COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 44 (2020) e12867

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12867

Unification by Fiat: Arrested Development of Predictive Processing

Piotr Litwin,[a,b] Marcin Miłkowski[b]

[a]*Faculty of Psychology, University of Warsaw*
[b]*Institute of Philosophy and Sociology, Polish Academy of Sciences*

Received 26 July 2019; received in revised form 25 April 2020; accepted 15 May 2020

Abstract

Predictive processing (PP) has been repeatedly presented as a unificatory account of perception, action, and cognition. In this paper, we argue that this is premature: As a unifying theory, PP fails to deliver general, simple, homogeneous, and systematic explanations. By examining its current trajectory of development, we conclude that PP remains only loosely connected both to its computational framework and to its hypothetical biological underpinnings, which makes its fundamentals unclear. Instead of offering explanations that refer to the same set of principles, we observe systematic equivocations in PP-based models, or outright contradictions with its avowed principles. To make matters worse, PP-based models are seldom empirically validated, and they are frequently offered as mere just-so stories. The large number of PP-based models is thus not evidence of theoretical progress in unifying perception, action, and cognition. On the contrary, we maintain that the gap between theory and its biological and computational bases contributes to the arrested development of PP as a unificatory theory. Thus, we urge the defenders of PP to focus on its critical problems instead of offering mere re-descriptions of known phenomena, and to validate their models against possible alternative explanations that stem from different theoretical assumptions. Otherwise, PP will ultimately fail as a unified theory of cognition.

Keywords: Predictive processing theory of cognition; Unification; Explanation; Bayesian just-so stories; Consistency fallacy

Correspondence should be sent to Piotr Litwin, Faculty of Psychology, University of Warsaw, Stawki 5/7, 00-183, Warsaw, Poland. E-mail: piotr.litwin@psych.uw.edu.pl.

1. Introduction

To adaptively guide behavior in a rapidly changing environment, nervous systems must unravel perplexingly complex relations between inherently ambiguous sensory inputs and their underlying causes—hidden states of internal and external milieux. Predictive processing (PP) theory has been gaining momentum, becoming a dominant candidate for a unified explanatory strategy of how the brain manages this task. Since it has been discussed at length multiple times (Clark, 2013, 2016; Hohwy, 2013), we will keep our introduction brief.

The central tenet of PP is that nervous systems predict their sensory inputs, following the imperative to minimize the mismatch between these predictions and the actual sensory signals. Through this recursive process, the statistical structure of intra- and intermodal sensory dependencies on various timescales is extracted to form an internal model of the causal matrix underlying the system's inputs that further informs consecutive predictions (Williams, 2018c).

Certain constraints must be imposed on the functioning and structure of the nervous system for the internal model to operate efficiently. To instantiate predictions, the brain uses approximate Bayesian causal inference, in which the mean value of the posterior is determined by prior probability of the hypothesis (as determined by the causal structure described in the model) and likelihood (probability of obtaining such evidence for a true hypothesis). However, rather than focusing on a moment-to-moment mirroring of the external causes of short-term activations in their sensorium, the brain aims at long-term prediction error minimization through development of a comprehensive model, allowing predictive control of the environment. This is held to be achieved through acquisition of a multi-tiered, hierarchical organization with multiple levels conceived of as ensembles of latent variables, simulating the underlying causal structure at varying levels of abstraction and spatiotemporal scaling (Kiefer & Hohwy, 2018; Williams, 2018b). At higher levels, general, relatively invariable (at the ontogenetic time scale), and context-independent assumptions about the world are conveyed. These assumptions constrain the gamut of possible context-dependent hypotheses (posterior probability distributions) at the lower levels of the model. In the process, inevitable mismatches (prediction errors) between these predictions and actual sensory signals occur. Prediction errors may either be explained away through optimization of lower-level model parameters or, if persistent, propagated up the hierarchy to update higher-order hypotheses or revise the generative model. The extent to which prediction errors influence posterior probability distributions depends on the ratio between inverse variances of prior and likelihood density distributions (precisions).

Originally inspired by the work on bidirectional, hierarchical processing in the visual cortex and Bayesian modeling in cognitive science, PP was developed within the realm of perception science (Clark, 2013). It quickly expanded to other domains of cognitive science and psychology to inspire theories of action (Adams, Shipp, & Friston, 2013; Friston, Daunizeau, Kilner, & Kiebel, 2010), attention (Hohwy, 2012), language and thought (Clark, 2016; Lupyan & Clark, 2015), interoception and emotion (Barrett &

Simmons, 2015; Seth, Suzuki, & Critchley, 2012), psychiatric disorders (Friston, Stephan, Montague, & Dolan, 2014), body perception (Apps & Tsakiris, 2014), pain perception (Bissell, Ziadni, & Sturgeon, 2018), and many others. On a tide of this great explanatory enthusiasm, PP was repeatedly anointed a grand unifying theory: "the first truly unifying account of perception, cognition, and action" (Clark, 2016, p. 2), bringing "perception, action, and attention into a single unifying framework" (Clark, 2013, p. 201), explaining "perception and action and everything mental in between" (Hohwy, 2013, p. 1), and bridging traditionally opposed theories of cognition and consciousness (e.g., Global Workspace Theory and sensorimotor accounts; Jęczmińska, 2017) and research traditions (e.g., representationalism and enactivism; Clark, 2015).

We argue that PP currently fails to stand as such a unifying theory, and that its failure is deeply rooted in its current theoretical structure. The interpretation of its mathematical underpinnings turns out to be ambiguous, and the PP hierarchy seems implausible as a general blueprint of a cognitive architecture. A formalization does not make a theory strict. As noted by Cooper and Shallice (1995), Clark Hull's theory of behavior and the cognitive architecture Soar are specified in a mathematically strict way but interpreted in a number of inconsistent ways. We believe the same point applies to PP. In this case, the under-determination of fundamentals results in a "horizontal" trajectory of PP development: It expands "to the side," being extrapolated in the form of many re-descriptions to particular psychological and cognitive phenomena prematurely. Instead of developing "vertically," or simply going deeper into fundamentals of the theory to increase its theoretical virtues, a plethora of proto-models and theories—frequently mutually exclusive and inconsistent with basic PP tenets—is being formulated in liberally interpreted PP terms. This explanatory rush creates conceptual confusion and effectively inhibits the theory's advance, as fundamental concepts of PP are either further blurred or, at best, remain ill-defined. We hope our criticism will contribute to removing these obstacles in the future work on PP.

The paper is structured as follows: In the next section, we briefly explicate the notion of unification, later used to analyze the tensions in the theoretical commitments of PP. Then, we focus on the major problem of PP: its lack of clear fundamentals. It is difficult to understand how some of its computational structures correspond to cognitive constructs, which creates a gap between the theory and its algorithmic underpinnings. Then, we turn to the issue of the current developmental trajectory of the theory. Rather than validating PP models empirically, researchers tend to either appeal to common rhetorical strategies or point to empirical evidence that is merely consistent with their models, which is insufficient to establish the grand unificatory nature of PP. In the penultimate section, we deal with possible objections to our critique and, finally, conclude by pointing out that the theory cannot reach unity without a substantial change in current practices.

2. Unification in cognitive science

As Colombo and Wright (2017, p. 5) observe, advocates of PP have left unspecified the conditions that make their assertion that PP is a grand unifying theory true.

Intuitively, a single theory capable of explaining all relevant cognitive phenomena by appealing to the same principle of Bayesian inference as approximated by PP in perception and action certainly seems to be unificatory. In the unificationist account of explanation, a genuinely unificatory theory should provide explanations of the widest possible range of phenomena using as few explanatory patterns as possible (Kitcher, 1989). Instead of simply collecting multiple explanations based on different approaches and reducing them to a single theory, PP aspires to offer a unified mathematical theory, composed of a limited number of explanatory patterns. In short, PP is supposed to be a single, homogeneous theory whose scope covers all phenomena of interest.

Here, we adopt a slightly more nuanced account of unification (Miłkowski, 2016; Miłkowski & Nowakowski, 2019) than that of Kitcher (1989). In short, there are four dimensions of unification in our view: generality, simplicity, homogeneity (or non-monstrosity), and systematicity (for more details, see Miłkowski & Nowakowski, 2019). While Kitcher's account of explanation follows the received view of explanation in terms of arguments, we do not assume as much. The argument view is not directly applicable to cognitive theories which may take the form of computer programs, flow charts, or verbal descriptions. Additionally, we believe that the account of unification in cognitive science should view unified theories of cognition, defended forcefully by Allen Newell (1990), as indeed unified. Newell's view was that one could provide a unified theory by proposing an outline of cognitive architecture. However, these architectures are not stated in terms of laws or invariant regularities but as functional structures. Because PP offers an outline of a functional structure (e.g., Kanai, Komura, Shipp, & Friston, 2015), in contrast to some other Bayesian approaches (Chater & Oaksford, 1999), such an account of unification should also be applicable to PP.

The account we adopt is similar in some respects to that defended by Kitcher: We posit that genuine unification requires the widest possible scope of a single, simple, and homogeneous theory. Generality and simplicity are the features stressed by Kitcher: A unified theory should be as general as possible to cover all relevant invariant structure in target phenomena; at the same time, it should remain as simple as possible. But, in addition, it is crucial that a unified theory remains a single, homogeneous theory. By homogeneity, we mean that a theory should be more than a conjunction of disconnected explanatory patterns, termed *non-monstrosity* by Votsis (2015). A scientific representation is monstrous when it contains parts which cannot be disconfirmed by a single fact, which implies that it contains something akin to isolated islands—collections of statements assumed to be true at the same time, without any deeper connections. Moreover, a scientific theory should be also systematic with respect to its domain of application (Hoyningen-Huene, 2013). There should be a systematic body of explanatory principles governing a class (or multiple classes) of phenomena; in cognitive science these could relate, for example, to various levels of explanation or stages of processing.

We believe this is the understanding that underlies most claims about the unificatory nature of PP; it is supposed to be a relatively simple theory with a large scope of application, systematically built to homogeneously cover the whole domain. The same set of assumptions also seems to be at play in Newell's account of unified theories of cognition.

The current account of unification is applicable to various kinds of scientific representation: models, families of models, theories, frameworks, and toolboxes. While these terms are used somewhat liberally in the cognitive science community, by *theories* we mean more general scientific representations that allow one to build models of particular phenomena (such as tasks), and by *frameworks*, we mean, following Newell, "conceptual structures (often computer systems) that are reputed to be relatively content free but that permit specific cognitive theories to be inserted in them in some fashion" (1990, pp. 16–17).[1] Lastly, *toolboxes* are (mostly computational) methods used in various theoretical approaches, although tool development may have crucial theoretical implications (Gigerenzer, 1991). As we will argue below, PP is usually assumed to be a unifying theory, but remains a computational framework at best.

3. Unclear fundamentals

Below, we aim to provide the examples of a fallacious explanatory pattern recurring in PP-based explanations: repeatable identifications of technical PP terms with phenomena in the cognitive domain. This renders the present unificatory power of PP questionable as meanings of its technical terms are untenably broad and oscillate between particular subtheories (Section 3.1). Moreover, these core concepts are introduced into PP-based explanatory schemes in a way that ignores major assumptions of the theory (Section 3.2). We argue that the inconsistent or incorrect usage only creates an impression of the existence of the limited set of core explanatory patterns constitutive for PP, and this impression is wrong: Resulting incompatibilities between models of individual phenomena and avowed tenets of PP are symptoms of heterogeneity and possible arrested development, not hallmarks of grand unification.

3.1. Explanation by equivocation: Fluid core concepts

Consider as an example the notion of precision. In PP, precision is inversely related to the variability or uncertainty of signals or priors (width of likelihood and prior probability distributions) and proportional to the influence that prediction errors or priors exert on the generative model (Kanai et al., 2015). Computationally, it is identified with the confidence (Allen et al., 2016, p. 9; Friston et al., 2012, p. 238; Kanai et al., 2015, p. 3) or salience (Barrett & Simmons, 2015, p. 2; Kanai et al., 2015, p. 3; Friston et al. 2012, p. 1; Friston et al., 2014, p. 149) of (sensory) signals. In this specific computational sense, such identifications are uncontroversial: Information conveyed by less variant signals is more certain and reliable. Therefore, it should be more "salient" to a cognitive system—it should influence further processing, for example, update the system's expectations, to a greater degree. However, PP proponents tend to switch to the domain of subjective experience, as if, due to this semantic identification on a computational level, precision were to be identified with subjective feelings of confidence or salience. This is unwarranted.

Let us consider following examples. Stephan et al. (2016, p. 88; our emphasis) propose that "neuronal encoding of confidence in the beliefs (. . .) corresponds to the salience or precision afforded to sensory evidence" and it does so "psychologically." Engström et al. (2018) identify high precision with the subjective feelings of high confidence (p. 165) or trust (p. 179). According to the theory linking aberrant functioning of the oxytocin system to autism spectrum disorders (Quattrocki & Friston, 2014), the lack of proper interoceptive signal precision attenuation "impair[s] the child's ability to prescribe precision (attention) to appropriate social cues" (p. 421) and, thus, engenders a failure to "assign salience to a mother's face" (p. 420). Finally, in a PP-based account of the dopaminergic system (Friston et al., 2012), salience is downright synonymous with precision, being defined as "an attribute of (probabilistic) representations that determines the confidence or certainty about what is represented" (p. 2). According to the authors, "by associating salience with precision we can also connect to constructs like *incentive salience* in psychology and *aberrant salience* in psychopathology" (Friston et al., 2012, p. 2). This tendency to define psychological phenomena as synonymous with precision (or other functional PP terms[2]) is widespread.

Such identifications are problematic because they could lead to a fallacy of equivocation—the confusion of several meanings of a single term in an argument (cf. Copi & Cohen, 1990; Section 3.3). For example, psychological salience and computational precision are semantically inequivalent, as incentive "magnetism" (or social salience, see Quattrocki & Friston, 2014) has no clear-cut relation to the quality of perceptual representation (Colombo & Wright, 2017) or sharpness of sensory attention. Without further explication of how they refer to each other, the putative connection to functional psychiatric definitions is merely declarative—especially since some PP proponents seemingly preserve the original meaning of the notion of salience in their works, for example, while writing about the "salience network" in the brain (e.g., Barrett & Simmons, 2015). A unified theory of anything cannot be created simply via a series of equivocations, which are then used to argue that the same theoretical construct underlies various phenomena.

These equivocations become even more perplexing when we consider that precision weighting mechanisms should actually correspond (psychologically) to sensory attenuation or attention (Friston, 2018, p. 1019). This has led some researchers to claim that "although the concepts of salience, confidence and attention may appear distinct, their intimate relationship can be interpreted as an integral part of perceptual inference—reflecting the different faces of precision" (Kanai et al., 2015, p. 9). However, while mechanisms of how attention may result from neuromodulatory gains in hierarchical predictive systems have been sketched (Feldman & Friston, 2010; Hohwy, 2012, 2017), there is a lack of such mechanisms for psychological salience and confidence. This leads to inevitable confusion: For example, Quattrocki and Friston (2014) suggest that the attenuated precision of exteroceptive cues results in an inability to pay attention to social objects (p. 421) and diminished salience of these objects (p. 420). The spatiotemporal order of this interplay between attention and salience is unclear; in the authors' words, sensory cues "take on a salience that will capture attention" (p. 414) but also "attention can enhance visual saliency" (p. 420).

Thus far in this section, we have aimed to show that there is a one-to-many mapping between the key technical PP term of precision (having strict computational sense) and the whole gamut of subjective phenomena. For instance, subjective feelings of trust, salience, confidence, or sharpened attention are all being attributed to "enhanced precision"; however, such attempts to cover various psychological phenomena by identifying them with a single computational term cannot result in informative explanations. As long as the process through which precision regulation gives rise to particular subjective phenomena is not specified exactly (e.g., enhanced precision of which signals? at which levels of the predictive architecture?), it will remain unclear what differentiates them from one another.

However, our concerns are not restricted to precision. The lure of equivocation may even be stronger for technical PP terms that have homonymic counterparts in the subjective domain, such as *predictions*, *expectations*, or *beliefs*. Humans are capable of predicting, expecting, or believing, and PP-based explanations seem to be ready made for these phenomena. Indeed, it has been argued that persisting negative expectations (despite positive feedback) in patients with major depressive disorders result from diminished precision of Bayesian expectations of positive events (Kube, Schwarting, Rozenkrantz, Glombiewski, & Rief, 2019), that sustained religious beliefs stem from a decreased ability to update higher-level Bayesian beliefs (van Elk & Aleman, 2017; see Section 4.1), and that both prospection and predictive dynamics of the brain rely on uncertainty mitigation (Gilead, Trope, & Liberman, 2019). However, despite the appealing simplicity of such explanations, there are important differences between subjective and computational (as envisioned by PP) domains (Litwin & Miłkowski, in press). Most importantly, according to PP, computational processes that implement prediction error minimization underlie *all cognition*. Should beliefs, certainty, confidence, expectations, or ability to prospect be explained by PP, models of these phenomena, distinguishing them from other aspects of our mental lives, must be proposed. One cannot rely on intuitive arguments based on mere terminological affinity, especially given that the ways in which human and Bayesian beliefs evolve often explicitly diverge (Kahneman & Tversky, 1972).

One could object that the meanings of the technical terms of a theory naturally fluctuate, that new experimental findings and theoretical considerations continuously morph their senses, extending the explanatory scope of the theory and specifying distinctions between explanations of particular phenomena. We do not deny this. Indeed, if defenders of PP had not claimed that the unique value of their theory is that it unifies a huge range of neurocognitive phenomena, the ambiguity or vagueness of the core concepts would be simply a by-product of the constant development of the theory, to be removed at later stages of inquiry. But as long as formal objects of the theory pertain to a range of widely differing phenomena and remain open to diverse interpretations, one cannot speak about unification. Instead, we see a conceptual gap between PP qua theory and its various implementations in models or verbal descriptions of potential models.

3.2. Models incompatible with the grand theory

Let us go back to precision, a central component of many PP-based explanations. Many recent models go far beyond simple identifications, focusing on the way dynamically changing ratios between prior and prediction error precisions give rise to particular cognitive and psychiatric phenomena. To efficiently orchestrate the selection of relevant information, precision estimation should be dependent on (among other factors) current goals, task demands, or higher-level knowledge and contextualization (Miller & Clark, 2018), and should also be relatively independent at different levels of the inferential hierarchy (Hohwy, 2017). PP proponents frequently refer to these contextualizing factors. This approach carries the risk of precision taking the form of a "magic modulator," a free variable that can be adjusted to fit every explanation (Miller & Clark, 2018, p. 2568). In this section, we seek to show that recourse to disturbed global dynamics of precision weighting indeed leads to explanations which are at odds with the core tenets of the grand paradigm.

We can point to contemporary computational models of psychosis and schizophrenia as an example (Corlett et al., 2019; Sterzer et al., 2018; Sterzer, Voss, Schlagenhauf, & Heinz, 2019). According to these models, the heterogeneous clinical picture of schizophrenia (e.g., simultaneous presence of delusions, hallucinations, and altered lower-level perceptual processing) results from disturbed global precision regulation: imbalances between prior and prediction error precisions, with skewness of the ratio varying throughout the levels of the inferential hierarchy. However, as we show below, these models are inconsistent with fundamental commitments of PP.

Sterzer et al. (2018) provide an analysis of an interplay between higher and lower levels of the hierarchy leading to the clinical manifestation of symptoms. Patients with schizophrenia are less susceptible to perceptual oddball effects and certain visual illusions (Friston, Brown, Siemerkus, & Stephan, 2016), such as the hollow mask illusion, in which healthy people perceive the concave side of the mask as convex (Dima et al., 2009). Thus, the percepts of patients in this population, even though more veridical, are abnormal, and reflect a diminished precision of low-level perceptual priors. Weakened priors also surface in studies using intermittent presentation of ambiguous stimuli that can be perceived as rotating either leftward or rightward. While in healthy subjects this leads to a late-trial dominance of one of the percepts (due to the steady build-up of prior beliefs informed by previous experiences), percepts do not stabilize (priors do not accumulate) in schizophrenic patients (Schmack, Schnack, Priller, & Sterzer, 2015). But how could weak perceptual priors give rise to hallucinatory percepts? They *should* emerge as a result of hyperprecise priors dominating inferences regardless of actual sensory evidence (Corlett et al., 2019). Indeed, priors seem to exert greater influence on perceptual inference in hallucination-prone individuals, regardless of whether or not they have received a clinical diagnosis (Powers, Mathys, & Corlett, 2017).

PP proponents have reconciled strong and weak prior accounts to form a single model of schizophrenia and psychosis arising from disturbed global precision regulation dynamics. In this take, inefficient structural priors, derived from natural scene statistics, are

compensated by precise higher-order priors semantically consistent with delusional beliefs (Sterzer et al., 2018, 2019). The latter simultaneously facilitate sensory activations to give rise to hallucinatory percepts semantically consistent with (delusional) expectations (Corlett et al., 2019). Note that this line of reasoning rests on the assumption that the structural priors driving visual illusions are hard-wired into lower-level perceptual systems and are relatively independent of higher-level influences (Teufel et al., 2015, p. 13405). However, this story is in stark contrast with major PP commitments on the structure and functioning of the nervous system.

Predictive processing entails certain indispensable assumptions regarding functional organization of the inferential structure. The generative model adopts a multi-level, hierarchical, sphere-like (or tree-like) form, with the parts receiving input occupying the outer edges (Williams, 2018b). The various levels are conceived of as latent variables, with higher levels capturing increasingly general regularities. The hierarchy is homogeneous and organized along the single continuum, although what exactly organizes the hierarchy (e.g., abstractness, spatiotemporal scaling, complexity, detachment from sensory processes) remains problematic (Williams, 2018a).[3] Information passes only between the adjacent levels, as each feeds predictions only to the level immediately below (providing constraints on the plausible parameter values/posterior distributions) and prediction errors to the level immediately above. Any interaction between non-adjacent levels is indirect—for example, upper levels may constrain plausible predictions at lower levels, which in turn determine what is represented at the level below (Williams, 2019). Top-down projections are inhibitory, as predictions suppress prediction errors (Denève & Jardri, 2016). Whether they succeed in doing so depends on the relative precisions of priors and errors.

PP-based models of schizophrenia cannot abide by most of these commitments. First, within PP, illusory percepts arise due to constraints imposed on perceptual processing by priors learned empirically over time (Hohwy, Roepstorff, & Friston, 2008; Notredame, Pins, Deneve, & Jardri, 2014), for example, that faces are convex. These priors capture long-term regularities in the perceptual stream and are acquired over a lifetime of experiences; as such, they should be represented at much higher levels than context-dependent, short-term, and newly learned associations (e.g., between wearing glasses and the direction of rotation of the stimulus; Schmack, Rothkirch, Priller, & Sterzer, 2017). Nonetheless, these contextual expectations are taken to be "higher-level cognitive beliefs" (Sterzer et al., 2019, p. 138). This should be impossible according to the PP principle that higher-level beliefs are learned over large timescales. Note that this holds even if we consider that structural priors may be easily overridden at relatively short time scales by context-dependent priors after repeated exposure inconsistent with the structural prior (see Colombo, 2018; Seriès & Seitz, 2013 for example). Regardless of the nature of the interplay between structural and context-dependent priors, the former should be represented at higher levels.[4] Moreover, PP-based models of psychopathology, highlighting "potentially different roles of high and low levels of the hierarchy" (Sterzer et al., 2018, p. 639), make recourse to a classic distinction between a "lower-level perceptual system" (with certain expectations

hard-wired within) and a "higher-level cognitive system," effectively doing away with the continuity of the hierarchy.

Second, proponents of PP-informed schizophrenia models do not specify how higher-level cognitive beliefs could compensate for imprecise low-level priors. Corlett et al. (2019, p. 4; our emphasis) suggest that their influence is direct: "*At the same time*, precise cognitive *priors at a higher hierarchical level will sculpt perception*, subtending hallucinations and the maintenance of delusions." This is, however, impossible from the PP perspective, as it would violate the assumption that the influence of non-adjacent representations must be mediated by the levels in between (i.e., lower-level perceptual priors). Moreover, the influence cannot be indirect, since imprecise predictions conveyed at lower-level sensory processing stages would fail at constraining in accordance with higher-level demands. Simultaneous compensatory influences from higher-order levels are not provided for in hierarchical architectures—all computational mechanisms proposed in dynamic predictive models of psychopathology, for example, deep Boltzmann machine (Corlett et al., 2019), hierarchical Gaussian filter (Powers et al., 2017), belief propagation algorithm (Denève & Jardri, 2016), and Bayesian predictive coding (Sterzer et al., 2018), only allow information to pass directly between neighboring levels.

Third, PP suggests that top-down influences should be essentially suppressive: Backward connections are inhibitory or modulatory, and forward connections are excitatory (Kanai et al., 2015). Conversely, Sterzer et al. (2018, p. 638) propose that higher-level abstract or semantic beliefs *enhance* or *facilitate* signals in sensory cortices, driving hallucinations. PP does not currently offer a computational mechanism that could underlie such top-down excitatory signaling.

Importantly, all these difficulties cannot be simply explained away by the fact that "priors at low and high levels may be differentially affected" (Sterzer et al., 2018, p. 638). Certainly, the predictive system could be biased (yet stable) due to systemic perturbations of inferential loops and the resulting local differences in prior/prediction error precision ratios. For instance, Jardri and Denève (2013) implemented a belief propagation algorithm in a hierarchical neural network and impaired its inferential loops by altering the excitatory-inhibitory balance, which resulted in a circular inferential pattern. Results of successive iterations of the perturbed network squared with actual behavior observed in schizophrenic patients (e.g., overconfidence in beliefs despite weak evidence). However, one cannot argue for the fruitfulness of a particular paradigm based on successful applications of similar yet distinct theories. Belief propagation differs from PP. In the former, inter-level connections are essentially excitatory, whereas inhibitory interneurons control the informational flow, canceling out reverberated signals (so prior beliefs could not be fed back to the upper levels, which would result in overcounting of the priors; Denève & Jardri, 2016). In PP, feedback connections are inhibitory. Thus, impaired inhibition in belief propagation networks leads to circular inference, resulting in either strong or weak priors (depending on the loops that are impaired), whereas in PP it necessarily leads to weak priors (Jardri, Duverne, Litvinova, & Denève, 2017).

4. Avoiding empirical validation

A recent surge of reports further suggests that PP models of psychopathology may actually be misled. For example, the acquisition of priors from visual scene statistics is not affected in patients with schizophrenia (Kaliuzhna et al., 2019; Valton et al., 2019). However, even though PP models of schizophrenia are not indicative of unificatory powers of PP, there are good reasons to applaud the actions of their proponents. These models offer computational rigor and generate clearly formulated predictions which are being experimentally tested; when they do not match the data obtained, the authors call for the revision of their models (e.g., Kaliuzhna et al., 2019). Iteration of this process may result in a narrowing of the set of possible interpretations of core terms or tenets, which may clarify the conceptual territory.

Unfortunately, this approach to modeling is an exception rather than a rule in the PP universe. As we show in the section below, researchers tend to pick low-hanging fruits, creating numerous theoretical re-descriptions of psychological, psychiatric, and cognitive models in liberally interpreted PP terms, and interpreting the results of their studies as *consistent* with loose commitments of PP, even in the case of available alternatives or internal inconsistencies of PP-based explanations. We argue that mere consistency does not validate the approach. Conversely, this "horizontal" trajectory of development of PP (theoretical re-descriptions and ubiquitous post-hoc just-so stories) only exacerbates the extant problems: The resulting models are mutually exclusive, inconsistent with PP tenets, or general enough to accommodate almost every empirical finding, which makes the theory even more heterogeneous, or monstrous. Certain argumentative strategies are also used to justify these re-descriptions that offer scarce, if any, new empirical predictions. In this section, we aim to expose them.

4.1. Horizontal development and just-so stories

PP is not easily applicable to cognition, which is frequently detached from immediate perceptual input. As a result, such applications face serious theoretical challenges (for a discussion, see Williams, 2018a, b) and mostly take the form of conceptual or verbal models (Kwisthout, Bekkering, & van Rooij, 2017), as it is unclear how to approach the problem computationally. While computational or mechanistic models are scarce, conceptual and verbal theories proliferate. To name just a few, PP-based theoretical models of distorted cognitive processing in depression (Kube et al., 2019), meditation (Lutz, Mattout, & Pagnoni, 2019), perceived injustice in chronic pain (Bissell et al., 2018), drivers' behavior (Engström et al., 2018), and religion and spirituality (van Elk & Aleman, 2017) have been recently proposed. The last model will be analyzed below.

van Elk and Aleman (2017) present PP as an alternative for dual-process accounts of religiosity. In particular, the authors propose that their model explains four phenomena associated with religiosity and spirituality, for which four corresponding "neurocognitive mechanisms" are outlined:

1. Religious visions and hallucinations. Drawing from PP approaches to schizophrenia, they propose that visions and hallucinations may arise from imprecise coding of efferent signals and low-level priors, rendering sensory activations unpredictable. The volatility of low-level priors results in overreliance on context- and culture-informed religious or delusional beliefs.
2. Mystical experiences. Self-transcendental feelings and associated reduced body awareness stem from differential weighing of exteroceptive and interoceptive signals, with significantly diminished precision of the latter.
3. Personal experiences of a godly presence. Feelings of supernatural presence come from simulated offline inferences of God's mental states, based on internal generative models linking social cues to the interoceptive states of others.
4. Acceptance and maintenance of religious beliefs. Reduced error monitoring processes facilitate increased reliance on high-level priors, which results in a tendency to accept and sustain prior religious and spiritual beliefs, and a decreased ability to update priors in the face of conflicting sensory information.

van Elk and Aleman (2017) provide an exceptionally comprehensive overview of neurocognitive studies on religiosity and spirituality to argue that PP allows theoretical integration of all past findings. However, their claim is unsubstantiated. The argumentation is based on a collage of reinterpretations of results from neuroscientific studies and re-descriptions of hypothetical religious visions in PP terms (e.g., mystical experience of a mountaineer, p. 366) as well as arguments intuitively speaking to the reader (e.g., the fact that contents of religious and spiritual experiences are culture-informed and shaped by previous experiences is *in line* with PP accounts predicting the important role of prior beliefs). Thus, the detailed neurocognitive mechanisms announced in the introduction are not actually specified, and PP merely provides a new language with which to talk about diverse phenomena. To make matters worse, the authors tell an inconsistent story using this language, for instance, their references to the distinction between rigid high-level priors sculpting perception and imprecise perceptual priors violate central PP tenets (see Section 3.2). Moreover, neuroscientific findings which are "broadly congruent" (p. 368), "broadly compatible" (p. 368), or "in line" (p. 371) with PP are taken to "substantiate [the] model with empirical evidence" (p. 360; see Section 4.2 for a thorough treatment of the consistency fallacy in PP models).

What matters for evidential support is whether the model actually fits the data better than alternatives. van Elk and Aleman (2017) do not show that—they simply use PP as a generic language to talk about disparate aspects of religiosity and spirituality. A conceptual model, which consists of speculative (and inconsistent) re-description of to-be-explained-phenomena, does not deepen our understanding of them. Therefore, PP proponents harness certain argumentative strategies to justify their introduction. These patterns of rhetoric are repeatable and surface in many recent PP models (see Table 1). In particular, the models are presented as follows:

1. *An answer for the need to unify the scattered field of study*—The model is not presented as an alternative or a new approach, but as an umbrella theory gathering all

Table 1
Repeatable Rhetorical Strategies in PP-Based Theoretical Models

PP-Based Model of	Answers the Need for Unification	Apparent or Post-Hoc Predictions	Model as a Starting Point
Religiosity and spirituality; van Elk and Aleman (2017)	"There is currently no up-to-date review and integrative framework that accounts for the different findings that have been reported in the literature (. . .) Our proposed model is unique as it provides a unifying account of the neurocognitive basis of religiosity and spirituality thereby integrating recent findings from different fields."	"Imprecise predictions may result in a failure to properly update one's prior models, while hyper-precise prediction error signals may result in a malfunctioning learning process potentially leading up to delusional beliefs. (. . .) Thus, our theoretical framework makes testable predictions about when we may expect the development and maintenance of fixed belief systems."	"(. . .) research is needed before stronger conclusions on the role of dopamine in religious beliefs can be reached. Thus, the notion that error monitoring mechanisms play a central role in adopting and sustaining religious and paranormal beliefs and the supposed involvement of the dopaminergic system in this process opens interesting avenues for future research."
Drivers' behavior; Engström et al. (2018)	"Several of the general concepts underlying the present framework are accounted for by existing human factors frameworks and models (. . .). Thus, predictive processing should not be viewed as a radical alternative to these existing models but rather as a framework for bringing together different strands of human factors research based on the unifying principle of prediction error minimisation."	"Finally, we discuss how predictive processing concepts may help to understand drivers' interaction with automatic steering interventions and AD functions. These examples are intended to provide a first illustration of how the proposed framework can improve our understanding, and generate testable hypotheses of different aspects of driver behaviour, and encourage its application to other driving-related phenomena."	"The present paper is intended as a first exploration of the application of predictive processing to driving and we hope that it will encourage others to apply and further develop these ideas. Efforts towards more specific quantitative driver behaviour models based on predictive processing concepts are currently underway and will be reported in future publications."

(continued)

Table 1. (continued)

PP-Based Model of	Answers the Need for Unification	Apparent or Post-Hoc Predictions	Model as a Starting Point
Self-recognition; Apps and Tsakiris (2014)	"Recent reviews of this literature have concluded that the absence of a unifying theoretical framework has resulted in a largely incoherent picture of the circuits and mechanisms which are engaged during self-recognition. (. . .) In this paper we attempt to highlight how the free-energy principle, a recent attempt at a unifying theory of the brain, can explain many previous findings in self-recognition research."	"Our third prediction was that that there will be a suppression of activity when a self-stimulus is predicted or when a self-stimulus leads to the expectation of a sensory event. Evidence is provided of such a notion by research examining self-touch. A seminal paper (. . .) found that participants cannot experience a tickling sensation when they apply tactile stimulation to their own skin, only when tactile stimulation is externally delivered."	"We have provided evidence in support of our theory, although we note that to date, empirical data neither largely supports nor refutes our account of self-recognition. However, this work does provide a broad range and extensive set of prediction about the nature of self-recognition that can be tested empirically. We hope that such empirical investigation will generate important and novel findings."
Emotion in action; Ridderinkhof (2017)	"These proposals are consistent with, on the one hand, Frijda's recent views on emotion vis-à-vis impulsive action (. . .), and, on the other hand, the principles and mechanisms of perception–action coordination laid out in a recent integrative theoretical framework (. . .) Thus, this article aims at a synthesis that integrates previous work and extends it with the notion of forward modeling."	"Forward modeling and its computational bases have been developed extensively in the literature on motor control (. . .) and have recently been elaborated in the literature on predictive processing (. . .) Such a framing may help generate novel hypotheses, that allow for empirical tests (. . .)."	"Rephrasing questions about emotional behavior in terms of underlying constructs and mechanisms of information processing may not in itself add much explanation. Our aim here was to evaluate whether, from an action control perspective, a meaningful integration and synthesis with emotion theory is feasible, at least at the level of the conceptual components. It is our hope that the present theoretical synthesis, and in particular its inclusion of forward modeling, may engender a deeper understanding of emotion in action (or at least a framework from which novel and testable predictions can be derived)."

present models of a given phenomenon. While specific models are good at explaining particularities, they cannot be generalized to the whole field of study. This is what its proponents claim PP does: allow explanatory pluralism to be replaced by a single unifying account.

2. *Generating new, specific and testable predictions*—This is certainly to be expected from a new model or theory. PP proponents claim to specify such predictions in their papers; however, upon a closer look, these "predictions" are not framed as hypotheses or measurable phenomena that are interpretable without additional PP assumptions. For example, van Elk and Aleman (2017, p. 371, our emphasis) refer the reader to a table (p. 364) for "*hypotheses* and *theoretical predictions* to be addressed in future research." However, the table contains only open research questions (such as "Do mystical and self-transcendent experiences rely on a differential weighting of exteroceptive compared to interoceptive information for an inferred model of the bodily self?").

Despite the authors' claim that their account "makes testable predictions" (p. 372), not a single specific, testable hypothesis is elaborated. Instead, they re-discuss their theoretical model and refer to assumptions stemming from other PP-based models (e.g., that biased learning processes arise from imbalances between precisions of priors and predictive errors, caused by a dopamine system malfunction). Thus, the model presupposes that the brain works in a certain way to guide empirical research intended to decide whether it actually works this way, by giving rise to particular cognitive phenomena. The account cannot be tested without these assumptions determining the interpretation of what the observables mean, for example, that top-down modulations reflect prior beliefs explaining away prediction errors. PP serves as a theory and interpretative context at the same time. As a result, its actual hypothesis-generating potential is minimal.

3. *A starting point*—re-description of observables (e.g., behavior, neuroscientific data) in terms of hypothetical and unobservable underlying processes is presented as a first step toward comprehensive understanding of a given phenomenon. Proponents of PP models postpone the verification of their models, hoping that their proposition will inspire other scientists to carry out further research.

Summing up, such "theoretical proposals" (van Elk & Aleman, 2017, p. 362) or "purely conceptual account[s]" (Engström et al., 2018, p. 174) do not extend our understanding of re-described phenomena. They may actually be a step backwards compared to heavily criticized Bayesian models (e.g., Bowers & Davis, 2012; Jones & Love, 2011). While Bayesian models and their assumptions tend to be fitted post-hoc around data, they are at least quantitative and allow simulations. On the other hand, PP-informed models predominantly take the form of theoretical reinterpretation of data in light of such post hoc assumptions. This reinterpretation is simply an application of a theory to interpret a wide range of phenomena, and does not entail a common explanation for that range (Colombo, Elkin, & Hartmann, 2018). In other words, PP is not applied systematically, and it could become monstrous if these applications are purely post hoc.

4.2. Consistency fallacy in PP models

In neuroscience, conclusions that observed data provide evidence for a given hypothesis often come from over-confirmatory research and data interpretation strategies, rather than being substantiated by data (Mole & Klein, 2010). In particular, researchers are prone to the "consistency fallacy" (Coltheart, 2013): They claim that they found *support* for a given theory after collecting data which are merely *consistent* (or *not inconsistent*) with this theory. However, this is not sufficient for confirmation. Consistent data might come from observations from an unrelated set (e.g., observation of black ravens is consistent with the theory that all bears are brown) or their consistency may rely on the theory's degree of adaptability. For example, let us consider the example of psychoanalysis. The fact that the client denies that his or her psychological struggle stems from a dysfunctional relationship with his or her father is consistent with the hypothesis that this relationship is actually a root problem, as associated adverse feelings may be psychologically repressed. On the other hand, acknowledgment and willingness to talk about father issues may be a sign of insight. Any kind of patient behavior can be consistent with the original hypothesis of a psychoanalyst.

Thus, a body of data cannot provide evidence for a theory only in virtue of its consistency. It must also weigh against the competing hypotheses. The data d support a hypothesis only if the conditional probability of hypothesis H, given that data d were accurately observed, exceeds the conditional probability of an alternative (i.e., $P(H|d) > P(\text{non-}H|d)$). If competing hypotheses are equally capable of accommodating the data, the study is non-diagnostic. Coltheart (2013) proposes that this can be avoided by the use of proper study and data analysis design: Prior to the experiment, one should specify observable outcomes that (a) would speak in favor of the tested theory (and against the contradictory one) and (b) are inconsistent with the tested theory (and accommodated by the contradictory one). When no alternative theories are present, at least one outcome irreconcilable with the tested theory should be possible. For non-experimental work, such as theoretically driven simulation, at least one alternative model that accounts for the same results should be specified. This alternative model could be extremely simplistic, but what is important is that all testing is comparative (Sober, 1999), so it requires a baseline for comparison.

Although the consistency fallacy is evident in neuroscience (see Mole & Klein, 2010 and Coltheart, 2013 for example), we argue that its prevalence in PP models leads to inevitable contradictions between studies on the same phenomenon. We focus on studies on neural underpinnings of the rubber hand illusion (RHI; Limanowski & Blankenburg, 2015; Zeller, Friston, & Classen, 2016), which supposedly support a PP model of body ownership and self-recognition (Apps & Tsakiris, 2014).

Limanowski and Blankenburg (2015) employed dynamic causal modeling (Friston, Harrison, & Penny, 2003) to test nine neurodynamic models of information exchange between neural structures involved during RHI. All models had a two-level structure, with visual (lateral occipital cortex, LOC) and somatosensory areas (secondary somatosensory cortex, SII) separately providing sensory input to a multisensory integration hub (intraparietal sulcus, IPS) at the upper level. The models also allowed for inter-level connectivity between

the IPS and premotor ventral cortex (PMv), and they were divided into three families, depending on the effective connectivities that were predicted to be strengthened during the illusion compared to the control condition: (a) bottom-up (from LOC/SII to IPS), (b) top-down (from IPS to LOC/SII), and (c) bidirectional. One of the bottom-up models strongly outperformed competing models. According to the authors, spatiotemporal congruence of visual and somatosensory signals drives the cognitive system to assume their common cause. This results in intersensory conflict between touch represented visually on a dummy and felt on one's real arm, which elicits prediction errors. These errors are streamed from the LOC to upper-level multisensory integration regions. The IPS tries to counter this mismatch via switching of somatosensory coordinates onto the visual reference frame. These adjusted coordinates are also signaled to the SII, where they do not match the somatosensory information (e.g., proprioceptively encoded position of one's arm), thus eliciting further prediction errors.

In another paper on the topic, Zeller et al. (2016) used the same approach and proposed their own neurodynamic model in which RHI arises as a result of attenuation of intrinsic connections in the primary somatosensory cortex (SI) and enhanced effective bottom-up connectivity from the contralateral occipital cortex (OC) to the contralateral ventral premotor cortex (PMC). The authors continue the story told by Limanowski and Blankenburg (2015): To suppress prediction errors arising in SI, top-down modulations from PMC attenuate their precision, which offers greater weight (enhanced attention) to visual signals.

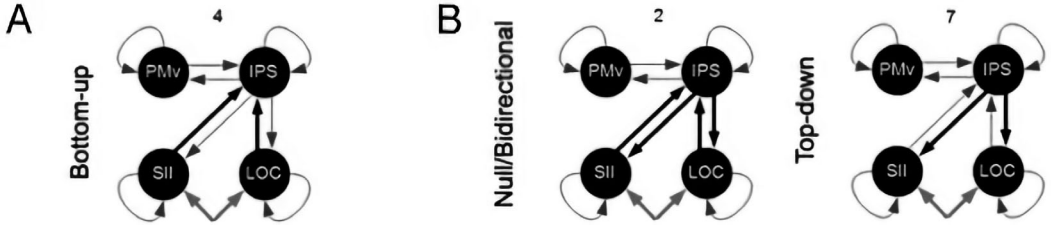

Fig. 1. Consistency fallacy in the study performed by Limanowski and Blankenburg (2015). All dynamic models reprinted from Limanowski and Blankenburg (2015). (A) A much simpler explanation could be proposed for the winning bottom-up model: Spatiotemporally congruent stimulation results in enhanced signaling from lower-level perceptual cortices to the multisensory integration area. This interpretation is particularly appealing given no differences in top-down signaling, which should appear if predictive processes were at work. (B) Kindred PP interpretations of the results could be proposed for bidirectional and top-down models. The narrative only makes particular effective connectivities more or less relevant. Consider the following examples from the authors: *Bidirectional*: Intersensory conflict between visual and tactile signals elicits visual prediction errors from LOC to IPS. Top-down signals from IPS counter the mismatch through recalibration of somatosensory coordinates on a rubber hand and signal new coordinates to SII. This elicits somatosensory prediction errors, as somatosensory signals do not match new coordinates. *Top-down*: Intersensory conflict between visual and tactile signals elicits visual prediction errors from LOC to IPS. For the illusion to arise, top-down signals from IPS suppress these errors through recalibration of somatosensory coordinates on a rubber hand, and signal new coordinates to SII. Somatosensory prediction errors—arising due to switched coordinates—are further suppressed by top-down signals from IPS. Otherwise, they would break the illusion.

Both research teams succumb to a consistency fallacy in a number of ways. First, data observed by Limanowski and Blankenburg (2015) do not make improbable a competing hypothesis that congruent stimulation simply evokes multisensory integration processes. The authors themselves acknowledge this, writing that "predictive coding is only one candidate explanation for the mechanisms underlying the brain's hierarchical inference about the world and the body" (p. 2301). Both the PP-based account and the traditional multisensory integration approach seem quite likely in the face of the data (Fig. 1A). However, the authors waive off this concern with a short remark that the alternative hypothesis is not inconsistent with PP. And Zeller et al. (2016) did not discuss any competing theory or hypothesis.

Second, models potentially incompatible with PP were not specified in either study. This raises the suspicion that post-hoc PP interpretations could be tailored to virtually all possible outcomes (see Fig. 1). After all, in line with the interpretation of Limanowski and Blankenburg (2015), one could also expect top-down modulations suppressing errors in perceptual areas to come forward. Yet, in the winning model, top-down effective connectivity from IPS to LOC was present regardless of the experimental context (illusion/control), and effective connectivity from IPS to SII was altogether absent. Finally, prediction errors from LOC to IPS should be suppressed for the illusion to arise (according to authors); if so, why does error-related bottom-up effective connectivity from LOC to IPS arise? None of these problems is even discussed in the article, which prompts us to ask: Could the very same interpretation be fitted post-hoc to each of the nine models tested (Fig. 1B)? Nonetheless, the authors "interpret these results as support for a predictive coding account of hierarchical inference in the brain" (Limanowski & Blankenburg, 2015, p. 2285) and other researchers follow (e.g., Tsakiris, 2017, p. 604).

The study by Zeller et al. (2016) succumbs to a consistency fallacy in a very similar fashion. None of the 32 tested models was defined beforehand as an outcome potentially inconsistent with PP, and the interpretation of the results is highly questionable. None of the chosen models allowed for effective top-down connectivities; therefore, such modulations should be excluded from explanation. Since "the choice of model families was motivated by our hypotheses" (p. 269), the authors must have earlier considered such top-down modulations irrelevant. Moreover, two effects of interest were specified: CONGRUENT-REAL (differences between RHI and condition in which participants were simply observing their own hand) and INCONGRUENT-REAL (differences between a control palm-up condition and a simple observation of one's own hand), with the latter "refer[ring] to the visual perception of an artificial hand without an associated feeling of ownership" (p. 268). Enhanced connectivity from left OC to left PMC was much more pronounced in the INCONGRUENT than the CONGRUENT condition and, accordingly, should be interpreted as associated with visual processing of the hand or incongruence rather than ownership. Therefore, claims that "*during illusory perception*, greater precision is afforded to visual input (from the OC)" (p. 270) or that the results "highlight an increase in ascending visual influences (e.g., prediction error) to multimodal sources *during illusory percepts*" (p. 272, our emphases) are simply unwarranted, even if the OC-PMC connection was stronger during the CONGRUENT than during the REAL condition.

The consistency fallacy is detrimental to scientific progress, since it creates an impression that the theory was empirically corroborated—when it was not. It also leads inevitably to contradictions between various parts of the theory-driven explanation. For example, the enhanced top-down effective connectivity from IPS to LOC, present regardless of the experimental condition in the first study (Limanowski & Blankenburg, 2015), was interpreted by the authors as enhanced top-down visual attention to the rubber hand. However, their winning model predicted the attenuation of intrinsic connectivities in both the LOC and SII. Attenuated intrinsic connectivity in the somatosensory cortex was also found and expressly interpreted as reduced precision weighting by Zeller et al. (2016). If this interpretation is true, top-down visual attention could not result in lowered intrinsic connectivity in the LOC, as it increases the gain of error units and precision weighting of bottom-up signals (Hohwy, 2012). Thus, even though both models were taken to support the same theory of body ownership, they are irreconcilable in its own terms. As opposed to other inconsistencies between them (e.g., pertaining to presence or absence of particular effective connectivities), this cannot be explained away by methodological differences between the studies.

The consistency fallacy may even lead researchers to claim that they have found support for PP in the face of results incongruent with the theory on the behavioral level. Once again, let us consider Schmack et al.'s (2017) study on schizophrenia that found that the effect of experimentally induced association between wearing glasses and the direction of the moving stimulus was *weaker* in patients than in healthy controls. Nevertheless, the authors focus on enhanced effective top-down connectivity from the orbitofrontal to visual cortex, which they assume, based on the very theory they committed to experimentally validate, to be belief-related. On that basis, they conclude that their results are in line with predictive coding accounts.

Summing up, the consistency fallacy is prevalent in empirical studies devised to test PP. As long as outcomes compatible and incompatible with both PP and competing theories are not specified prior to the experiment (e.g., in the form of a preregistered study), PP cannot obtain legitimate empirical support. Assumptions about what a neural signal actually means should also be clarified to avoid post-hoc just-so stories and resulting contradictions. Otherwise, PP-based neuroscience will remain a muddled territory with theories that are mutually inconsistent and irreconcilable with PP's core tenets.

5. Possible objections

Defenders of PP might object to our criticisms in several ways. Here, we provide short replies to possible objections.

All research programs are chaotic when they grow, and faulty application is to blame. One could object that our criticism is too harsh: All new theories incite new work, which is not always of the highest quality. Hence, our criticism misses the mark, as there is nothing special about PP. All grand research programs must deal with this problem. Some researchers are simply insufficiently cautious in its application to various phenomena.

Indeed, this is a reasonable point—rapidly developing approaches may incite many scientists to jump on the bandwagon. However, we are unaware of any criticism from within the PP community of overeager attempts to stretch the theory to fit every possible application. Thus, we believe that for this objection to be successful, PP proponents should proceed with more caution—as of today, they neither control the inflow of theoretical models nor critically evaluate them. The mere growth of the scope of theory—the growing number of phenomena "covered" by PP—is actually being used as an argument in the debate over the unificatory potential or explanatory usefulness of PP (e.g., Clark, 2016; Kiefer & Hohwy, 2018).

PP should not be considered in isolation from the FEP (the free energy principle; Friston, 2009) *or other Bayesian approaches*. Our critic might also stress that predictive coding models are usually motivated by free energy considerations or Bayesian approaches to rationality or adaptive behavior.

Indeed, it was argued that FEP may be viewed as a first principle that makes PP a cognitive architecture of all living systems (Colombo & Wright, 2018); as such, it grants PP its unificatory credentials. However, such universal constraints on cognitive organization —specifying, for instance, what "prior beliefs" or "sensory states" generally are—would have to be extremely liberal to apply to all organisms (from single-cell organisms through to humans). PP provides details on cognitive architecture that go beyond the assumptions of FEP (Gładziejewski, 2019). Moreover, even though there are several versions of PP, for example, stated in terms of Bayesian brain dynamics, or merely as predictive coding algorithms (Spratling, 2017), none logically requires the FEP to be true, even if it could provide a theoretical background for solving problems with the fundamentals of PP. Even supposing the main job of the brain is not to minimize (informational) free energy, or even to provide accurate Bayesian inference, or form predictions for motor control, one could still adhere to PP to account for the way cognitive processes are performed (see, e.g., Thornton, 2017). Thus, the potential unifying power of PP does not seem to depend on external theoretical virtues.

Not all proponents of Bayesian approaches defend the idea of the grand unification. Our criticism is aimed at the claim that PP can satisfactorily unify theories of cognition. But not all proponents of the theory necessarily have such ambitions (Clark, 2013).

We do not deny this. Among PP proponents, we may distinguish "neats"—who posit that all cognition arises from the imperative to minimize informational uncertainty in hierarchical PP architectures—from "scruffies," who perceive PP as a framework or even a toolbox. This toolbox could be a common stock of algorithmic specifications and concepts that find their precise empirical meanings in individual models (Clark, 2013). If PP is a toolbox, then its main virtue lies in its fruitfulness in displaying commonalities across phenomena. We actually sympathize with the scruffies: PP might be a useful theory even if it is not necessarily unifying to a high degree. However, tools from any toolbox are used for particular purposes. In this case, the issues with the processing hierarchy we indicated above (see also Williams, 2018a, 2018b, 2018c) could be circumvented by limiting the intended scope of the theory to perception and action,

and by excluding cognitive or psychopathological phenomena which remain problematic for PP.

However, "scruffiness" does not relieve one from a duty to provide good and coherent scientific explanations. Our critique is aimed at showing that many PP proponents fail to do so: They confuse subjective and computational orders of description, provide inconsistent and mutually exclusive models, do not understand (or regard) the core tenets of the theory, ignore the fact that these tenets are at odds with empirical data, propose untestable theoretical re-descriptions generating only apparent predictions, and serially adhere to the consistency fallacy. These are the signs of the misapplication of a theory's tools, regardless of one's individual view on the unificatory potential or the scope of applicability of a given theory. We agree that it is particularly problematic if one is anointing PP to the role of a grand overarching theory in cognitive science, but we do not believe that our concerns may be easily waived off in this way. A "scruffy" should also be committed to work with reliable explanations.

Fundamental and conceptual problems may not be targeted first—the progress goes the other way around in science. Our opponents may insist that our professed solution to problems that plague PP—focusing on theoretical fundamentals—is premature.

While it is certainly true that there is no special preference to be given to one sequence of solving scientific problems ("general first" or "particular first"), our focus is on the unificatory claims of PP. They will remain dubious unless the theoretically fundamental issues are addressed. These include not only the conceptual and theoretical confusions we point out in Section 3 but also crucial problems with neuroscientific evidence. For example, it is questionable that top-down modulation in the brain is essentially suppressive; it often facilitates neural responses (Denève & Jardri, 2016). It is unlikely that one could also define one homogeneous processing hierarchy for all perceptual and cognitive processes (Williams, 2018b), in either theoretical or neurobiological terms.

We agree that solving particular problems is essential to the progress of PP. This is why we urge its supporters to develop empirically validated models of cognitive phenomena that remain especially problematic to PP (such as thinking and its pathological forms; cf. Williams, 2018a, 2018b). The success in developing them may depend on addressing deeper theoretical issues as well, however.

Unifying accounts may be desirable even if they do not diverge or add any explanatory advantage, as they bring all explanations under a common unifying theory.

The core of this objection is probably aimed at the potential virtue of unification: A unified theory may fail to provide new empirical predictions but offer pure theoretical progress (Gładziejewski, 2019). One could, for example, point out that Copernicus did not provide better calculations or new predictions than Ptolemy, but his theory was much more systematic. This is in line with our take on unification: Systematization of a class of phenomena may indeed increase unification without providing any new predictions or explanations. However, it should contribute to a deeper understanding of phenomena, for example by providing their taxonomy. While there are debates over what exactly scientific understanding is, at a minimum it should provide substantially new inferences about phenomena in question, for example about their connections (for a comprehensive review

of the issue of unificatory understanding, see Regt, 2017). Mere re-description is insufficient for this; all it provides are new synonymous terms and no other inferentially relevant information. For example, we cannot fathom any new inferences licensed by the models of religiosity or drivers' behavior discussed in Section 4.

6. Conclusions

In this paper, we have argued that the development of PP as a unificatory theory of perception, action, and cognition is stalled. This is because of two gaps: one between PP as a theory in cognitive science and its computational implementation, and another between this theory and its biological underpinnings. The mere fact that predictive algorithms are specified mathematically does not suffice to make a strict theory; on the contrary, the fundamentals of PP remain unclear, and the trajectory of its development seems to be in stark contrast to the unificatory claims of its proponents.

In our analysis, we assumed that unified theories should display at least four features. They should be as general as possible, while remaining simple, homogeneous, and systematic. PP, however, fails to be homogeneous because its computational framework does not sufficiently constrain numerous and diverse interpretations in theoretical and biological terms. Technical terms and computational structures are posited in a fairly ad hoc manner, and while it is certainly possible to re-describe a number of known models in PP terms, these re-descriptions are not driven by sufficiently precise theoretical considerations. This has led to a proliferation of various models that apply PP terms liberally, or even contradict its basic assumptions. Unfortunately, the known problems of PP remain unaddressed, and defenders of PP as a unified theory seem satisfied with merely noting that the number of PP models grows over time. But this growth, we claim, is a source of conceptual confusion rather than evidence of ongoing unification.

To argue for these points, we showed that crucial PP concepts, such as precision, are interpreted in diverse ways. Moreover, the modeling practices of PP defenders are quite lax regarding the theoretical principles of PP, which gives rise to models that contradict the theory's avowed fundamentals. A deep unified theory cannot be compatible with contradictory computational models that are supposed to implement it; otherwise, it is either underspecified or contradictory itself. It is definitely not homogeneous in such a case.

We also showed that PP models are not validated empirically and are usually stated as just-so stories, or re-descriptions of various phenomena. Worse, instead of validating them against possible alternatives, proponents of PP are content to point out that PP does not seem to be inconsistent with (selected) empirical evidence, usually from neuroscience. But the assumption that consistency is sufficient for evidential support is fallacious.

We stress that our focus is on PP as a unified theory, and our criticisms are not directed against using it as a computational framework. Indeed, PP is frequently used in this way in scientific practice, without any concern for its theoretical underpinnings. This framework, however, fails to be a unified theory of cognition, action, and perception, and it is not being developed currently in the way that could license a reasonable expectation

that it would become such a theory in the future. What it actually might become, given the current diversity of approaches within the PP community, is rather a research program or tradition in the sense of Laudan (1977), encompassing multiple alternative and mutually exclusive theories. While these theories may still have some unifying power (Miłkowski & Nowakowski, 2019), diversified research traditions are unlikely to provide most general, simple, homogenous, and systematic explanations of their phenomena of interest. Certainly, it is possible for one particular theory to attain a dominant position within the PP research tradition, but it will not be unificatory as long as the gap between the mathematical formalism of PP and its (neuro) cognitive interpretations remains unaddressed.

Acknowledgments

The authors would like to thank Mark Bickhard, Paweł Gładziejewski, Tomasz Korbak, Jacek Malinowski, Vincent Müller, Maxwell Ramstead, and Leszek Wroński for helpful comments on earlier versions of the manuscript. We would also like to express our gratitude to the Reviewers and the Editor for their in-depth comments and suggestions which led to a major improvement of the paper.

Funding Information

The work in this paper was funded by a National Science Centre (Poland) research grant under the decision 2014/14/E/HS1/00803. The funding source had no involvement in the writing of the manuscript or decision to submit it for publication.

Conflict of interest

None.

Notes

1. Note that in cognitive science, the terms *theory* and *framework* are often used interchangeably, as evidenced by citations of defenders of PP who consider it theory in our sense but call it a framework (e.g., citations in Section 4.1, Table 1).
2. For example, Veissière and Stendel (2018, pp. 4–5) write: "Cravings, on this view, could be conceptualized as prediction errors."
3. Note that this makes the main theoretical term of PP vague: the scope of the proper application of *hierarchy* is unclear, as its conditions of accuracy are only dimly

sketched. Even worse, the organization of computational mechanisms in the brain that could be responsible for hierarchical processing is unclear.

4. The fact that they are so easily overridden by context-dependent and newly learned associations is actually another problem for PP.

References

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

Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., & Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *ELife*, *5*, e18103. 10.7554/eLife.18103

Apps, M. A. J., & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, *41*, 85–97. https://doi.org/10.1016/j.neubiorev.2013.01.029

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*(7), 419–429. https://doi.org/10.1038/nrn3950

Bissell, D. A., Ziadni, M. S., & Sturgeon, J. A. (2018). Perceived injustice in chronic pain: An examination through the lens of predictive processing. *Pain Management*, *8*(2), 129–138. https://doi.org/10.2217/pmt-2017-0051

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

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences*, *3*(2), 57–65.

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

Clark, A. (2015). Radical predictive processing. *The Southern Journal of Philosophy*, *53*(S1), 3–27. https://doi.org/10.1111/sjp.12120

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

Colombo, M. (2018). Bayesian cognitive science, predictive brains, and the nativism debate. *Synthese*, *195*(11), 4817–4838. https://doi.org/10.1007/s11229-017-1427-7

Colombo, M., Elkin, L., & Hartmann, S. (2018). Being realist about Bayes, and predictive processing. *The British Journal for the Philosophy of Science*. Advance online publication. https://doi.org/10.1093/bjps/axy059

Colombo, M., & Wright, C. (2017). Explanatory pluralism: An unrewarding prediction error for free energy theorists. *Brain and Cognition*, *112*, 3–12. https://doi.org/10.1016/j.bandc.2016.02.003

Colombo, M., & Wright, C. (2018). First principles in the life sciences: The free-energy principle, organicism, and mechanism. *Synthese*. Advance online publication. https://doi.org/10.1007/s11229-018-01932-w

Coltheart, M. (2013). How can functional neuroimaging inform cognitive theories? *Perspectives on Psychological Science*, *8*(1), 98–103. https://doi.org/10.1177/1745691612469208

Cooper, R., & Shallice, T. (1995). Soar and the case for unified theories of cognition. *Cognition*, *55*(2), 115–149. https://doi.org/10.1016/0010-0277(94)00644-Z

Copi, I. M., & Cohen, C. (1990). *Introduction to logic*. New York: Macmillan Publishing Company.

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K., & Powers, A. R. (2019). Hallucinations and strong priors. *Trends in Cognitive Sciences*, *23*(2), 114–127. https://doi.org/10.1016/j.tics.2018.12.001

Denève, S., & Jardri, R. (2016). Circular inference: Mistaken belief, misplaced trust. *Current Opinion in Behavioral Sciences*, *11*, 40–48. https://doi.org/10.1016/j.cobeha.2016.04.001

Dima, D., Roiser, J. P., Dietrich, D. E., Bonnemann, C., Lanfermann, H., Emrich, H. M., & Dillo, W. (2009). Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. *NeuroImage*, *46*(4), 1180–1186. https://doi.org/10.1016/j.neuroimage.2009.03.033

Engström, J., Bärgman, J., Nilsson, D., Seppelt, B., Markkula, G., Piccinini, G. B., & Victor, T. (2018). Great expectations: A predictive processing account of automobile driving. *Theoretical Issues in Ergonomics Science*, *19*(2), 156–194. https://doi.org/10.1080/1463922X.2017.1306148

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215. https://doi.org/10.3389/fnhum.2010.00215

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

Friston, K. J. (2018). Does predictive coding have a future? *Nature Neuroscience*, *21*(8), 1019–1021. https://doi.org/10.1038/s41593-018-0200-7

Friston, K. J., Brown, H. R., Siemerkus, J., & Stephan, K. E. (2016). The dysconnection hypothesis (2016). *Schizophrenia Research*, *176*(2–3), 83–94. https://doi.org/10.1016/j.schres.2016.07.014

Friston, K. J., Daunizeau, J., Kilner, J. M., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics*, *102*(3), 227–260. https://doi.org/10.1007/s00422-010-0364-z

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302. https://doi.org/10.1016/S1053-8119(03)00202-7

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., Dolan, R. J., Moran, R., Stephan, K. E., & Bestmann, S. (2012). Dopamine, affordance and active inference. *PLOS Computational Biology*, *8*(1), e1002327. https://doi.org/10.1371/journal.pcbi.1002327

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

Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, *98*(2), 254–267. https://doi.org/10.1037/0033-295X.98.2.254

Gilead, M., Trope, Y., & Liberman, N. (2019). Above and beyond the concrete: The diverse representational substrates of the predictive brain. *Behavioral and Brain Sciences*, 1–63. https://doi.org/10.1017/S0140525X19002000

Gładziejewski, P. (2019). Mechanistic unity of the predictive mind. *Theory & Psychology*, *29*(5), 657–675. https://doi.org/10.1177/0959354319866258

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, *3*, 96. https://doi.org/10.3389/fpsyg.2012.00096

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

Hohwy, J. (2017). Priors in perception: Top-down modulation, Bayesian perceptual learning rate, and prediction error minimization. *Consciousness and Cognition*, *47*, 75–85. https://doi.org/10.1016/j.concog.2016.09.004

Hohwy, J., Roepstorff, A., & Friston, K. J. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, *108*(3), 687–701. https://doi.org/10.1016/j.cognition.2008.05.010

Hoyningen-Huene, P. (2013). *Systematicity: The nature of science*. New York: Oxford University Press.

Jardri, R., & Denève, S. (2013). Circular inferences in schizophrenia. *Brain*, *136*(11), 3227–3241. https://doi.org/10.1093/brain/awt257

Jardri, R., Duverne, S., Litvinova, A. S., & Denève, S. (2017). Experimental evidence for circular inference in schizophrenia. *Nature Communications*, *8*(1), 14218. https://doi.org/10.1038/ncomms14218

Jęczmińska, K. (2017). Global workspace theory and sensorimotor theory unified by predictive processing. *Journal of Consciousness Studies*, *24*(7–8), 79–105.

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169–231.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

Kaliuzhna, M., Stein, T., Rusch, T., Sekutowicz, M., Sterzer, P., & Seymour, K. J. (2019). No evidence for abnormal priors in early vision in schizophrenia. *Schizophrenia Research*, *210*, 245–254. https://doi.org/10.1016/j.schres.2018.12.027

Kanai, R., Komura, Y., Shipp, S., & Friston, K. J. (2015). Cerebral hierarchies: Predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370* (1668), 20140169. https://doi.org/10.1098/rstb.2014.0169

Kiefer, A., & Hohwy, J. (2018). Content and misrepresentation in hierarchical generative models. *Synthese*, *195*, 2387–2415. https://doi.org/10.1007/s11229-017-1435-7

Kitcher, P. (1989). Explanatory unification and the causal structure of the World. In P. Kitcher & W. C. Salmon (Eds.), *Scientific explanation* (Vol. *505*, pp. 410–505). Minneapolis: University of Minnesota Press.

Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2019). Distorted cognitive processes in major depression: A predictive processing perspective. *Biological Psychiatry*, *87*(5), 388–398. https://doi.org/10.1016/j.biopsych.2019.07.017

Kwisthout, J., Bekkering, H., & van Rooij, I. (2017). To be precise, the details don't matter: On predictive processing, precision, and level of detail of predictions. *Brain and Cognition*, *112*, 84–91. https://doi.org/10.1016/j.bandc.2016.02.008

Laudan, L. (1977). *Progress and its problem: Towards a theory of scientific growth*. Berkeley, CA: University of California Press.

Limanowski, J., & Blankenburg, F. (2015). Network activity underlying the illusory self-attribution of a dummy arm. *Human Brain Mapping*, *36*(6), 2284–2304. https://doi.org/10.1002/hbm.22770

Litwin, P., & Miłkowski, M. (in press). Prospection does not imply predictive processing. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X19002991

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, *24*(4), 279–284. https://doi.org/10.1177/0963721415570732

Lutz, A., Mattout, J., & Pagnoni, G. (2019). The epistemic and pragmatic value of non-action: A predictive coding perspective on meditation. *Current Opinion in Psychology*, *28*, 166–171. https://doi.org/10.1016/j.copsyc.2018.12.019

Miłkowski, M. (2016). Unification strategies in cognitive science. *Studies in Logic, Grammar and Rhetoric*, *48*(1), 13–33. https://doi.org/10.1515/slgr-2016-0053

Miłkowski, M., & Nowakowski, P. (2019). Representational unification in cognitive science: Is embodied cognition a unifying perspective? *Synthese*. Advance online publication. https://doi.org/10.1007/s11229-019-02445-w

Miller, M., & Clark, A. (2018). Happily entangled: Prediction, emotion, and the embodied mind. *Synthese*, *195*(6), 2559–2575. https://doi.org/10.1007/s11229-017-1399-7

Mole, C., & Klein, C. (2010). Confirmation, refutation, and the evidence of fMRI. In S. J. Hanson & M. Bunzl (Eds.), *Foundational issues in human brain mapping* (pp. 99–112). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262014021.003.0010

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA and London: Harvard University Press.

Notredame, C. E., Pins, D., Deneve, S., & Jardri, R. (2014). What visual illusions teach us about schizophrenia. *Frontiers in Integrative Neuroscience*, *8*, 63. https://doi.org/10.3389/fnint.2014.00063

Powers, A. R., Mathys, C., & Corlett, P. R. (2017). Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science*, *357*(6351), 596–600. https://doi.org/10.1126/science.aan3458

Quattrocki, E., & Friston, K. J. (2014). Autism, oxytocin and interoception. *Neuroscience & Biobehavioral Reviews*, *47*, 410–430. https://doi.org/10.1016/j.neubiorev.2014.09.012

de Regt, H. W. (2017). *Understanding scientific understanding*. New York: Oxford University Press.

Ridderinkhof, K. R. (2017). Emotion in action: A predictive processing perspective and theoretical synthesis. *Emotion Review*, *9*(4), 319–325. https://doi.org/10.1177/1754073916661765

Schmack, K., Rothkirch, M., Priller, J., & Sterzer, P. (2017). Enhanced predictive signalling in schizophrenia. *Human Brain Mapping*, *38*(4), 1767–1779. https://doi.org/10.1002/hbm.23480

Schmack, K., Schnack, A., Priller, J., & Sterzer, P. (2015). Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. *Schizophrenia Research: Cognition*, *2*(2), 72–77. https://doi.org/10.1016/j.scog.2015.03.005

Seriès, P., & Seitz, A. (2013). Learning what to expect (in visual perception). *Frontiers in Human Neuroscience*, *7*, 668. https://doi.org/10.3389/fnhum.2013.00668

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, *2*, 395. https://doi.org/10.3389/fpsyg.2011.00395

Sober, E. (1999). Testability. *Proceedings and Addresses of the American Philosophical Association*, *73*(2), 47–76.

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

Stephan, K. E., Binder, E. B., Breakspear, M., Dayan, P., Johnstone, E. C., Meyer-Lindenberg, A., Schnyder, U., Wang, X.-J., Bach, D. R., Fletcher, P. C., Flint, J., Frank, M. J., Heinz, A., Huys, Q. J. M., Montague, P. R., Owen, M. J., & Friston, K. J. (2016). Charting the landscape of priority problems in psychiatry, part 2: Pathogenesis and aetiology. *The Lancet Psychiatry*, *3*(1), 84–90. https://doi.org/10.1016/S2215-0366(15)00360-0

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., & Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological Psychiatry*, *84*(9), 634–643. https://doi.org/10.1016/j.biopsych.2018.05.015

Sterzer, P., Voss, M., Schlagenhauf, F., & Heinz, A. (2019). Decision-making in schizophrenia: A predictive-coding perspective. *NeuroImage*, *190*, 133–143. https://doi.org/10.1016/j.neuroimage.2018.05.074

Teufel, C., Subramaniam, N., Dobler, V., Perez, J., Finnemann, J., Mehta, P. R., Goodyer, I. M., & Fletcher, P. C. (2015). Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proceedings of the National Academy of Sciences*, *112*(43), 13401–13406. https://doi.org/10.1073/pnas.1503916112

Thornton, C. (2017). Predictive processing simplified: The infotropic machine. *Brain and Cognition*, *112*, 13–24. https://doi.org/10.1016/j.bandc.2016.03.004

Tsakiris, M. (2017). The multisensory basis of the self: From body to identity to others. *The Quarterly Journal of Experimental Psychology*, *70*(4), 597–609. https://doi.org/10.1080/17470218.2016.1181768

Valton, V., Karvelis, P., Richards, K. L., Seitz, A. R., Lawrie, S. M., & Seriès, P. (2019). Acquisition of visual priors and induced hallucinations in chronic schizophrenia. *Brain*, *142*(8), 2523–2537. https://doi.org/10.1093/brain/awz171

van Elk, M., & Aleman, A. (2017). Brain mechanisms in religion and spirituality: An integrative predictive processing framework. *Neuroscience & Biobehavioral Reviews*, *73*, 359–378. https://doi.org/10.1016/j.neubiorev.2016.12.031

Veissière, S. P., & Stendel, M. (2018). Hypernatural monitoring: A social rehearsal account of smartphone addiction. *Frontiers in psychology*, *9*, 141.

Votsis, I. (2015). Unification: Not just a thing of beauty. *THEORIA. An International Journal for Theory, History and Foundations of Science*, *30*(1), 97. https://doi.org/10.1387/theoria.12695.

Williams, D. (2018a). Hierarchical Bayesian models of delusion. *Consciousness and Cognition*, *61*, 129–147. https://doi.org/10.1016/j.concog.2018.03.003

Williams, D. (2018b). Predictive coding and thought. *Synthese*, *197*(4), 1749–1775. https://doi.org/10.1007/s11229-018-1768-x

Williams, D. (2018c). Predictive processing and the representation wars. *Minds and Machines*, *28*(1), 141–172. https://doi.org/10.1007/s11023-017-9441-6

Williams, D. (2019). Hierarchical minds and the perception/cognition distinction. *Inquiry*. https://doi.org/10.1080/0020174X.2019.1610045

Zeller, D., Friston, K. J., & Classen, J. (2016). Dynamic causal modeling of touch-evoked potentials in the rubber hand illusion. *NeuroImage*, *138*, 266–273. https://doi.org/10.1016/j.neuroimage.2016.05.065