iterative. The inputs to each stage will change as the informational states change at the adjacent stages, and these changes will mandate a new update.[3]

Different versions of hierarchical Bayes then differ on whether such a Bayesian description is appropriate *only* at the task level or also at the algorithmic level. As we see it, "algorithmic bayesianism" involves three components: (a) the relevant probability distributions are represented by the system (b) the input-output function computed by the algorithm over these representations approximately conforms to ideal Bayesian inference and (c) a uniform computational strategy is used across contexts. Task-level Bayesians need not, and often do not, endorse such a further interpretation (Griffiths et al. 2012). The feature representations at each stage might encode single values, not probability distributions. The conditional distributions linking stages might be seen as constraining the system's proper function without being represented; for example, they might be environmental statistics that the system is designed to conform to (Geisler 2008, Orlandi 2014). Finally, approximately optimal results might be achieved using a context-dependent "bag of tricks" rather than a uniform algorithm across contexts.

In contrast with such weaker versions of Bayesianism, the predictive processing framework, as interpreted by Clark, is Bayesian in the stronger sense that "neural representations (…) encode probability density functions and the flow of inference respects Bayesian principles." (Clark 2015, p. 39) Additionally, Clark thinks that a uniform type of algorithm (predictive coding) is used across contexts. In these respects, Clark follows Friston (2009). Probabilistic representations in perception do have some empirical support (Knill and Pouget 2004, Ma and Jazayeri 2014), as does predictive coding (Huang and Rao 2011, Rao and Ballard 1999). Still, it's important to note that neither are an inevitable part of the hierarchical Bayes package.

An attraction of a Bayesian system is its informational richness. Taking into account uncertainty, rather than using crude conditionals linking variables, allows for much more accurate feature estimates. However, a common problem for Bayesian models is that the probabilistic computations that would give optimal results are often intractable. This leads to one motivation for a Hierarchical model. Even with a relatively small number of stimulus variables, we have the problem of a combinatorial explosion of probabilistic dependencies that might have to be taken into account to compute a posterior on a particular feature, because each feature might depend on any combination of other features in arbitrary ways. A Hierarchical model makes this more tractable by making *conditional independence assumptions*: conditional on the fixing of variables adjacent in the network, the variable of interest is independent of other variables (see Domingos 2015, ch. 6 for helpful discussion). This only works if the conditional independence assumptions are roughly correct, putting interesting constraints on the sets of features that can be successfully computed using a model of this kind: the external features must form a dependence hierarchy that mirrors the processing hierarchy.

[3] Given that we have stable percepts, this iterative process should converge on a solution in normal circumstances, although convergence need not be guaranteed in every case (e.g. binocular rivalry).

So, a distinctive feature of this model is the way it *restricts* the kind of priors drawn on, by restricting the domain to feature sets with the right hierarchical dependencies. Inferential systems typically rely on background assumptions to infer a representation of the environment from sensory stimulation; here, all the work is done by conditional probability distributions linking adjacent variables in the hierarchy. In a specific sense, then, hierarchical Bayes is actually *more* similar to a traditional "bottom-up" approach than a less restricted model that allows direct inferences from the highest-stages to lowest-stage variables.

Another key feature of this model is the way that it allows the system to take into account *past evidence* as well as current evidence in assessing the distribution of a certain feature. One way you could do this is by simply storing recent sensory information in a buffer, and directly taking it into account in perceptual inference. This model instead takes into account past evidence by incorporating it, through a step-wise recursive process, into its posterior estimates of the features it represents. These posterior estimates become the new priors at each step, to be combined in a rationally weighted way with new evidence to produce a new posterior.[4] We think this shows that a description of the visual system as integrating sensory evidence that comes in *over a period of time* is more fundamental than a description of the system as combining top-down prior information with sensory evidence (which Clark emphasizes): the latter is a mere means to the former.

Another simple point to stress in this regard is that the current bottom-up signal, and the top-down, past-evidence-based priors will be weighted differently depending on their relative reliability/informativeness. In particular, in a novel environment, at least initially, the visual system's priors will be neutral between many possibilities, and the bottom-up signal will do most of the work. That is, there's nothing in the model ruling out current evidence often being much more informative than past evidence (again, contrary to Clark's emphasis).

So far, we've described what is distinctive about the Hierarchical Bayesian picture. We now consider predictive coding.

2. Predictive Coding in Perception

"Predictive coding" refers to "a strategy for the efficient encoding and transmission of information" that is used by a family of similar algorithms that can implement (i.e. further unpack) Hierarchical Bayes to give us "predictive processing" (*ibid*, p. 27, see also Spratling (2017)). In particular, predictive coding involves the use of error-correction units in the updating of representations at each stage of the Bayesian hierarchy (more detail below).

Predictive coding, according to Clark, provides the most striking contrast with traditional models of perception.

[4] In this way, hierarchical Bayes is an example of a broader class of *recursive estimators*, including the well-known Kalman Filters.

"What is most distinctive about the predictive processing proposal (and where much of the break from tradition really occurs) is that it depicts the forward flow of information as solely conveying error, and the backward flow as solely conveying predictions." (p. 38).

By "tradition" here Clark means models that posit a "feedforward (even if attention-modulated) cascade of simple-to-complex feature detection" (p. 45). For example, Marr envisaged perceptual processing as sequentially constructing first lower-level feature representations (e.g. edge maps), *then* (on this basis) intermediate levels representations (e.g. surface representations) *then* high-level feature representations (e.g. 3D object representations).

Before we describe the predictive coding framework in further detail, we want to stress that it is only one possible implementation of hierarchical Bayes, and that it is a fairly speculative empirical proposal. Why is so much weight put on this version (Rescorla 2017)? One could explore hierarchical Bayesian models of perception in a more implementationally neutral way, including taking seriously the possibility of a "scruffy" mixture of implementional strategies. This pluralism would seem to be more in the spirit of Clark's remarks in his 2013 (p. 194) that "considerable distance still separates such [Bayesian] models from the details of their implementation in humans or other animals. It is here that the apparent triumph of the neats over the scruffies may be called into question."

Further, we need to be clear whether any alleged feature of the model depends on this particular implementation. Consider Clark's gloss of the predictive framework, on which "the *flow* of representational information (the predictions), at least in the purest versions, is all downwards (and sideways)." (*ibid*, p. 46). This is meant to be supported by the fact that a predictive coding algorithm uses error-correction units connected to higher-stage representation units, and so (allegedly), the forward flow "solely conveys error" (*ibid*, p. 38).

This gloss introduces a puzzle. Suppose that (contrary to intention) we regard it as a gloss about the hierarchical Bayesian model considered independently of its predictive-coding implementation. Hierarchical Bayes, intuitively, involves *bottom-up* flow of information, in the sense that it involves evidence about the world coming in from the bottom, over a period of time, that results in Bayesian updating of priors through conditionalization. It is true that the system can have strongly weighted priors, and in that sense it has a "top-down" component. This makes it different from a bottom-up, Marrian, picture. Nonetheless, there is no obvious sense in which there is *only* top-down information "flow" on this picture. If that is right though, how could merely specifying how the hierarchy is implemented make any difference to our interpretation?

That it does make such a difference is clearly Clark's picture; again, the use of error correction signals is the critical factor. We'll now argue that this is the wrong interpretation, focusing on the canonical version of predictive coding from Rao and Ballard's (1999) paper.

Here's Rao and Ballard's helpful diagramming of the model ((b) shows the details inside the predictive estimator shown in (a)):

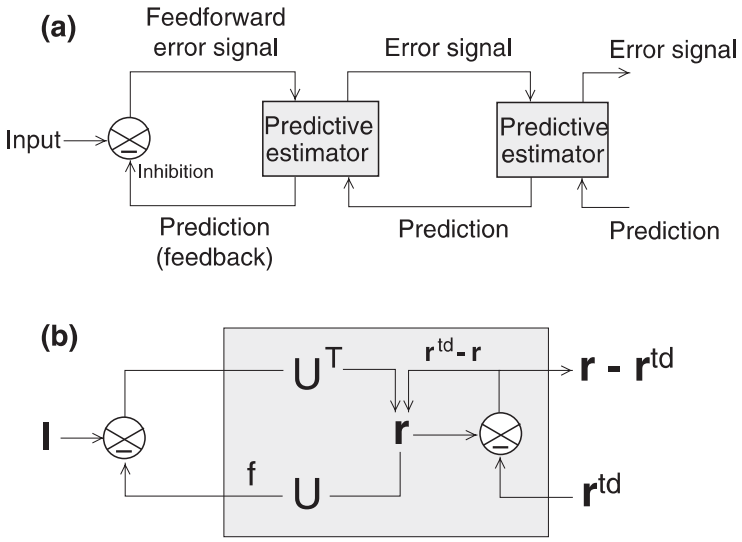

Figure 1. Hierarchical network for predictive coding, from Rao and Ballard 1999, p. 80.

In diagram (b), **r** is the state of the representation unit at that stage, representing a stimulus feature; $\mathbf{r}^{td}$ is the "prediction" generated by the representation unit one stage up. U (modified by f to represent non-linearities in the transformation, and realized by the weights of neural connections), is a function converting **r** into a prediction of the lower stage, which is compared with the representation at the lower stage at a subtraction unit (the crossed circle in the diagram) producing the "error representation". This is then converted, by the inverse of U, U^T, into a signal that is used to update **r**, along with the prior coming from the stage *above* in the form of an error signal generated by comparing **r** with $\mathbf{r}^{td}$ (so contrary to common glosses, there actually is a "downwards error signal" in this model, albeit inside the estimator). The model also includes a learning algorithm, which, over larger time scales, updates the generative model (i.e. U and the analogous functions at other stages).

Note that the system involves both representation units *and* error units – the representation units were already postulated as components of the hierarchical Bayesian model; we are adding the error units to flesh out the implementation.

Now let's compare this with a different diagramming of a Hierarchical Bayesian model, from Lee and Mumford's influential (2003) paper:

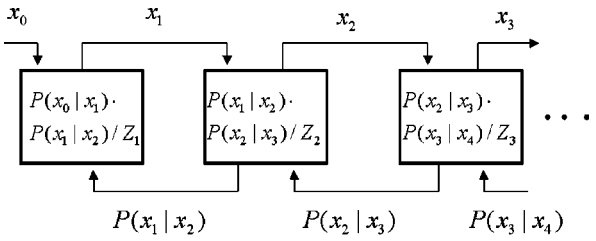

Figure 2: Schematic of Hierarchical Bayesian Inference from Lee and Mumford 2003, p. 1436.

$x_1 \ldots x_n$ are variables ranging over values of the stimulus features (e.g. edges), and x_0 ranges over values for sensory evidence (e.g. retinal stimulation). Inside each box, the priors and likelihoods for the relevant variable are combined to calculate the maximum posterior value of the feature variable, which is fed to the next step. Superficially, then, this looks quite different from the "error correction" model, because a representation of a stimulus feature is being fed forward from one step to the next, not an error signal, as on the Rao and Ballard model.

Despite this appearance though, we think that the right interpretation of the predictive coding model is that it *does* involve information about stimulus features being fed forward. Indeed Lee and Mumford themselves explicitly say that their model is compatible with predictive coding. Moreover, what fig 2 really illustrates is just a version of Hierarchical Bayes, which, as we have said, can be *implemented* by predictive coding.

Clark might respond by insisting that "information flow" really is very different on the error correction model. He might dramatize this by considering the case where predictions from above match the lower stage representations, so there is no error signal. The absence of error signal means that the lower stage is prevented from causally influencing the higher stage. If you're thinking of "information flow" as a partly causal notion, it's natural to read this as meaning that there is no information flow from the lower stage to the higher stage.

We think there are two problems with this reading, however. First, the absence of a physical signal clearly *can* carry information. For example, I might tell you that if you don't receive a telephone call from me at 6pm, this signals that the coast is clear, and the heist can begin! We can read the absence of an error signal as passing forward the information that a certain stimulus feature is present. So, forward information flow doesn't require a forward signal.

Second, it's not even true in cases like this that there is no "feed forward" physical signal, representing information about the stimulus. To see what we have in mind, compare again the model diagrams in figs 1 and 2. Following Spratling (2008), we think that the apparent difference here is superficial, depending on an arbitrary (at least from an information processing perspective) choice of stage-individuation, that is, of which components to parse together as a "stage". As is illustrated in figure 3 below, one can equally well parse the stages in Rao and Ballard's model in such a way that it is a representation of stimulus features that is fed forward, and it is an error signal that is passed down. It's (roughly) in this way that Spratling argues that predictive coding and biased competition models can be reconciled with one another.

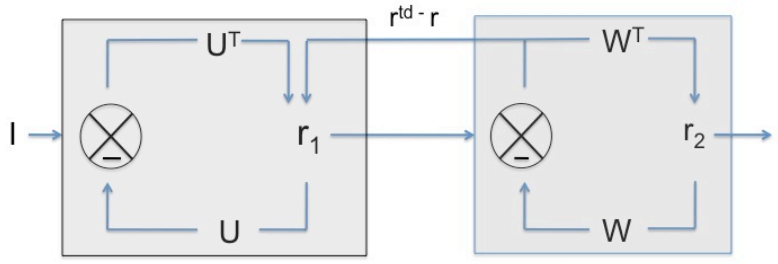

Figure 3. Reparsing of levels in the Rao and Ballard diagram, inspired by Spratling 2008.

Viewed this way, in the case where there is a match at the error unit, what happens is that there is no signal from the higher stage to the lower stage, although there *is* a physical signal coming "bottom up" in the other direction.

Perhaps if "stage" is given a neurophysiological reading, there could be an interesting difference between these two parsings of the model (Spratling *ibid* p.7). For example, do neural feedback connections between cortical areas have the function of inhibiting the activity of error units, or enhancing the activity of representation units? (see Bastos et al 2012, Friston 2005, Lamme et al 1998). But this is not a difference in the flow of information. It is merely a difference in how this flow is realized neurophysiologically (and as we said, we think it could even be realized by *absences* of signals).

Talk of "controlled hallucination" (Clark 2015, p. 14), as well as the terminology of "error correction", suggests that the bottom-up signal only does the work of adjusting and fine-tuning the system's representation. Certainly, the bottom-up signal *is* indicating that the prior needs adjusting, and it is specifying the appropriate direction of adjustment. But even in a case where the prior is much more weighted than the bottom-up estimate, we are still just combining two separate estimates of the stimulus in a way that is essentially symmetrical between processing directions. We could equally well think of the prior as fine-tuning the bottom-up estimate. That would be more natural in a case where the bottom-up signal gets more weight; but such cases are ubiquitous (Clark concedes as much in several passages, *ibid* pp. 41-42, p. 57).

Our interpretation is further supported by considering an analogy with inter-modal integration. Suppose vision and touch deliver information about a common feature, such as the size of a ridge on an object (Ernst and Banks 2002). On a Bayesian model, the system's task is to combine the information from these sources in a probabilistically rational way, weighting them according to their relative reliability. We are in fact able to accommodate reliability in this way: for example, Fetsch et al (2012) provide neural evidence that visual-vestibular integration in macaque monkeys is able to flexibly take into account the reliability of cues on a trial by trial basis.

Now suppose that such a Bayesian cross-modal integration process is implemented using a "visual error correction" code, obtained from comparing the visual and auditory codes, feeding back and forth with the initial visual code to produce an integrated estimate. It would clearly be a mistake to interpret this as making the system in some way more "visuo-centric" than a version of the system implemented in a different way. This is made vivid by considering the case where audition is given much more weight than vision; this is perfectly compatible with the suggested implementation, even though it is, if anything, most naturally described as "audio-centric" processing.

3. What is being represented in the hierarchy?

As we have set things up, the features represented at each level are external stimulus features. However, Clark sometimes writes as if it is really intrinsic features of neural states that are represented; "One key task performed by the brain, according to these models, is that of guessing the next states of its own neural economy." (Clark 2013 p.183). Each stage is depicted as attempting to predict the intrinsic neural state of the

stage below, and adjusting its guess based on the bottom-up error signal. This suggests it is *representing* the neural state of the stage below. What to make of this?

Suppose, for example, that we are representing edges at the lower stage of visual processing and shapes at the higher stage. My current shape representation "predicts" what is happening at the lower stage in the sense that it is *converted into an edge representation* (the "edge prior"), which is compared with the lower-stage edge representation at the subtraction unit, producing the error representation. This description only mentions representations of external features. However, you could adopt a kind of testimonial metaphor here, and regard the top-down generated edge representation as "predicting what the lower stage will say about edge properties". Or, since the edge representations are in a neural code, you could also regard the top-down edge representation as predicting what the neural code of the edge representation will be at the lower stage.

Clark clearly thinks that the "predicting the neural code" gloss is helpful for getting a feel for how the system is able to "bootstrap" its way to the right conditional priors linking stages, via the learning algorithm it uses. One rationale for this would be if we interpret the learning algorithm not as simply improving a probabilistic model of a predetermined feature set, but rather as also *changing the feature set*, by changing the response profiles of the representations. Then, over the longer time scale of learning, the only common currency is the intrinsic features of neural codes, not their contents. It is therefore (perhaps) helpful to think of the system as learning to predict neural codes rather than predicting contents; the learned contents (e.g. the edge and bar representations learned in Rao and Ballard's simulation (p. 81)) are a useful side-effect of the learning process.

This interpretation of the learning algorithm, and the more general idea of perceptual learning as motivated by the reduction of prediction error, warrant further discussion. The point we want to stress here is that even if perceptual *learning* (i.e. improving the transitions between stages, by e.g. changing connection strengths) is best understood in a way that ignores environment-directed contents, that doesn't suggest that we could similarly understand perceptual *updating* (changing representations at individual stages) at short time-scales without thinking of it as operating on representations of external stimulus features. If we are trying to understand how the system achieves an accurate representation of the environment by applying a Bayesian analysis, then the top down predictions are helpfully understood as priors concerning the stimulus features, not representations of neural states.

4. Perception and Action

We have just been critiquing Clark's interpretation of the predictive processing model of perception, partly emphasizing ways in which the model is less divergent from orthodoxy than he suggests. We now want to make a similar point about Clark's account of action and perception in part II of *Surfing Uncertainty*. The account certainly has some radical features, nonetheless we think that some of the interpretative commentary is misleading.

On p. 65 Clark says: "(…) PP makes a strong proposal concerning the cognitive centrality of a complex looping interplay between perception and action. In fact, so complex, central and looping is the interplay that perception will emerge (Part II) as

inseparable from action, and the theoretical divisions between sensory and motor processing will themselves be called into question."

Later Clark, quoting Friston, says: "The perceptual and motor systems should not be regarded as separate but instead as a single active inference machine that tries to predict its sensory input in all domains." (p. 121)

Although these passages make it sound as though accepting Clark's version of PP threatens the perception/action boundary both empirically and theoretically, the picture we get from part II need not be given this radical interpretation (see also Shea's commentary on Clark 2013).

Action control, in this context, mostly means motor control, or the ability to control one's own body. Theories of motor control tend to be concerned both with cases of full-on intentional action, where one tries to achieve an explicitly set goal, and with more reflex-like bodily activities such as catching a baseball or dodging a punch.

In explaining motor control, researchers face problems that in some ways parallel the problems that arise in perception. Motor control, like perception, happens in conditions of uncertainty. This is a reason why Bayesian approaches are attractive here, subject to the usual provisos about computational plausibility.

One source of uncertainty in motor control is given by delays in sensory and proprioceptive feedback (*ibid*, p. 114). Transmittance limitations in nerves and synapses produce signaling delays that would seem to impede fluid motion. In fast reaching, for example, the brain has to receive, and respond to, a stream of proprioceptive information concerning the position and trajectory of one's arm and hand, as well as information about the location of objects (Clark and Toribio 1994 p. 402). The problem is that, for fast motions, there are signaling delays in proprioceptive feedback coming from nerve endings, as well as in the sensory systems that inform the brain about states of the world. Yet we seem capable of reaching objects, generally successfully, even when the movement is fast.

A particular kind of Bayesian model is often put to work to understand how this happens. The idea is that the brain uses a ``forward model'' – a neural network whose interconnected units model our arm-hand apparatus. The network emulates the interplay between arm-hand parameters, and it provides mock feedback in place of absent proprioceptive feedback. In this way, the limits of real time transmission of information can be circumvented to produce skilled reaching, by using top-down expectations concerning the projected position of the body, in combination with predictions concerning what the proprioceptive feedback should be. These top-down priors are integrated with bottom-up proprioceptive information, but the role of bottom-up information is limited in two ways relative to the perceptual case. First, unlike in a case of perceiving a novel environment, in the case of a novel motor command, the system *does* have a fairly accurate top-down model of what will happen right from the beginning, which can therefore carry a lot of weight in the Bayesian calculation. Second, because the system has this model, it can start putting it to use *before* getting bottom-up feedback, thus solving the time-delay problem. In this model, then, the bottom-up signal is "correcting" the forward model.

Uncertainty is also present when we introduce goals in our understanding of action. The uncertainty here is given by a redundancy problem. Once a goal is set, there are innumerable ways of achieving it. There are innumerable trajectories my arm and

hand can follow once I decide to reach out for a cup of coffee. The motor system has to select a motor command – a command of muscle activation – from among many possible sequences of motor commands, discounting the commands that are not advantageous.

Bayesian models of action control are again useful in this context. Such models – one strand of which is *Optimal Feedback Control* (or OFC) – propose to explain how the redundancy problem is solved by thinking that the motor control system engages in an unconscious form of decision-making (ibid pp. 117-119). The system uses Bayesian inference to estimate environmental conditions, and it then selects "optimal" motor commands for achieving the set goal. Optimality is characterized as relative to a *cost function* that rewards achievement of the goal, and efficiency. Roughly, in this picture, the motor system picks motor commands that, given the goal – reaching for coffee – and the hypothesized state of the world – where the coffee cup is with respect to the body – minimize expected costs for performing the action. The notions of a goal and of a cost function are central to this type of explanation.

Since predictive processing is a particular implementation of Bayesian models, we might naively expect Clark to be sympathetic to Optimal Feedback Control. Instead Clark, following Friston, proposes an alternative that he calls "Active Inference". Active inference is an extension of the predictive processing framework to the case of action:

"The core idea is thus that there are two ways in which biological agents can reduce prediction error. The first (as seen in Part I) involves finding the predictions that best accommodate the current sensory inputs. The second is by performing *actions that make our predictions come true* – for example, moving around and sampling the world so as to generate or discover the very perceptual patterns that we predict." (*ibid*, p. 121).

And:

"Active inference names the combined mechanism by which perceptual and motor systems conspire to reduce prediction error using the twin strategies of altering predictions to fit the world, and altering the world to fit the predictions." (p. 122)

The most radical idea here is that we can dispense with goals / cost functions, and instead do all the work with sensory predictions (*ibid*, p. 124). For example, instead of appealing to a goal of picking up a coffee cup to explain a bodily movement, and a motor plan for carrying out this goal, the account appeals directly to my expectation that I will move in a certain way. When this expectation isn't met, the motor system accommodates this fact, not by changing my belief, but by changing the world – by moving my hand.

There's a question here whether this is a kind of eliminativism about goals, or whether they are "folded into" the predictions. We'll treat that as a verbal issue: what matters is whether solely appealing to predictions gives us the conceptual and empirical resources to explain action control correctly.

There are several problems in this context. First, there is the obvious point that there is a big difference between wanting something to happen and expecting it to happen. The kind of expectations involved in perception don't have a desire-like functional profile, even taking into account the point that perception involves an expectation-driven process of moving around to better sample the world. For example, if

I expect to see my keys on the table, this might inform the kind of active sampling that I engage in, but it won't make me *put my keys on the table*. This is true even if what is expected is a bodily movement – I can expect my body to move in a certain way, without intending it to move that way (Colombo 2016, Klein 2016).

Second, as Rescorla (2017) emphasizes, the account is less explanatory than the OFC account in a vital respect. On both accounts, my reaching for the coffee cup in a certain way is partly explained *by* the fact that, on launching the motion, I expect it to unfold in a certain way. But why do I expect that? OFC has an answer to this question that has to do with the calculated optimality of the specific bodily trajectory given my goals. Active inference, by contrast, just takes these expectations as given without explaining them.

Given these problems, we wonder whether Clark should favor Friston's radical story over OFC or other models. Additionally, we have a concern about Clark's claim that a predictive processing account of action threatens to dissolve the distinction between perceptual and action-oriented processing. We don't see this as motivated by the framework. Clearly, the fact that perception and action use the same *kind* of processing (they are "computational siblings" (*ibid,* p.120)) is perfectly compatible with their being quite separate systems. Perception and action might deal with operating under conditions of uncertainty in analogous ways – by combining top-down models with sensory information – and they might use a similar predictive coding algorithm to approximate Bayes optimality. If action processing operates without a cost function, that deepens the analogy (although some models of perceptual processing do appeal to cost functions. See Mamassian et al. 2002). But why do these computational similarities ground the claim that perception and action are now "inseparable"?

Clark himself seems to admit separability when he alludes to the fact that perception and action have different "directions of fit" (ibid, p. 121). One system aims to progressively change its representations to better fit the world. The other system aims to, roughly, change the world to conform to its representations. The extreme view that there is only one kind of prediction, a "sensory-motor prediction", in which errors can be accommodated either by changing the world *or* by changing the representation of the world, has the implausible implication that perception can always appropriately respond to error by changing the world to fit it (for example, putting my keys on the table to accommodate not seeing them there when I expect to). We also get the appealingly Buddhist, but psychologically unrealistic, conclusion that unsatisfied desires can always be dealt with by changing the desires.

One could try to accommodate this point by distinguishing the *contents* of the predictions: for example, errors in predictions about proprioceptive feedback and sensorimotor contingencies are dealt with through initiating/altering bodily movement, whereas errors in predictions about external stimulus features are dealt with through representational changes. Indeed, Clark alludes to this difference in content (*ibid,* p. 121). Perception for action predicts how an object looks like given some movement, while perception considered alone predicts how an object looks independently of one's movement (*ibid*, ch. 1). This would only deepen the case that we are dealing with separate systems however, since we now have both a functional *and* a content difference between sensory and action predictions. We doubt that Clark would accept such a sharp distinction between the contents of say, vision-for-action and vision-for-perception, but if

he did, then perception and action would be separate. They should not be described as inseparable strategies in a combined mechanism to reduce prediction error.

References

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron*, *76*(4), 695–711. http://doi.org/10.1016/j.neuron.2012.10.038

Blake, A., Kohli, P., & Rother, C. (2011). Markov random fields for vision and image processing. Mit Press.

Clark, A. (1998). Being there: Putting brain, body, and world together again. MIT press.

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

Clark, A. (2015). Surfing uncertainty: Prediction, action, and the embodied mind. Oxford University Press.

Clark, A. and Toribio, J. (1994). Doing without representing? Synthese, 101:401{431.

Colombo, M. (2016). Social motivation in computational neuroscience: Or if brains are prediction machines, then the Humean theory of motivation is false. *Routledge Handbook of Philosophy of the Social Mind*.

Domingos, P. (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. Basic Books.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. http://doi.org/10.1038/415429a

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. Nature neuroscience, 15(1), 146-154.

Friston, K. J. (2005). A theory of cortical responses. Philosophical Transactions of the Royal Society B: Biological Sciences, 360(1456):815–36.

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

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, *59*, 167–192. http://doi.org/10.1146/annurev.psych.58.110405.085632

Griffiths, T. L., Chater, N., Norris, D., & Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are): comment on Bowers and Davis (2012). *Psychological Bulletin*, *138*(3), 415–422. http://doi.org/10.1037/a0026884

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

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580–593. http://doi.org/10.1002/wcs.142

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. http://doi.org/10.1016/j.tins.2004.10.007

Klein, C. (2016). What do predictive coders want?. *Synthese*, 1-17.

Lamme, V. a F., Super, H., Speckreijse, H. (1998). Horizontal and Feedback Processing in the Visual Cortex. *Current Opinion in Neurobiology*, *8*, 529–535.

Lee, T., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *20*(7), 1434–48. http://doi.org/10.1364/JOSAA.20.001434

Ma, W. J., & Jazayeri, M. (2014). Neural Coding of Uncertainty and Probability. *Annual Review of Neuroscience*, *37*, 205–220. http://doi.org/10.1146/annurev-neuro-071013-014017

Mamassian, P., Landy, M. S., Maloney, L. T., Rao, R., Olshausen, B., and Lewicki, M. (2002). Bayesian modelling of visual perception. Probabilistic models of the brain: Perception and neural function, pages 13{36.

Rao, R. P. N., & 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. http://doi.org/10.1038/4580

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

Spratling, M. W. (2008). Reconciling Predictive Coding and Biased Competition Models of Cortical Function. *Frontiers in Computational Neuroscience*, *2*(4), 1–8. http://doi.org/10.3389/neuro.10.004.2008

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, *112*, 92–97. http://doi.org/10.1016/j.bandc.2015.11.003