Affective Value in the Predictive Mind

Sander Van de Cruys

Although affective value is fundamental in explanations of behavior, it is still a somewhat alien concept in cognitive science. It implies a normativity or directionality that mere information processing models cannot seem to provide. In this paper we trace how affective value can emerge from information processing in the brain, as described by predictive processing. We explain the grounding of predictive processing in homeostasis, and articulate the implications this has for the concept of reward and motivation. However, at first sight, this new conceptualization creates a strong tension with conventional ideas on reward and affective experience. We propose this tension can be resolved by realizing that valence, a core component of all emotions, might be the reflection of a specific aspect of predictive information processing, namely the dynamics in prediction errors across time and the expectations we, in turn, form about these dynamics. Specifically, positive affect seems to be caused by positive rates of prediction error reduction, while negative affect is induced by a shift in a state with lower prediction errors to one with higher prediction errors (i.e., a negative rate of error reduction). We also consider how intense emotional episodes might be related to unexpected changes in prediction errors, suggesting that we also build (meta)predictions on error reduction rates. Hence in this account emotions appear as the continuous non-conceptual feedback on evolving — increasing or decreasing — uncertainties relative to our predictions. The upshot of this view is that the various emotions, from "basic" ones to the non-typical ones such as humor, curiosity and aesthetic affects, can be shown to follow a single underlying logic. Our analysis takes several cues from existing emotion theories but deviates from them in revealing ways. The account on offer does not just specify the interactions between emotion and cognition, rather it entails a deep integration of the two.

Keywords:

Affect | Dark room | Emotion | Prediction error | Predictive processing | Reward | Uncertainty | Valence

Acknowledgements

Sander Van de Cruys is a postdoctoral fellow of the Research Fund Flanders (FWO). Partly supported by a Methusalem grant by the Flemish Government (METH/08/02) to Johan Wagemans. I want to thank Johan Wagemans, Jan Lauwereyns, Andrey Chetverikov, Agnes Moors, Tomer Fekete, Chris Trengove, Omer Van den Bergh, and three anonymous reviewers for helpful comments on an earlier draft of this paper. I am very grateful to the organizers and the participants of the MIND23 workshop (Frankfurt, May 2016) for inspiring discussions, and to the editors and three anonymous reviewers for constructive comments. The views expressed (and any remaining errors) are solely my own responsibility.

> *Happiness is neither virtue nor pleasure nor this thing nor*
> *that, but simply growth. We are happy when we are growing.*
> *— W.B. Yeats*

In his seminal work on the relation between cognition and emotion, Robert Zajonc (Zajonc 1980) wrote that "preferences need no inferences". If this were true, affect would never find a place in the currently popular attempt at a unified model of cognition, called predictive processing (PP), that holds that inference is all the brain does (Clark 2013; Friston 2010). Indeed, there is so far little work on affective value and affective experience in PP (but see Barrett and Simmons 2015; Seth 2013). Nevertheless, PP has in recent years been shown to hold a lot of promise in blending perception, action and cognitive beliefs into a coherent, well-founded framework, pleasingly taking down the walls between these subfields. Although plenty of fundamental issues concerning its computational articulation and biological implementation remain (e.g., see commentaries on Clark 2013), as a unified theory of cognition it arguably fares better than any other alternative we have. Crucially, if it is to become an overarching framework of the mind-brain, emotions, so central to existence and survival, somehow

Van de Cruys, S. (2017). Affective Value in the Predictive Mind.
In T. Metzinger & W. Wiese (Eds.). *Philosophy and Predictive Processing:* 24. Frankfurt am Main: MIND Group. doi: 10.15502/9783958573253

have to fit in, including their experiential aspects. However, the unifying logic of PP—a single computational principle for the whole brain—seems directly opposed to the popular notion in emotion theorizing that emotions are a bricolage of modules adapted to very specific challenges in our ancestral environment. Rather than built around a single neat, optimal logic, emotions are assumed to be a messy, ad hoc bag-of-tricks. In practice, it has, however, proven difficult to distinguish different emotion 'modules' in the brain, even at the subcortical level, which has led some emotion theorists (e.g., Barrett 2014; Carver and Scheier 1990; Moors 2010; Russell 2003) to abandon this route in favor of a view that assumes fewer fundamental affective 'building blocks'. The aim of the current paper is to show that this movement may afford new ways to integrate emotion in PP. Much of this is, as we will see, thanks to the clear evolutionary rationale that is at the core of PP. At first blush, this may seem to lead to a concept of value or emotion that seems rather alien or counter-intuitive to how we usually think about emotional value, but it will turn out to have much in common with existing theories of emotions. Most importantly, cognition (information processing) and emotion will be shown to be entangled from the start.

First a brief note about terminology. Throughout this paper, we will generally use the broader term "affect" because it does not constrain our explanandum to so-called "basic emotions", to conscious "feelings", to "moods" or to "reward". While we will suggest how to differentiate between these affective concepts in the course of the paper, we assume a basic "affective value" that is shared by all of these. It is the cause of this core of emotions that we attempt to explain. The exercise we undertake in this paper is to examine how the affective dimension of our mental life can be understood *without* positing more principles than those provided by PP. To anticipate a key thesis of this work, we propose the origin of emotion does not lie in being able to infer or predict (the causes of) bodily changes, as accounts of emotion as "perception of the body" argue (Barrett and Simmons 2015; Seth 2013). Rather, it is situated in how the brain succeeds or fails to do so over time (i.e. prediction error dynamics). We do not deny the importance of bodily arousal in the resulting emotions, but we identify error dynamics as the fundamental cause of emotions. Given that, following PP, prediction errors are ubiquitous processing products, the implication will be that emotions can emerge from any processing, not just that about the body.

1 Predictive Processing

PP holds that an organism is constantly, proactively predicting the inputs from its environment. Since it has no independent access to objective features in the world, all an organism can do is learn patterns in its input generated by statistical regularities in the environment and by its own actions (Clark 2013). While in principle there may be different ways in which prediction could modulate perceptual processing, PP proposes a well-defined computational scheme and a single guiding principle (Friston 2010). The scheme posits that every level of the perceptual hierarchy predicts activity in the level below, in effect explaining away input that is consistent with it such that only mismatches remain. These mismatches, called prediction errors, are sent upwards to update future top-down predictions. Much of the brain's predictive activity has a limited time frame. It predicts current input by inferring assumed causes of these inputs. The higher up in the hierarchy, the more time and space predictions can span, because they can work with regularities defined on lower levels. In this way lower level predictions model the faster changing dynamics, while those higher up track and recreate slower changing dynamics.

PP thus completely inverts the classical bottom-up view of the perceptual hierarchy. The brain actively generates the perceptual world (predictions are based on generative models, i.e., models that can generate the input), and perceptual input is in fact the feedback on how good these constructed models are. Although anatomically prediction errors are conveyed by feedforward connections, functionally they are the feedback, sanctioning the models we construct of the outside world, in line with the view of perception as controlled hallucination (e.g., Horn 1980). The fundamental underlying principle guiding this process of iterative, hierarchical matching of predictions with inputs is that of prediction error minimization (PEM). Perception is inference to the best prediction, the one that minimizes prediction errors. Simultaneously, learning will use remaining prediction errors to home in on the best predictions for the current context, thereby reducing future prediction errors. Hence, we perceive what led to successful predictions in the past (see also Purves et al. 2011).

1.1 Action

There is one other, complementary way of minimizing prediction errors, which does not focus on improving predictions, but rather on modifying the things predicted, through action. In this framework, movements serve to bring the input closer to our prior expectations (often called 'active inference'). More specifically, they are induced by their expected exteroceptive and proprioceptive consequences (Friston et al. 2010), much in line with James' "anticipatory image" (James 1890) and with the ideomotor principle (Hoffmann 2003). Like object-level, conceptual predictions ("an apple") unpack to a myriad of lower-level featural predictions ("green", "curved", …), so can high-level expected states ("goals") be unpacked in specific component predictions and eventually in expected proprioceptive states. When the latter are compared to afferent signals of muscle spindles at the spinal level, they generate sensory prediction errors to be reduced by motor neuron activation, in a classical reflex arc. Hence, motor commands are replaced by expectations about the state of proprioceptive sensors. At a higher level these 'commands' stem from beliefs about state transitions (active inference). A certain perceptual stimulus may be predictive of a state transition through the agent's intervention (an affordance, if you will), that can be actualized by unpacking this prediction to proprioceptive states.

Bear in mind that, from this inference system's perspective, there is no intrinsic difference between the external and the internal milieu. With the same predictive machinery, generative models can be learned about changes in interoception, based on input from somatovisceral sensors (Seth 2013). Likewise, internal 'actions', such as autonomic responses, are brought about by similar principles as 'external' actions. They consist of changing a bodily set-point or expectation (e.g., temperature) so autonomic reflexes (e.g., shivering) can be elicited.

We limit ourselves to this very brief sketch of PP and refer to the many in-depth resources for more details about its computational mechanisms and how these may map onto neural circuits and their plasticity (Bastos et al. 2012; Friston 2003, Friston 2010). Further implications of the framework will be discussed to the extent that they connect to value and emotional relevance.

1.2 Prediction and Homeostasis

Organisms maintain their own organization (order) in the face of constant fluctuations in the environment through homeostasis. The bioenergetic regulation of homeostasis in essence makes sure that the organism is bounded in the physiological states it can be in, which allows it to resist the dispersive effect of the second law of thermodynamics (Schrödinger 1992). This can be considered the most fundamental goal of any organism, though of course, it is not an intentional one. It is just a result of the fact that organisms that do not tend to homeostasis will lose existence as a unit. One can view homeostasis as a limited set of *expected* interoceptive states 'discovered' by evolution, because they have proven to enable continued existence (Friston 2010; Pezzulo et al. 2015). Therefore, survival "depends upon avoiding surprising encounters and physiological states that are uncharacteristic of a given phenotype" (Friston 2009).

The key problem is that organisms generally do not have direct ways to change these internal states (Pezzulo et al. 2015). They may be able to shiver when their body temperature drops, but they cannot replenish glucose levels without interaction with their environment. This implies that they actually need to build a generative model for these expected internal states, by learning about the different ways these internal states can be 'caused'. It is only because the organism needs to go through its environment to fulfill the interoceptive expected states, that it needs (exteroceptive) perception and action.

It needs perception to infer hidden states of affairs in the world that may cause the expected internal states. And it needs action to control those states of affairs. As hinted in the section on action above, actions are also represented as predictions or beliefs, specifically about transitions that one happens to be able to control (control beliefs). As we saw earlier, this generative model, in service of the body, is constructed with the PP scheme. The models needed for this may be rudimentary and fixed or complex and flexible, depending on complexity and volatility of the organism's *Umwelt*. For some organisms the causal chain of expected internal states might go through very high-level and context-dependent states (e.g., social interactions). At that point, it pays to build deep models incorporating this strong contextualization of interoceptive states (Pezzulo et al. 2015).

Simpler organisms may not need such strong, flexible contextualization, because internal states are reliably caused by stable states in their environment. Even in that case homeostasis is not actually static or merely reflexive. If reliable predictive information is present, it is more efficient to anticipate changes with compensatory action (Heylighen and Joslyn 2001). For example, when the single-celled gut bacterium E. coli is ingested by mammals it will respond to the temperature shift by not only upregulating heat shock genes (to compensate for temperature) but also by downregulating genes for aerobic respiration. They use the temperature information (when entering the mouth) to predict that they will end up in a low oxygen environment, i.e., gastrointestinal tract (Freddolino and Tavazoie 2012). They encode something about the cause of this particular stimulation (ingestion), however rudimentary. As Freddolino & Tavazoie (Freddolino and Tavazoie 2012, p. 369, my emphasis) describe: "microbial behaviors are as much responses to the *meaning* of a specific environmental perturbation (viewed in the context of their evolved habitat) as they are responses to the direct consequences of that perturbation".

For E. Coli the predictive "learning" of these regularities does not take place within the organism but within populations. Through natural selection the predictive environmental relation can be embodied in the molecular regulatory networks of the cell (Freddolino and Tavazoie 2012). Environmental regularities left their imprint on the organism's constitution, just because a constitution embodying these regularities increases fitness in a Darwinian sense. In analogy to PP, evolution can be considered as an error-correcting code, except that the errors are not represented at the level of a single organism[1]. But note that the normative character—the value—originates in the organism (that maintains its internal states better or worse), not in the process of evolution (Deacon 2011). The boundedness in the homeostatic set, the 'mother-value of all values' (Weber and Varela 2002) also gives the whole predictive endeavor its normativity. Once the organism engages itself to make a prediction, there is something at stake (cf. value), because of the link from the quality of predictions to basic organismic functioning. There is a vested interest for the prediction to materialize. In complexer animals, "the gross bodily form, biomechanics, and gross initial neural architecture of the agent all form part of the (initial) 'model' and […] this model is further tuned by learning and experience" (Friston et al. 2012a), using the general-purpose PP mechanisms. Reducing prediction error can be a proxy for fitness, because prediction error minimization is the proximal, local mechanism that makes sure that in the long run organisms stay within physiological bounds (Friston 2010).

2 Prediction and Reward

The fact that predictive models are grounded in homeostasis by no means implies that only predictions about favorable outcomes can be formed. For example, even though some perceptual predictions may seem not to be consistent with biologically expected states (e.g. "the rattling that I hear is caused by a snake nearby"), it is all the more important to make them accurately, and not hallucinate more

---

[1] Recent work suggests there is a deeper, formal equivalence between PP and evolutionary population dynamics (Harper 2009).

agreeable alternatives. A negative stimulus only has really detrimental consequences for survival if the system was not able to adequately prepare for it, by marshaling the necessary compensatory mechanisms—if necessary by acting to avoid it. Once this is taken care of, what could be a threatening stimulus for bodily integrity, becomes harmless. Conversely, predicting (and hence preparing for) a future negative stimulus that turns out not to occur, is often very wasteful for an organism. So we see that there are good biological reasons for why prediction confirmation should be good, while failures should be bad. In fact, even an appetitive stimulus such as food could, for an unprepared body, be unpleasant.

In a PP account, reward stimuli are just expected or familiar sensory states (Friston et al. 2012b). Intuitively we feel that we avoid punishment or seek reward and therefore visit these states less or more frequently, respectively. PP turns this intuition around, describing frequently visited states as rewards because they are expected. The reward value of a stimulus can be defined as the frequency with which it is chosen (Moutoussis et al. 2015). Rewards do not "attract behavior", but attainment of rewards is the result of prediction error minimization, exactly as described for perception and action in general. Specifically, while in classical reinforcement learning goal-directed decision making consists of finding the policy that maximizes expected reward, when framing it in terms of Bayesian inference one assumes reward attainment and finds the policy (state-action pair) that best explains or causes that effect (see Schwartenbeck et al. 2014; Solway and Botvinick 2012; Srivastava and Schrater 2015). If one redescribes utility of outcomes as prior beliefs about states one will end up in, one can use the same PP machinery to minimize errors along the road to the expected state. This boils down to building a generative model of rewards (same as for any other stimulus). Importantly, it requires that we have prior beliefs about what the world will be like and about expected final states or goals (Moutoussis et al. 2014). The latter are the alternative outcomes that we expect to be reachable with policies we can apply. The key is to reduce the discrepancy between the likely and the expected outcomes. Note that within this approach, one could still make a distinction between "greedy" and epistemic actions (Friston et al. 2015). Greedy or pragmatic actions use prediction errors to directly fulfill expected "rewarding" states, e.g. food consumption. This is possible when there is little uncertainty about the path leading up to the expected state. In case there *is* considerable uncertainty, epistemic actions are directed at acquiring more information, that allows greedy action in the future. This implies that action may lead to increase of the distance to a goal (prediction error), in order to move to a familiar position where one can approach the goal with larger certainty (Friston et al. 2015). However, any actual behavior will have a combination of pragmatic and epistemic elements, with prediction errors as the common currency.

The repertoire of innate expected states is specified and extended by learning throughout an animal's life. In fact, within this view, there are no distinct reward or punishment stimuli (Friston et al. 2009; Wörgötter and Porr 2005). Any sensory signal has a cost, namely the prediction error. It tells something about the success (failure) of the generative model we used for predicting the input. This also implies that habits or 'rituals', i.e., predictable sequences of behavior, are in fact a form of reward. There is usually no tremendous pleasurable experience to habits (we will come back to this point later on), but not performing habits when the appropriate eliciting context is present seems to produce some negative affect. It speaks to the self-sustaining nature of habits (Egbert and Barandiaran 2014; Egbert and Canamero 2014). Indeed, for over-learned behavioral patterns, devaluation of the reinforcer that was originally used to establish the behavior will not lead to reduction in behavior (Wood and Neal 2007). The wider implication is that organisms do not only preserve their life (homeostatic predictions) but also their *way of life*, as a set of expected (preferred) behaviors (Di Paolo 2003).

2.1 Problems with the Classical Reward Concept

Several developments started eating away at the concept of reward as absolute, stable representation of utility, that guides decision making. We highlight two theoretical problems, and two empirical ones.

First, Friston et al. (Friston et al. 2012b) criticize the inherent circularity in the definition of reward. Reward is often defined as a stimulus that elicits (reward seeking) behavior (Schultz 2007). Evidently, one cannot invoke rewards to explain that same behavior later on. Second, recent theorizing suggests that rewards and punishments are always subjective and internal, meaning they "are constructions of the subject rather than products of the environment" (Dayan 2012, p. 1089). They are dependent on the position relative to expected states. Reward is not something in the environment, much less an external critic such as often assumed in computational reinforcement learning (see a similar critique in Singh et al. 2009). Psychological theories too, often incorporate a semi-hidden homunculus. Here, the value (or cost) function applied to perceptual or cognitive output hides an 'evaluator', an unanalyzed 'agent' that can assign the values, within an allegedly objective, quantifiable construct. Misled by our intuition that rewards are self-evident, these homunculus remnants too often go unquestioned.

Third, Chater & Vlaev (Chater and Vlaev 2011) argue on empirical grounds that, similar to sensory judgment in psychophysics, value is a not represented as an absolute magnitude but rather as a comparison, relative to the local context. Chater & Vlaev conclude that "to the extent that people have a grasp of their own, more global, values, this must be inferred from sampling their own past choices and other memories, thus revealing their preferences" (Chater and Vlaev 2011, p. 96). In other words, humans can easily infer the reward value based on experience sampling, but these values are constructed predictions that best explain the sampled experiences. Generally, however, these representations are not necessary to enable adaptive behavior.

Finally, after conditioning reward-modulated activity is found throughout the visual hierarchy, including the primary visual cortex and the lateral geniculate nucleus (Gershman and Daw 2012; Serences 2008). Conversely, "neutral" perceptual prediction errors elicit activity in striatal and midbrain regions, usually connected to reward/punishment and motivational functions (e.g., Schiffer et al. 2012; Iglesias et al. 2013; den Ouden et al. 2009). Other studies find that dopamine neurons also code for sensory uncertainty in a rewarded sensory decision-making task (de Lafuente and Romo 2011), possibly because expected reward will decrease when sensory uncertainty increases (Bach and Dolan 2012). In sum, these developments indicate that the strict segregation of probabilities (perceptual processing) and utilities (cost-reward processing) is untenable (Gershman and Daw 2012), and suggest that the PP concept of reward merits further examination, because it does not suffer from these shortcomings.

2.2 Pleasant Surprises and Other Objections

The PP framing of reward does not mean that learning or behavior is not as constrained as in conventional models of reward and punishment. To take the extreme example, even if at the agent-level a pain stimulus is perfectly expected, across all levels of predictions this will never become an expected state. Tissue damage can be seen as a violation of a bodily expected state (bodily integrity) that is not compatible with continued existence. On the other hand, this approach has no difficulty explaining why humans seem to find reward in endlessly varying idiosyncratic 'niches', based on the wide flexibility in predictions they can generate.

An obvious counter-argument to the thesis that prediction errors always have a negative value is the existence of 'pleasant surprise', e.g. when one receives an offer that is better than expected. However, even in such cases there is some evidence in humans and monkeys that the initial reaction to prediction error or surprise is generally negative (however short-lived) (Knight et al. 2013; Noordewier and Breugelmans 2013). But note that the agent-level emotion of surprise encompasses more than a single, momentary prediction error (surprisal) at some level of the brain. As we will see in later sections, the dynamics of the failures and successes in prediction are more important here.

Still, intuitively, the identification of prior probabilities (predictions) with utilities seems wrong-headed. For example, Gershman & Daw (Gershman and Daw 2012, p. 306) ask: "Should a person immersed in the 'statistical bath' of poverty her entire life refuse a winning lottery ticket, as this would necessitate transitioning from a state of high equilibrium probability to a rare one?" To start to defuse this argument, one has to acknowledge that in such complex cases there is not just one prediction (e.g., of poverty) at play, but rather a complete predictive hierarchy. The person growing up in poverty does not lose his or her expectation to be well-fed and to provide for kin. There might be interiorized social expectations as well, that could also urge the person to accept the winning lot. That said, once accepted, the new situation may create quite some prediction errors given a predictive system unadapted to that new state of affairs (indeed, most lottery winners like to continue their life, including job, as before; Kaplan 1987). Later sections will hopefully shed a different light on these forms of 'upward mobility' (Gershman and Daw 2012).

A related counter-intuitive idea in this proposal is the lack of distinction between core concerns, desires, needs or goals of an organism versus just any predictions. Part of the answer has to be found in the hierarchy, with likely states being about what happens when I see (or do) this. In contrast, desired states are about what I can, more abstractly, expect given my biological constitution, experience and sensorimotor capacities (i.e., predictive models)[2]. For example, while I, at some level, expect a food reward (desire), the way to "generate" this involves lower-level sensorimotor predictions about states and state transitions (some of which we control), that are navigated through by PEM. The assumption is that some expectations (e.g., about homeostatic states that, for example, a food reward can fulfill; see the section on homeostasis) are hardwired, installed by evolution. So another part of the answer must lie in the lower flexibility of goals/desires, compared to sensorimotor predictions which are flexibly updated when inputs change. However, because the whole hierarchical model is relevant to the organism (if it was not it would not have been formed), any prediction has cognitive (belief) and conative (desire) elements which cannot be disentangled (Millikan 2004). Predictions are never motivationally neutral, but represent states and direct behavior at the same time. More work will be needed to explain our intuitive distinction between likely and desired states if both are ultimately predictions, but the hierarchy is bound to play an important role. The lower level likely states mostly pertain to faster changing dynamics in inputs (regularities in shorter time frames). The higher level desired states link to slower changing dynamics (e.g., 'I am a good person'). If evidence mounts that undermines the latter type of predictions, a full-blown existential crisis may occur. Luckily, there are often ways to shield such prediction errors, i.e., to explain them away with 'auxiliary predictions'.

At this point one might object that reward (or positive affect) as defined so far is too 'conservative' a concept[3]: we basically aim to return to familiar, overlearned states or situations and resist anything that deviates from those expected states (but see dissonance or conflict theories of emotion; Mandler 2003; Festinger 1962; Hebb 1946). This approach may explain the familiarity bias (cf. the mere exposure effect) that is often reported (Lauwereyns 2010), but it does not even remotely seem to capture our *experience* of reward in general. We easily get bored—a loss of reward value—with very familiar or repetitive stimuli. More so, we seem to actively explore departures from well-trodden paths and expected situations. How do we explain that our motivations often lie outside of predictable ruts? And how can we more fully account for rewards as hedonic, pleasurable experiences derived from these different situations? That is what we will discuss next.

3 Affective Valence

If we define drives as prediction errors or discrepancies between current and expected state (Keramati and Gutkin 2011), we end up a new way of looking at the affective, experiential value of rewards as

---

[2] At least in humans this seems to have an important social comparative component as well: our predictions are formed based on what people that one considers to be similar to oneself, could attain.

[3] But it is far from a passive notion: to keep the organism within some expected range of a variable often means elaborate and vigorous activity (cf. allostasis; Egbert et al. 2013).

well. Reward value is directly dependent on drives in the sense that the reward value of, say, a drop of water depends on the internal drive state of the organism (e.g., a thirsty rat). It is easy to see that what is critical then is the change in prediction errors (drive states). Hence, positive value is defined as a decrease of prediction errors, while negative value can be equated with an increase in prediction errors.

More generally, we propose that the affective valence is determined by the change in (or first derivative of) prediction error over time (Joffily and Coricelli 2013; Van de Cruys and Wagemans 2011)[4], with positive valence linked to active reduction of prediction errors, and negative to increasing prediction errors. This makes sense because these temporal dynamics signal whether the organism is making progress (or regress) in predicting its environment, which in the long run translates in proper functioning of the processes of life (fitness) (Damasio 2003). It is easy to see that the reward value of food very much depends on how large the prediction error initially was (i.e., how hungry you were), and hence how big a change the food consumption induced, but we propose this is a general pattern.

Importantly, emotional valence is not something added to these error dynamics, it is those dynamics. They are a reflection of quality of processing, so they do not have to be evaluated in turn. We connect positive and negative affect here to general purpose processing characteristics, detached from particular utility or motivations. They are purely determined by how the organism interacts with its environment (see also Polani 2009).

This goes beyond the simple view that prediction confirmation results in positive affect, while violations of predictions are negative. Once homeostasis, rather than being reactive, relies on predictive models, errors often do not have direct effect on homeostasis (or fitness). It then becomes equally important to monitor prediction error dynamics, as it is to monitor the errors as such. Mere presence of instantaneous prediction error does not seem to be an adequate basis of emotional valence. Positive affect might still occur for a large instantaneous error as long as this error is (or has been) in the process of being reduced.

It is no stretch for humans to imagine that making progress in predicting various sensorimotor domains can be very rewarding (e.g. see Hsee and Ruan 2016). More challenging is to show those 'informational' rewards in nonhuman animals. However, Bromberg-Martin & Hikosaka (Bromberg-Martin and Hikosaka 2011) have managed to show that monkeys too are prepared to work to receive cues that reduce their uncertainty (reduce errors), even though their choice had no influence whatsoever on the actual reward subsequently received. The animals even chose the information cue more consistently than they typically choose a high probability reward over a low probability reward (Niv and Chan 2011). Moreover, these informational gains elicited dopaminergic neural activity in midbrain regions similar to that for conventional rewards. Our account would predict that such effects generalize to other animal species, but of course, for there to be changes in prediction errors there need to be predictions formulated. Therefore, the specific instances of predictive gain will depend on the kind of models an animal constructs about its world.

Behavioral testing of these ideas is challenging because these dynamics are subject to learning and because it can be difficult to determine the predictions participants apply. Suggestive evidence comes from a recent study looking at the affective consequences of conflict resolution (Schouppe et al. 2014). These authors build on the priming study by Dreisbach & Fischer (Dreisbach and Fischer 2012) which showed that incongruent Stroop stimuli, as opposed to congruent ones, can prime people to more quickly evaluate negative words or pictures than positive ones (an indirect measure of negative affect). Schouppe et al. (Schouppe et al. 2014) report that, while incongruent stimuli are indeed aversive, once they are successfully solved more positive affect will follow than for congruent stimuli. The original prediction error (conflict) seems conducive to later reward from resolution, consistent with what we

---

[4] Note that the model by Joffily & Coricelli strictly speaking is not about prediction errors but rather about the more general concept of (variational) free energy.

propose here. Future studies should attempt to induce, violate and resolve new predictions in the lab to see if these dynamics have the hypothesized emotional effects.

3.1 Specifying Predictive Progress

Even though the current view entails that emotions can arise wherever errors are compared, there are good computational and ecological reasons why change in errors is computed and compared within the limits of one and the same input domain. Comparing errors from very different perceptual levels or sensorimotor situations would be very demanding to the system, and, more importantly, unproductive. As Oudeyer, Kaplan & Hafner (Oudeyer et al. 2007, p. 8) remark with regard to an artificial agent, such a system may "attribute a high reward to the transition between a situation in which a robot is trying to predict the movement of a leaf in the wind (very unpredictable) to a situation in which it just stares at a white wall trying to predict whether its color will change (very predictable)." PP proposes that specialization (functional segregation) in the brain stems from conditional independence of different representations—representations that have predictive relations organize into regions with tight interconnections (Friston et al. 2013; Stansbury et al. 2013). This architecture may also be used to evaluate changes in errors relating to predictions that actually belong to the same domain.

Predictive progress has already been used to understand and implement intrinsic rewards in the domain of artificial intelligence (Kaplan and Oudeyer 2007; Schmidhuber 2010). More recently a decrease in prediction errors (or equivalently a predictive learning gain) was assumed to underlie intrinsic rewards in humans as well (Kaplan and Oudeyer 2007). Agents that at each point try to maximize predictive progress, will avoid losing time in regions of sensorimotor space that are too difficult to predict with the current capacities and regions that do not contain any learnable differences anymore, either because the domain is known or because what is left is noise variation. Hence, they will automatically focus on situations and stimuli that contain learnable differences, just above their current state of predictive knowledge, where the largest gain can be made. This guiding principle enables the agent to explore and proceed through stages of increasing predictive difficulty ('developmental phases').

There is some debate about the extent to which such an imperative to maximize prediction error reduction and PEM are one and the same thing (Clark 2013; Froese and Ikegami 2013; Little and Sommer 2013). Proponents of the 'maximizing learning gain' position contend that an organism driven by PEM will seek a dark room and stay there, because prediction error is maximally reduced there. However, a dark room is not actually a maximally expected situation, or does not stay so for long, in a PP framework (see also Friston et al. 2012a). Prediction errors are always computed relative to an agent's possibly complex, embodied model, with its specific organism-defining expectations, quickly rendering the dark room unexpected. While this seems to answer the 'negative' objection (why not stay in the dark room), can PEM also fully account why we humans 'positively' seek out prediction errors? This seems to depend on the kind of multi-level and second-order predictions we generate. As an example, if, at an abstract level, you expect yourself to be friendly, confirmation of this prediction will sometimes entail prediction errors on other, possibly lower levels. The key is to predict the violations as well, such that their impact can be reduced (see discussion on precision above). Similarly, if you expect to be a good darts player, you will need to tolerate some lower level sensorimotor errors to get there, usually because you can also reasonably expect the errors encountered to be reducible, based on previous experience. In short, a good predictive agent will always expect to be surprised. We seek prediction errors that are reducible, given our models, including the actions (as beliefs about inputs we control) we can rely on.

3.2 Non-Conceptual Metacognition

An operation performed on the prediction errors can be considered a form of metacognition. Similar to precision, the temporal comparison of prediction errors is a second-order operation. In the first-order process, prediction errors are information used to update predictions, while in the second-order process the prediction errors as outputs of the first order process are in turn compared in time, which provides new information that, we argue, is phenomenally experienced as valence and that may become available for processes beyond the predictive chain that created the errors. The result is a form of nonconceptual information about uncertainty that increases or decreases in the current situation. It is not about the (propositional) content but about the content-forming processes. The thesis here is that emotions are the qualitative experience (quale) of this kind of nonconceptual information. In a related view, Reisenzein (Reisenzein 2009)[5] argues that emotions non-conceptually convey important changes in an experiencer's belief system in interaction with the world. This is indeed what prediction errors signal. Their dynamics are a form of feedback on the system's own functioning as it deals with external and internal challenges, so a conception of affect as a continuous "neurophysiological barometer of the individual's relationship to an environment at a given point in time" (Duncan and Barrett 2007) is nicely consistent with this. Similarly, Frijda (Frijda 2006, p. 82) notes: "pleasure is the positive outcome of constantly monitoring one's functioning". For affect and motivation, the attainment of the "object" is of lesser importance, considering that predictive progress is zero then. This provides an interesting perspective on what Cantor and Kihlstrom (Cantor and Kihlstrom 1987, p. 179) called the paradox of goal-setting, namely "that people are often less intrigued or impressed with an end-state the closer they come actually to achieving it".

In the current view, the non-conceptual information is available in terms of the positive or negative affective tone of experiences. Note the connection with the concept of cognitive or perceptual fluency (Reber et al. 2004) as the ease with which stimulus material is processed. In its different operationalizations (e.g., by increasing the symmetry and contrast of visual stimuli, or the readability of words), it has been repeatedly shown to positively affect the appreciation of stimuli. Fluency should also be seen as a metarepresentation (Alter and Oppenheimer 2009) and is arguably well characterized as the experience of actively reducing prediction errors (and disfluency as increases in prediction errors). Moreover, if one identifies emotion with the way of processing rather than end-products, perceptual (dis)pleasures and 'proper' emotions might be subsumed under the same principles. Specifically, (dis) fluency with regard to approaching high-level goals or biological concerns (bodily expected states) is what we usually associate with emotions. This idea is barely new. In a very influential control-theoretic approach to emotions, Carver & Scheier (Carver and Scheier 1990) linked dynamics in mismatch between goals and actual state of affairs to dynamics in emotion. They described how multi-level goals should be interpreted as hierarchical reference values, from abstract idealized goals (e.g., having a self-image of a good person at the highest level), to more concrete actionable expectations (e.g., shoveling snow off of walks). In PP terms, actions have to make sure that the agent can harvest the inputs that conform to "trickled down" expectations. So, analogously to PP, these expected values can generate errors at every level. Our own actions (or external circumstances) cause changes over time in discrepancies relative to these values. Carver and Scheier already argued that emotion is about monitoring the rate of discrepancy (prediction error) reduction, as we propose above. However, their analysis suggests a pertinent extension of what we presented so far. They suggest that the rate of mismatch reduction is in turn subject to a control loop, comparing actual with expected rate of change. Only when the current rate of prediction error reductions deviates from the expected rate of reduction, so Carver and Scheier argue, one experiences emotion. This will of course be positive affect if the rate of progress to the goal is higher than expected, negative if it is lower than expected.

---

[5] Note that in Reisenzein's theory desires still have a status categorically different from other beliefs.

Based on PP, this makes a lot of sense. As we described, prediction error minimization is the way we perceive and act, so we are reducing errors all the time, e.g., when we successfully use our sensorimotor system to walk the street. Generally, little positive or negative emotion is involved despite these constant error reductions. This may mean that these changes in sensorimotor errors are not large enough, but most likely what rate is substantial depends on the expected rate of reduction for the current sensorimotor context. Where do the predictions of rates come from? These might very well be contextually learned through the same predictive machinery as for 'first-order' predictions. In fact, PP already includes second order expectations about precisions, which can be considered as (inverse squared) expected prediction errors (Mathys et al. 2014). The higher order prediction errors that are used to update these expected precisions (so-called volatility prediction errors or VOPEs) compare predicted prediction errors (predicted uncertainty) with observed prediction errors (actual uncertainty). For example, if first level prediction errors (also called value prediction errors or VAPEs[6]) are reduced, but the corresponding VOPEs do not decrease, the error continues to be lower than expected, possibly providing a basis for positive affect. It still needs to be clarified to what extent the temporal derivatives and expected rates as described here can be realized using expected volatilities per se (Joffily and Coricelli 2013), but these developments at least suggest learning about such second order states is possible within a PP system.

However, genetic factors might also contribute to these expected error reduction rates. Individual differences in expected rates of error reduction may account for certain dispositional affective traits. Indeed if the predicted rate of progress is set too high, an individual will tend to experience more negative affect than positive, because the prediction will rarely be matched (Carver and Scheier 1990). This may happen, even if this person's actual rate of progress is very high. Furthermore, if the expected rates of progress are indeed at least partly learned specifically for different sensorimotor situations, this may constitute a form of emotion regulation. Specifically, the system may, through updating the expected rates, remain within a given range of emotional experience by adapting this criterion of expected rate of change (the neutral point).

Interestingly, once an agent can track and learn to expect certain rates of change in prediction errors, it arguably will show a distinct propensity to explore and learn. This continuous, active search for reducible prediction errors (satisfying an expected error reduction rate) may in turn have enabled the development of rich social relations and culture. As such, this may form another counter-argument for the dark room objection against the principle of PEM: There will never be a stationary stimulus or situation satisfactory for an agent that expects some non-zero rate of prediction error minimization.

3.3 Varieties of Affect

Looking back, we have first encountered reasons to attribute emotion to prediction errors (mismatch) or confirmation as such, then we have shown it may be better attributed to changes over time in prediction errors, and finally to errors about expected rates of change. Importantly, these three can be independent. Borrowing an analogy from Carver & Scheier (Carver and Scheier 1990); if we make the parallel with distance, speed (first derivative of distance over time) and acceleration (second derivative), we can see that any rate of progress can be associated with any instantaneous prediction error, and further any change in rate of progress can co-occur with any instantaneous rate. The rate of error minimization seems to provide the necessary signal for valence. However, in mammals, especially humans, rate may be subjected to predictions of its own, moving important emotional dynamics to that level. Still, rate may determine the continuous hedonic tone of what is sometimes called 'background emotions' (Damasio 2000). A steady rate of progress may induce a diffuse feeling of well-being, a sense of properly functioning bodily and sensorimotor systems, akin to what is sometimes described as experience of flow (Csikszentmihalyi 1996). The usually brief episodes of (intenser changes in) emotions

---

[6] This is not about emotional value, but about a quantitative mean value of the estimated state.

in our daily life, i.e., emotions as commonly understood, seem linked to unexpected changes in rates of progress.

Emotions are notoriously volatile, comparative, and subject to habituation (Frijda 2006 [1988]). These characteristics naturally follow from the current framework. By definition, prediction errors and their temporal dynamics are dependent on learning. Pleasures from increased rates of predictive progress only last as long as this progress is possible. Kaplan & Oudeyer (Kaplan and Oudeyer 2007) note that "progress niches are nonstationary". Meanwhile, the contrastive property of emotion entails that a suboptimal state (sizable prediction error) may still be pleasurable depending on the starting position, because there is a positive, possibly higher than expected, rate of error reduction. Emotions emerge as situated in perpetually moving regions of state space in a system that grounds them in predictive dynamics. Rather than being associated with particular "target" objects or states, they are the concomitant of successful (or unsuccessful) striving (Duncker 1941). Note that this type of system does not aim to maximize the frequency of positive affect (nor would that be particularly adaptive; Carver and Scheier 1990). Rather, it may redistribute frequencies of positive and negative affect so as to preserve the range.

One might object that the view we propose runs the risk of 'intellectualizing' emotions. Indeed we essentially described affect as a specific form of cognition. In that sense, it is somewhat related to previous emotion theories about the 'need to resolve uncertainty', most aptly formulated by Kagan (Kagan 1972). But conceptualizing this as the 'wish to know' (Kagan 1972) or the 'need for cognition' (Cacioppo and Petty 1982) seems to suggest that this capacity is aimed at finding out some 'ground truth' (and exclusive to so-called higher animals). Knowledge captured in the predictive models is always subjective and constructed (Heylighen and Joslyn 2001), i.e., the agent has no direct access to the 'real world', but can only 'negotiate' its conditions by actively predicting (constructing) its characteristics. As von Glasersfeld (von Glasersfeld 1995) stated we meet the world only in our failures. In contrast to these existing related approaches, ours centers on prediction errors, rather than *any* uncertainty, and more specifically their dynamics (rather than static uncertainty). Still, our account underscores the role of uncertainty and unpredictability in emotion and motivation (Anselme 2010; Jackson et al. 2014; Whalen 2007). For example, rats seem more motivated to work for a reward in conditioning experiments that introduce some uncertainty in the predictive link between conditioned stimulus and unconditioned stimulus (reward) (Anselme et al. 2013). Conversely, the exacerbating effect of uncontrollability[7] and unpredictability on stress and anxiety is well-documented (Hirsh et al. 2012; Mineka and Hendersen 1985).

Error dynamics are common to all processing, be it interoceptive, exteroceptive, abstract goal-related or low-level sensorimotor. This may better account for the very broad range of situations that can engender positive or negative affects. For example, apart from biologically relevant things, positive emotions may be experienced from scary movies, abstract perceptual stimuli (e.g., in art), acquired tastes such as piquant foods (Rozin and Kennel 1983) or painful stimulation such as masochistic pleasures (Klein 2014). These instances may be difficult to explain from the viewpoint that pleasure is only attached to biologically instrumental situations (or appetitively conditioned stimuli). Below we review those emotions, subtle and intense, that are usually considered to be atypical, for that reason. We try to show that, when taking into account the error dynamics relative to (learned) expected states, they are very representative emotions.

3.3.1 Intrinsic Pleasure and Curiosity

Development is a rich source of emotions. For example, the baby that wants to keep on playing peek-a-boo (Parrott and Gleitman 1989) till predictions of object constancy are fully formed and the situation

---

[7] In the current account the distinction between unpredictability and uncontrollability largely dissolves—actions (to exercise control) are predictions as well, with concomitant expected levels of prediction error decrease or increase.

contains virtually no dynamics in prediction errors anymore. Or the child that is excited to hear the same bed-time story again and again, until errors are driven down by learning its structure (not only of the plot but also of lower level sensory patterns, as is for example clear from toddlers' preference for repetitive rhymes). Later in life, emotion theorists emphasize the centrality of the emotion of interest, for development and beyond (Izard 2007; Silvia 2001). The two factors that have been shown to determine interest can easily be translated to our approach, arguably gaining some specificity in the process. First, only new, unexpected or complex stimuli ("novelty-complexity appraisal") can elicit interest (Silvia 2008), implying that prediction errors are required[8]. The second factor needed to evoke interest is roughly described as comprehensibility or coping potential (Silvia 2008), an appraisal of one's capacity to deal with or understand the (unexpected) stimulus. In our terms, this would be an expectation of a positive rate of error reduction for the current sensorimotor context. We continually, implicitly probe our coping potential by predicting performance (sensory consequences of actions) and computing errors. In fact, making progress (actively reducing errors) in predicting a certain activity domain would be a good indicator of adequate coping potential in this domain in the near future. Hence, the importance of expected rates of progress. In this way, important elements of appraisal theories of emotion can fit within this PP account (Ellsworth and Scherer 2003; Moors 2010).

One of the most influential views on curiosity and exploration is Berlyne's optimal level account (Berlyne 1970). He argued that organisms seek out stimuli with medium level complexity or novelty, to keep their arousal at an optimal, pleasing level. This preference for optimal level of complexity is corroborated in experiments with infants that looked longer at stimulus items that were neither very simple nor very complex (Kidd et al. 2012). Rats too, prefer to spend time in arms of a maze of which the patterns on the walls were slightly more complex relative to the walls they preferred earlier (Dember et al. 1957). The latter studies emphasize the crucial role of experience, which can lower complexity (increase predictability). We would argue that organisms are very much tuned to reducible uncertainty in input. They explore stimuli with medium levels of prediction errors, because they predict a positive rate of error reduction in these inputs. Indeed, they have had experience of error reduction with slightly simpler but similar inputs. In agreement with this, 18 month old children already attend longer to learnable compared to unlearnable linguistic grammar, strongly suggesting they make good estimates of their future predictive learning progress (Gerken et al. 2011).

In adults, through experience, these dynamics, and the pleasures or displeasures derived from them, are not so much situated on the purely perceptual level, but rather on the conceptual level, e.g., stories, jokes or soaps. Although a complete treatment of social emotions will not be given here, observe that they often involve a convergence or divergence in opinions or 'worldviews' (expected states and beliefs). We make models of ourselves and others, like we do for the rest of our environment (Moutoussis et al. 2014), so similar error dynamics are in play in this context. Moreover, the rewarding sense of (em)power(ment) can be interpreted as the result of actively bringing about anticipated sensory effects through action execution (Polani 2009). This seems consistent with our idea that rewards derive their rewarding capacity from reductions in pre-existing prediction errors. Beneficial, motivational effects of a sense of control ('mastery') (Klein and Seligman 1976) may similarly be explained as positive affect from a high expected rate of error reduction.

3.3.2 Humor

In general, the positive emotional mark on unexpected progress towards predicted states is stronger (than just progress). This is consistent with the view proposed here, that a higher than expected rate of error reduction determines positive emotion. This is best illustrated in laughter. In a poignant analysis, Sroufe & Waters (Sroufe and Waters 1976) observe that laughter results when a rapid, maximal tension build-up is followed by a rapid 'release' or 'recovery'. The ill-defined term 'tension' was often used to

---

[8] Note that complexity is also dependent on predictability.

denote some incongruity in perceptual input, assumed to cause some negative arousal. Of course, prediction error can take its place, gaining not only specificity, but also integration in a plausible theory of cognitive processing. More important to stress is that a steep, sudden gradient of prediction error will lead to a prediction of low rate of error reduction. If errors can in fact be reduced, e.g. through an appeal to different predictions (restructuring of input), the reduction rate will be much higher than expected, resulting in intensely positive affect (laughter). This is the typical processing profile, not only for peek-a-boo-like fun in children, but for instances of humor in general (Rozin et al. 2006). Consistent with our approach, both the gradients and the unexpectedness are crucial. In earlier work (Van de Cruys and Wagemans 2011), we developed a similar account for "aesthetic emotions", where artists allow for unexpected increases in error reduction rates for greater appreciation, consistent with the emphasis on *relative* fluency in recent experimental psychoaesthetics (Topolinski and Reber 2010; Wänke and Hansen 2015).

4 Feelings and Function

Our account implies a fundamental misattribution in emotion. The intuition that emotions are entirely caused by objects in the world is misguided, because in fact they are linked to processing characteristics (see also Reber et al. 2004) rather than content of processing itself. The evaluation of error dynamics seems to provide a parallel (affective) dimension to experience, that is not strictly linked to the content (predictions) or particular prediction errors taking part in those dynamics. Still, the specific and diverse forms emotions can take, seem largely dependent on the conceptual context (sensorimotor or cognitive domain) in which the error dynamics appear. But such attributions are always constructions, they will never be directly about what caused the emotions, the (changes in) error reduction rates. One might argue that conscious emotions (or feelings) thereby acquire an intentional object or propositional content, but this does not seem to be a strict requirement, as the existence of conscious moods illustrate. As a related side note, Picard (Picard 2013, p. 2496) reports on two patients experiencing a feeling of intense bliss during epileptic seizures originating in the insula. One patient describes the experience: "...all the ordinary facts about the environment seem suddenly to become infused with certainty and a sense of inevitability... The sense that I had when I was experiencing some of these seizures was not unlike a continuous series of profound "a-ha!" moments. [...] Instead of merely being justified by one or several other considerations or observations, [my beliefs] seemed to be irrefutably supported by literally everything in the world." Such reports suggest first that certainty (lack of prediction errors) is an important factor in bliss, and second that those affective experiences can happen without concrete object or propositional content. Both conclusions are very much in line with the current view. This raises the empirical question of whether the insula, which is known to be both involved in uncertainty or risk processing and in emotions (Singer et al. 2009), might be responsible for some of the computations on error (or uncertainty) dynamics that we propose underlie the reports of intensely positive aha-experiences.

In the current view, the full-fledged (semantically rich), conscious emotions, in all their heterogeneity and object-directedness, appear as a form of construction, a "making sense of" underlying affect (Barrett 2014; Russell 2003). This could already be seen as a form of coping, a reaction to affect: categorizing or labeling an emotion to make it predictable (explaining away errors). The underlying affect consists, in our reasoning, of the changes in error reduction rates. The first-order errors that determine these dynamics can be multisensory (combined interoceptive and exteroceptive). Conscious feeling will amount to finding predictions that best explain the co-occurrence (regularity) of situational context (exteroceptive input) and bodily states, together with changes in rates of error reduction (second-order). Hence, the intentional content of feelings is the product of inferences, but the generation of emotion lies in error dynamics. Differently put, if emotions are categorized, a kind of understanding is attained, which explains away part of the unexpected changes in errors, hence removing some of the emotionality.

To give one simple but concrete example of how a similar emotion might result in different feelings: an unexpected increase of prediction errors may be associated with both fear and shame (given that both are negative emotions). But on the basis of the different conceptual, situational context (e.g., shame probably concerns internalized social expectations, fear not necessarily so) they are differently interpreted and experienced. Interestingly, just by conceptualizing it as shame, one might activate coping strategies that in related situations helped returning to more expected states. The shame prediction for this constellation of inputs is predictive of certain actions or thought strategies that are in turn predictive of a reduction of prediction errors (e.g., actions to restore one's reputation with others).

4.1 Functions of Affect

The computations outlined above should be understood as building a model about how uncertainties evolve in the current context. It seems plausible that these models are crucial in guiding choices (implicitly or explicitly) about whether to continue to engage with the current sensorimotor activities or whether to disengage and switch. Specifically, (unexpected) decreases in prediction errors should raise predictive engagement, in line with how emotion motivates us to remain involved in activities. Confidence[9], as a rather emotional form of metacognition (Chetverikov and Filippova 2014), also seems to stem from these processes. In contrast, (unexpected) increases of prediction errors may change the balance in favor of performing actions to control input (conform to predictions; assimilation), instead of a continued search for revised predictions (accommodation). An action to avoid perceptual input (by averting the eyes), or even a mental switch to leave a certain way of thinking can also be ways to (temporarily) return to a more expected rate of error reduction.

This is where arousal and action tendencies, often considered to be core components of emotion (e.g., Frijda 1987), come in. Rather than being causally constitutive components, we would put them at the output side. If, as we argued, emotions are caused by (unexpected) changes in prediction errors, these computations indeed seem especially important in tipping the balance from updating predictions—a strategy that may be inadequate when confronted with increasing, precise errors—to acting to change the things predicted. Arousal is then derived from such action preparations. Of course, dynamics in autonomic and action-related prediction errors can give rise to emotional valence as well, given that they are governed by the same PP principles. In fact, it seems that the closer to action or autonomic responses these error dynamics are situated, the more intensely negative or positive emotions induced by these dynamics are. This may, however, have more to do with the precision of the predictions than with discrete differences in weight or importance in these predictive systems.

The brain predicts external stimuli in service of the body. It allows anticipation of what the body will need in terms of resources. Hence, it is important to accurately represent bodily states and their causes (Hohwy 2011). However, just recruiting bodily resources, or representing bodily states (and their causes) isn't emotion. If the body perfectly predicts the need for resources based on external input and prior knowledge, there can be bodily activation (arousal) without much emotion. Again, momentary prediction errors do not imply much, it is the changes in (in this case somatovisceral) prediction error, especially the unexpected ones, that should lead to notable emotion. So, while we agree with the models by Seth (Seth 2013), Gu, Hof, Friston & Fan (Gu et al. 2013) or by Barrett & Simmons (Barrett and Simmons 2015) that hold that emotions have to do with somatovisceral prediction errors, we stress that those accounts may not sufficiently explain the *causes* of emotions. The distinction should be clear: those accounts argue emotion is exactly like perception except of somatovisceral instead of exteroceptive inputs. Emotion is then inference to causes that explain (generate)

---

[9] We mean confidence as the common sense personal-level phenomenological construct here, not the subpersonal computational concept.

somatovisceral inputs. We do not deny that these somatovisceral models are constructed, but focus on the dynamics in discrepancies of bodily state as the causes of emotions. The origin of emotion lies not in being able to infer or predict (a cause of) bodily changes, but rather in how we succeed or fail to do so over time (error dynamics).

Apart from actually instigating action (preparation), another function of negative affect (increasing rates of prediction errors) is inducing disengagement from current predictive activity in order to move to a more predictable 'set', what one could call compensatory progress and order. Preliminary evidence for this idea has come from studies finding increased predictable pattern perception when confronted with ambiguity, inconsistency or lack of control (Greenaway et al. 2013; Proulx and Heine 2009; van Harreveld et al. 2014; Whitson and Galinsky 2008). More broadly, uncertainty or inconsistencies may lead people to reaffirm their own (predictable) worldviews, such as nationality, ideology or religion (Inzlicht et al. 2011; Proulx et al. 2010). The negative affect thought to drive these effects is, according to our theorizing, a direct reflection of the higher than expected increase in prediction errors. These examples may then all boil down to efforts to return to an expected, positive rate of uncertainty reduction. The reverse may also hold: positive mood seems to induce a greater reliance on default prior or top-down knowledge, as indicated by an increased influence of prior judgments, scripts or stereotypes in event or person perception (Bless 2000; Bodenhausen et al. 1994). This dovetails with the proposed view that positive mood is linked to high predictive progress, implying that the models the organism has about its world have improved and so are adequate. A rational conclusion for the system would then be to increase reliance on these prior, top-down models (and reduce the influence of prediction errors).

A last function of these affective computations relates to learning and attention. Joffily & Coricelli (Joffily and Coricelli 2013) formally show that the first derivative over time of prediction errors can fulfill a similar function as the one usually assigned to the precision mechanism. An increase in prediction errors (negative valence) may indicate that actual, important changes in the world have taken place, so input (incoming prediction errors) should be weighted more heavily compared to top-down predictions (that apparently need to be updated). In other words, the error rates can be used as a meta-learning signal, tuning the learning rate for new inputs, depending on whether there is much too learn (i.e., in a changing world) or not. What we defined as expectations of error reduction rates then take on the role of expectations on learnability of particular input domains. These are models about what we do not know yet about the structure of the world (Joffily and Coricelli 2013) and how these uncertainties will evolve, i.e., to what extent we estimate these uncertainties will be reducible. Joffily & Coricelli argue that a model that uses rate of change in errors is more parsimonious than one including precisions (a conventional PP model), but more work will be needed to clarify both differences in computational realizability and biological plausibility.

5 Outlook

Throughout history, visual perception —the 'noblest sense'—was considered our main route to find the "ground truth" about the world out there. But physical differences in the world only become information (meaning) by the way we probe them, with our organism-specific predictions. This means that value and information are intertwined by construction —courtesy of our existence as biological organisms. Taking the organism as an (extensible) model of its environment, epistemic coherence is paramount and emotions emerge as the dynamics of attaining this predictive coherence or error reduction.

In large part, the plausibility of this framework for affect hinges on the success of PP proponents in pinning down the physiological basis of the computational scheme. As discussed, prediction errors are sub-personal processing products, which means that we will need to rely on neural measures for tracking these dynamics. So far, there is no direct neural evidence for the existence of the proposed computational operations (or their products). There is only very preliminary and indirect evidence to date for the separable error and prediction populations of neurons (de Gardelle et al. 2012), as postulated by PP. However, the general idea that there are different levels in the hierarchy, with separable prediction errors has recently received support (Diuk et al. 2013; Wacongne et al. 2011). Once we succeed in properly localizing those on different hierarchical levels of processing, we can start looking for dynamics in these errors and neural populations or regions that track these changes and generate expectations of error reduction rates. Most likely, these computations are performed distributed in the brain (similar to first-order PP), given the widely distributed encoding of uncertainty in the brain depending on the domain concerned (Bach and Dolan 2012). In this regard, the overlap in regions found to be important in processing uncertainty and those active for emotional processing, is promising.

All in all, this approach shows some promise for the PP framework to become a common, well-specified language for psychology, from low-level sensorimotor issues to emotional and existential issues. The convergence between computational neuroscience and psychology as seen through the PP account is encouraging. However, this is only a preliminary exploration of how emotions and related aspects of experience may be reframed within PP. Many challenges lie ahead, but the question of whether the brain indeed tracks error increases and decreases and forms predictions about those, is open to empirical and computational investigation. If an emotion theory along the lines presented here is right, we might be getting some formal grasp on affective value and intrinsic motivation, key characteristics of proactive, living organisms.

References

Alter, A. L. & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Pers. Soc. Psychol. Rev., 13* (3), 219–235.

Anselme, P. (2010). The uncertainty processing theory of motivation. *Behav. Brain Res., 208* (2), 291–310.

Anselme, P., Robinson, M. J. F. & Berridge, K. C. (2013). Reward uncertainty enhances incentive salience attribution as sign-tracking. *Behav. Brain Res., 238*, 53–61.

Bach, D. R. & Dolan, R. J. (2012). Knowing how much you don't know: A neural organization of uncertainty estimates. *Nat. Rev. Neurosci., 13* (8), 572–586.

Barrett, L. F. (2014). The conceptual act theory: A précis. *Emot. Rev., 6* (4), 292–297.

Barrett, L. F. & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nat. Rev. Neurosci., 16* (7), 419–429.

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.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Percept. Psychophys., 8* (5), 279–286.

Bless, H. (2000). The interplay of affect and cognition: The mediating role of general knowledge structures. *Feeling and thinking: The role of affect in social cognition* (pp. 201–222). New York, NY: Cambridge University Press.

Bodenhausen, G. V., Kramer, G. P. & Süsser, K. (1994). Happiness and stereotypic thinking in social judgment. *J. Pers. Soc. Psychol., 66* (4), 621–632.

Bromberg-Martin, E. S. & Hikosaka, O. (2011). Lateral habenula neurons signal errors in the prediction of reward information. *Nat. Neurosci., 14* (9), 1209–1216.

Cacioppo, J. T. & Petty, R. E. (1982). The need for cognition. *J. Pers. Soc. Psychol., 42* (1), 116–131.

Cantor, N. & Kihlstrom, J. F. (1987). *Personality and social intelligence.* Englewood Cliffs, N J: Prentice-Hall.

Carver, C. S. & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychol. Rev., 97* (1), 19–35.

Chater, N. & Vlaev, I. (2011). The instability of value. In M. Delgado, E. A. Phelps & T. W. Robbins (Eds.) *Decision making: Attention and performance XXIII* (pp. 81–100).

Chetverikov, A. & Filippova, M. (2014). How to tell a wife from a hat: Affective feedback in perceptual categorization. *Acta Psychol., 151*, 206–213.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci., 36* (03), 181–204.

Csikszentmihalyi, M. (1996). Flow and the psychology of discovery and invention. New York: Harper Collins.

Damasio, A. R. (2000). *The feeling of what happens: Body and emotion in the making of consciousness.* Harvest Books.

——— (2003). *Looking for Spinoza: Joy, sorrow, and the feeling brain.* New York: Houghton Mifflin Harcourt.

Dayan, P. (2012). How to set the switches on this thing. *Curr. Opin. Neurobiol., 22* (6), 1068–1074.

de Gardelle, V., Waszczuk, M. & Egner, T. A. (2012). Concurrent repetition enhancement and suppression responses in extrastriate visual cortex. *Cereb. Cortex.*

de Lafuente, V. & Romo, R. (2011). Dopamine neurons code subjective sensory experience and uncertainty of perceptual decisions. *Proc. Natl. Acad. Sci. U. S. A., 108* (49), 19767–19771.

Deacon, T. W. (2011). *Incomplete nature: How mind emerged from matter.* W. W. Norton & Company.

Dember, W. N., Earl, R. W. & Paradise, N. (1957). Response by rats to differential stimulus complexity. *J. Comp. Physiol. Psychol., 50* (5), 514–518.

den Ouden, H. E. M., Friston, K. J., Daw, N. D. A. & Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cereb.Cortex, 19* (1460-2199 (Electronic)), 1175–1185.

Di Paolo, E. A. (2003). Organismically-inspired robotics: Homeostatic adaptation and teleology beyond the closed sensorimotor loop. *Dynamical Systems Approach to Embodiment and Sociality*, 19–42.

Diuk, C., Tsai, K., Wallis, J., Botvinick, M. & Niv, Y. (2013). Hierarchical learning induces two simultaneous, but separable, prediction errors in human basal ganglia. *J. Neurosci., 33* (13), 5797–5805.

Dreisbach, G. & Fischer, R. (2012). Conflicts as aversive signals. *Brain Cogn., 78* (2), 94–98.

Duncan, S. & Barrett, L. F. (2007). The role of the amygdala in visual awareness. *Trends in Cognitive Science, 11* (5), 190–192.

Duncker, K. (1941). On pleasure, emotion, and striving. *Philosophy and Phenomenological Research, 1* (4), 391–430.

Egbert, M. & Barandiaran, X. E. (2014). Modeling habits as self-sustaining patterns of sensorimotor behavior. *Front. Hum. Neurosci., 8.*

Egbert, M. & Canamero, L. (2014). In H. Sayama, J. Rieffel, S. Risi, R. Doursat & H. Lipson (Eds.) *Habit-based regulation of essential variables*. Cambridge, MA: MIT Press.

Egbert, M., Virgo, N., Egbert, M. D., Froese, T., Kampis, G., Karsai, I. & Szathmáry, E. (2013). For biological systems, maintaining essential variables within viability limits is not passive. *Constructivist Foundations,* 9 (1), 109–111.

Ellsworth, P. C. & Scherer, K. R. (2003). Appraisal processes in emotion. *Handbook of Affective Sciences, 572.*

Festinger, L. (1962). Cognitive dissonance. *Sci. Am., 207* (4), 93–107.

Freddolino, P. L. & Tavazoie, S. (2012). Beyond homeostasis: A predictive-dynamic framework for understanding cellular behavior. *Annu. Rev. Cell Dev. Biol., 28*, 363–384.

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cogn. Emot., 1*, 115–143.

——— (2006). *The laws of emotion.* Mahwah, N.J: Psychology Press.

Friston, K. J. (2003). Learning and inference in the brain. *Neural Netw., 16* (9), 1325–1352.

——— (2009). The free-energy principle: A rough guide to the brain? *Trends Cogn. Sci., 13* (7), 293–301.

——— (2010). The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci., 11* (2), 127–138.

Friston, K. J., Daunizeau, J. & Kiebel, S. J. (2009). Reinforcement learning or active inference? *PLoS One, 4* (7), e6421.

Friston, K. J., Daunizeau, J., Kilner, J. & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biol. Cybern., 102* (3), 227–260.

Friston, K. J., Thornton, C. & Clark, A. (2012a). Free-energy minimization and the Dark-Room problem. *Front. Psychol., 3.*

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M. A., Brown, H., Dolan, R. J., Moran, R. A. & Bestmann, S. (2012b). Dopamine, affordance and active inference. *PLoS Comput. Biol., 8* (1), e1002327.

Friston, K. J., Schwartenbeck, P., Fitzgerald, T. M., Behrens, T. & Dolan, R. J. (2013). The anatomy of choice: Active inference and agency. *Front. Hum. Neurosci., 7*, 598.

Friston, K. J., Rigoli, F., Ognibene, D. M., Fitzgerald, T. & Pezzulo, G. (2015). Active inference and epistemic value. *Cogn. Neurosci., 6* (4), 187–214.

Froese, T. & Ikegami, T. (2013). The brain is not an isolated ‘‘black box,” nor is its goal to become one. *Behav. Brain Sci., 36* (03), 213–214.

Gerken, L., Balcomb, F. K. & Minton, J. L. (2011). Infants avoid ‘labouring in vain’ by attending more to learnable than unlearnable linguistic patterns. *Dev. Sci., 14* (5), 972–979.

Gershman, S. J. & Daw, N. D. (2012). Perception, action and utility: The tangled skein. *Principles of Brain Dynamics: Global State Interactions*, 293–312.

Greenaway, K. H., Louis, W. R. & Hornsey, M. J. (2013). Loss of control increases belief in precognition and belief in precognition increases control. *PLoS One, 8* (8).

Gu, X., Hof, P. R., Friston, K. J. & Fan, J. (2013). Anterior insular cortex and emotional awareness. *J. Comp. Neurol., 521* (15), 3371–3388.

Harper, M. (2009). The replicator equation as an inference dynamic.

Hebb, D. O. (1946). On the nature of fear. *Psychol. Rev., 53* (5), 259–276.

Heylighen, F. & Joslyn, C. (2001). Cybernetics and second order cybernetics. *Encyclopedia of Physical Science & Technology, 4*, 155–170.

Hirsh, J. B., Mar, R. A. & Peterson, J. B. (2012). Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychol. Rev., 119* (2), 304–320.

Hoffmann, J. (2003). Anticipatory behavioral control. In M. V. Butz, O. Sigaud & P. Gérard (Eds.) *Anticipatory behavior in adaptive learning systems* (pp. 44–65). Springer Berlin Heidelberg.

Hohwy, J. (2011). Phenomenal variability and introspective reliability. *Mind Lang., 26* (3), 261–286.

Horn, B. K. P. (1980). *Derivation of invariant scene characteristics from images* (pp. 371–376). New York, NY, USA: ACM.

Hsee, C. K. & Ruan, B. (2016). The pandora effect: The power and peril of curiosity. *Psychol. Sci., 27* (5), 659–666.

Iglesias, S., Mathys, C., Brodersen, K. H. A., Piccirelli, M. & den Ouden, H. E. M. A. (2013). Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron, 80* (2), 519–530.

Inzlicht, M., Tullett, A. M. & Good, M. (2011). The need to believe: A neuroscience account of religion as a motivated process. *Religion Brain Behav., 1* (3), 192–212.

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspect. Psychol. Sci., 2* (3), 260–280.

Jackson, F., Nelson, B. D. & Proudfit, G. H. (2014). In an uncertain world, errors are more aversive: Evidence from the error-related negativity. *Emotion*.

James, W. (1890). The principles of psychology. *Harvard UP, Cambridge, MA*.

Joffily, M. & Coricelli, G. (2013). Emotional valence and the Free-energy principle. *PLoS Comput. Biol., 9* (6), e1003094.

Kagan, J. (1972). Motives and development. *J. Pers. Soc. Psychol., 22* (1), 51–66.

Kaplan, H. (1987). Lottery winners: The myth and reality. *J. Gambl. Stud., 3* (3), 168–178.

Kaplan, F. & Oudeyer, P.-Y. (2007). In search of the neural circuits of intrinsic motivation. *Front. Neurosci., 1* (1), 225–236.

Keramati, M. & Gutkin, B. S. (2011). A reinforcement learning theory for homeostatic regulation. *Advances in neural information processing systems* (pp. 82–90).

Kidd, C., Piantadosi, S. T. & Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One, 7* (5), e36399.

Klein, C. (2014). The penumbral theory of masochistic pleasure. *Rev.Phil.Psych., 5* (1), 41–55.

Klein, D. C. & Seligman, M. E. (1976). Reversal of performance deficits and perceptual deficits in learned helplessness and depression. *J. Abnorm. Psychol., 85* (1), 11–26.

Knight, E. J., Klepac, K. M. & Kralik, J. D. (2013). Too good to be true: Rhesus monkeys react negatively to better-than-expected offers. *PLoS One, 8* (10), e75768.

Lauwereyns, J. (2010). *The anatomy of bias: How neural circuits weigh the options.* Cambridge, MA: MIT Press.

Little, D. Y.-J. & Sommer, F. T. (2013). Maximal mutual information, not minimal entropy, for escaping the "dark room". *Behav. Brain Sci., 36* (03), 220–221.

Mandler, G. (2003). Emotion. In D. K. Freedheim & I. B. Weiner (Eds.) *History of psychology*. John Wiley and Sons.

Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J. & Stephan, K. E. (2014). Uncertainty in perception and the hierarchical Gaussian filter. *Front. Hum. Neurosci., 8*, 825.

Millikan, R. G. (2004). *Varieties of meaning: The 2002 Jean Nicod lectures.* Cambridge, MA: Mit Press.

Mineka, S. & Hendersen, R. W. (1985). Controllability and predictability in acquired motivation. *Annu. Rev. Psychol., 36* (1), 495–529.

Moors, A. (2010). Automatic constructive appraisal as a candidate cause of emotion. *Emot. Rev., 2* (2), 139–156.

Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R. J. & Friston, K. J. (2014). Bayesian inferences about the self (and others): A review. *Conscious. Cogn., 25*, 67–76.

Moutoussis, M., Story, G. W. & Dolan, R. J. (2015). The computational psychiatry of reward: Broken brains or misguided minds? *Front. Psychol., 6*, 1445.

Niv, Y. & Chan, S. (2011). On the value of information and other rewards. *Nat. Neurosci., 14* (9), 1095–1097.

Noordewier, M. K. & Breugelmans, S. M. (2013). On the valence of surprise. *Cogn. Emot., 27* (7), 1326–1334.

Oudeyer, P.-Y., Kaplan, F. & Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Trans. Evol. Comput., 11* (2), 265–286.

Parrott, W. G. & Gleitman, H. (1989). Infants' expectations in play: The joy of peek-a-boo. *Cogn. Emot., 3* (4), 291–311.

Pezzulo, G., Rigoli, F. & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Prog. Neurobiol., 134*, 17–35.

Picard, F. (2013). State of belief, subjective certainty and bliss as a product of cortical dysfunction. *Cortex, 49* (9), 2494–2500.

Polani, D. (2009). Information: Currency of life? *HFSP J., 3* (5), 307–316.

Proulx, T. & Heine, S. J. (2009). Connections from Kafka exposure to meaning threats improves implicit learning of an artificial grammar. *Psychol. Sci., 20* (9), 1125–1131.

Proulx, T., Heine, S. J. & Vohs, K. D. (2010). When is the unfamiliar the uncanny? Meaning affirmation after exposure to absurdist literature, humor, and art. *Pers. Soc. Psychol. Bull., 36* (6), 817–829.

Purves, D., Wojtach, W. T. & Lotto, R. B. (2011). Understanding vision in wholly empirical terms. *Proceedings of the National Academy of Sciences, 108* (Supplement_3), 15588–15595.

Reber, R., Schwarz, N. & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Pers. Soc. Psychol. Rev., 8* (4), 364–382.

Reisenzein, R. (2009). Emotional experience in the computational belief-desire theory of emotion. *Emot. Rev., 1* (3), 214–222.

Rozin, P. & Kennel, K. (1983). Acquired preferences for piquant foods by chimpanzees. *Appetite, 4* (2), 69–77.

Rozin, P., Rozin, A., Appel, B. & Wachtel, C. (2006). Documenting and explaining the common AAB pattern in music and humor: Establishing and breaking expectations. *Emotion, 6* (3), 349–355.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev., 110* (1), 145–172.

Schiffer, A.-M., Ahlheim, C. & Wurm, M. F. S. (2012). Surprised at all the entropy: Hippocampal, caudate and midbrain contributions to learning from prediction errors. *PLoS One, 7* (5), e36445.

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Trans. Auton. Ment. Dev., 2* (3), 230–247.

Schouppe, N., Braem, S., Houwer, J. D. A., Verguts, T. & Ridderinkhof, K. R. N. (2014). No pain, no gain: The affective valence of congruency conditions changes following a successful response. *Cogn. Affect. Behav. Neurosci.*, 1–11.

Schrödinger, E. (1992). *What is life? With Mind and Matter and Autobiographical Sketches.* Cambridge University Press.

Schultz, W. (2007). Reward. *Scholarpedia J., 2* (3), 1652.

Schwartenbeck, P., FitzGerald, T. H. B., Mathys, C., Dolan, R. & Friston, K. (2014). The dopaminergic midbrain encodes the expected certainty about desired outcomes. *Cereb. Cortex*, bhu159.

Serences, J. T. (2008). Value-based modulations in human visual cortex. *Neuron, 60* (6), 1169–1181.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci., 17* (11), 565–573.

Silvia, P. J. (2001). Interest and interests: The psychology of constructive capriciousness. *Rev. Gen. Psychol., 5* (3), 270–290.

——— (2008). Interest—The curious emotion. *Curr. Dir. Psychol. Sci., 17* (1), 57–60.

Singer, T., Critchley, H. D. & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends Cogn. Sci., 13* (8), 334–340.

Singh, S., Lewis, R. L. & Barto, A. G. (2009). *Where do rewards come from* (pp. 2601–2606).

Solway, A. & Botvinick, M. M. (2012). Goal-directed decision making as probabilistic inference: A computational framework and potential neural correlates. *Psychol. Rev., 119* (1), 120–154.

Srivastava, N. & Schrater, P. (2015). Learning what to want: context-sensitive preference learning. *PLoS One, 10* (10), e0141129.

Sroufe, L. A. & Waters, E. (1976). The ontogenesis of smiling and laughter: A perspective on the organization of development in infancy. *Psychol. Rev., 83* (3), 173–189.

Stansbury, D. E., Naselaris, T. & Gallant, J. L. (2013). Natural scene statistics account for the representation of scene categories in human visual cortex. *Neuron, 79* (5), 1025–1034.

Van de Cruys, S. & Wagemans, J. (2011). Putting reward in art: A tentative prediction error account of visual art. *Iperception, 2* (9), 1035–1062.

van Harreveld, F., Rutjens, B. T., Schneider, I., Nohlen, H. U. & Keskinis, K. (2014). In doubt and disorderly: Ambivalence promotes compensatory perceptions of order. *J. Exp. Psychol. Gen., 143* (4), 1666–1676.

von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning. studies in mathematics education series: 6.* Bristol, PA: Falmer Press, Taylor & Francis Inc..

Wacongne, C., Labyt, E., Wassenhove, V., Bekinschtein, T., Naccache, L. & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci. U. S. A., 108* (51), 20754–20759.

Weber, A. & Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenol. Cognitive Sci., 1* (2), 97–125.

Whalen, P. J. (2007). The uncertainty of it all. *Trends Cogn. Sci., 11* (12), 499–500.

Whitson, J. A. & Galinsky, A. D. (2008). Lacking control increases illusory pattern perception. *Science, 322* (5898), 115–117.

Wood, W. & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychol. Rev., 114* (4), 843–863.

Wänke, M. & Hansen, J. (2015). Relative processing fluency. *Curr. Dir. Psychol. Sci., 24* (3), 195–199.

Wörgötter, F. & Porr, B. (2005). Temporal sequence learning, prediction, and control: A review of different models and their relation to biological mechanisms. *Neural Comput., 17* (2), 245–319.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *Am. Psychol., 35* (2), 151.