
Mental distress through the prism of predictive processing

Sander Van de Cruys^{1*}, Pieter van Dessel^{2*}

¹Laboratory of Experimental Psychology, KU Leuven

²Department of Experimental-Clinical and Health Psychology, Ghent University

Abstract

We review the predictive processing theory's take on goals and affect, to shed new light on mental distress and how it develops into psychopathology such as in affective and motivational disorders. This analysis recovers many of the classical factors known to be important in those disorders, like uncertainty and control, but integrates them in a mechanistic model of adaptive and maladaptive cognition and behavior. We derive implications for treatment that have so far remained underexposed in this burgeoning field, specifically with regard to the model-dependent construction of value, the importance of model validation, and the introduction and learning of adaptive predictions that relieve suffering.

Keywords: *predictive processing, mental distress, psychopathology, emotion, depression, anxiety, active inference, learning, psychotherapy, computational psychiatry*

* Corresponding authors: Sander Van De Cruys (Sander.Vandecruys@kuleuven.be), Pieter Van Dessel (Pieter.VanDessel@UGent.be).

Introduction

In recent decades, the influential theory of predictive processing (PP) (also called Bayesian brain or active inference) [1–3] has increasingly gained traction. It states that organisms do not respond to stimuli directly but proactively model and predict sensory inputs and the opportunities and challenges they represent. This view implements pioneering ideas of Jerome Bruner [4] and George Kelly [5] that describe mental processing in terms of the formation and testing of hypotheses, as proto-scientists. As true scientists, we make predictions about the world and meet reality only in our failures [6], that is, when sensory input deviates from our predicted or constructed versions. Such deviations are called prediction errors in PP lingo. Because we perceive through our own constructs (in the form of predictions), perception is often described as ‘controlled hallucination’, which means that we de facto live in a constructed reality [7], kept in check by prediction errors.

The computational details of PP are beyond the current scope [3,8], but the central thesis is that the mental system realizes all its functions through prediction error minimization. Organisms embody an evolutionarily ‘discovered’ set of expected interoceptive (homeostatic) states that are required for continued existence but that they can only realize via action in their environment (e.g. foraging food to increase blood glucose levels). This requires them to build a generative model about the structure of their environment and how to ‘generate’ the expected states through action. This model allows them to not only react on homeostatic disturbances, but also predict and so preempt those. It is shaped by prediction errors, which can be minimized either by acting to bring predicted events in line with the current model (cf. assimilation), or by perception and learning, that is, by changing our models and inferences so they correspond better with incoming evidence (accommodation). In this way, the organism maximizes the evidence for the model it embodies, often called self-evidencing [9]. Importantly, our sense of self can be seen as just another inferred cause in the generative model, as the explanation for the regularities in multimodal perceptual consequences created by one’s own actions, and as a very efficient prediction of (regularities in) one’s future behavior, emotions, and thoughts (e.g. the narrative and social self). From this perspective, we construe and infer not only what is there in the world, but also who we are, what we can do, and what we want (i.e. our goals), as also described in self-perception theories [10].

The PP framework hence entails that a person’s behavior is canalized by the ways in which

s/he construes and anticipates events and its own reactions to these events [5]. Evidently, this holds for mental distress and disorder too. It is clear that people can end up with (sometimes drastically) diverging constructs (or beliefs) and mental models, first because of the vagaries of the environment that shape the models (i.e., our phylogenetic and ontogenetic learning history), second, because the updating of constructs using prediction errors is a fallible job that requires disentangling different types of uncertainty [11**], and third, because the very gathering of evidence is steered by the current constructs. Because of its unique capacity to account for the variety and idiosyncrasy of models (with mental constructs such as beliefs, goals, and desires), we will argue that the PP framework is well-placed to help us understand mental distress and, hopefully, its treatment.

Goals in predictive processing

The overarching principle of prediction error minimization means that the mental system is geared towards *model validation*, maximizing the evidence for one's model or, equivalently, reducing uncertainty relative to this model. At first glance, this conflicts with established reward-based or goal-directed theories [12] that see reward maximization or goal-realization as the core organizing principle for the mind. But, as we explained above, both biological needs and more personal goals are folded into the model. Goals are expected states (desired outcomes) that can be attained ('generated') by our own actions. Some (core) goals are defined at a higher level (longer temporal horizon) and are persistent against obstacles, i.e. these expectations will *not* be readily updated by momentary prediction errors. Evidently, this resistance is crucial for homeostatic expected states, where "updates" (e.g., lowering glucose expectation) would lead to loss of organismic integrity. However, humans construe a myriad of other goals and values in (putative) support of those interoceptive expectations. We infer our goals and values from the history of our choices and actions (e.g., approach and avoidance)[13], as well as from what people like us (parents, peers, and mentors) want. What makes these expectations resistant to updates is the precision they have accumulated through experience. Precision is the technical term in PP for the estimate of expected uncertainty associated with expected sensory states, in this particular case the expected outcomes of actions. In the literature on goal-directed behavior [14], this expected uncertainty about mapping between actions and outcomes is known as feasibility, efficacy or control(ability).

Given that any behavior is driven by expected (interoceptive or exteroceptive) states, habits

in the sense of automatic stimulus-response mapping have no place in PP. However, “habit-like behavior” can be accounted for as actions that are based on strong stimulus-stimulus predictions (one sensory input predicts another proprioceptive and exteroceptive input) on hierarchically lower levels. While these actions were originally driven by higher level goals, through experience, the lower levels expectations involved (sub-goals), accumulated enough precision (reduced uncertainty) so that the behavior is now sustained by ‘shallow models’ [15,16]. In contrast, behavior commonly referred to as “goal-directed behavior” can be seen as behavior that depends on higher level goals and involves lower level expectations that are contextualized by deep hierarchical models. These higher level goals are only activated when there is uncertainty about actions and whether they lead to desired outcomes. For instance, when sensory input activates different alternative actions, deliberation will be more extensive and will involve weighing whether one already has reliable (precise) actions to attain desired outcomes, or whether one needs to gather more information about the environment (including the consequences of one’s actions) through so-called epistemic actions. This deliberation may remain implicit, but will include counterfactual beliefs: “What if I do this, instead of that, how well does that reduce my expected prediction errors (uncertainty)?”. Hence, whether some behavior is goal-directed or habit-like, is a matter of degree, dependent on how shallow or (counterfactually) ‘deep’ the model is that led to the behavior.

This conceptualization of goals casts a new light on clinically relevant cases where verbally reported goals conflict with (possibly maladaptive) actual behavior that is more habit-like but possibly still driven by lower level “goals” or desired outcomes [16]. From this perspective, the key towards effective treatment is to target expected uncertainties, resolving uncertainty with regard to adaptive actions and goals, and introducing uncertainty for maladaptive actions or beliefs through new learning and epistemic actions.

Mental distress in predictive processing

The PP framework has already been applied to several psychopathologies, such as autism spectrum disorder [17], schizophrenia [18], and obsessive-compulsive disorder [19], usually by characterizing these conditions as stemming from a general deficiency in prediction error updating. However, to explain mental distress in affective and motivational disorders with their strong link to life events, we first need to look at the way PP understands emotions.

Dominant views of affect cast it as a continuous “neurophysiological barometer of the

individual's relationship to an environment at a given point in time" [20]. Similarly, Frijda [21] noted: "pleasure is the positive outcome of constantly monitoring one's functioning". Translated to PP, this kind of meta-monitoring of one's own performance in predicting internal and external disturbances, needs to be an operation on prediction errors. However, given that prediction error minimization is really all the brain does (when perceiving, acting, and thinking), any momentary prediction error will not be very informative or 'affective'. Instead, meta-expectations about the temporal average of prediction errors and how that changes —whether uncertainty increases or decreases relative to one's models— might serve as this meta-monitoring function that we associate with emotional relevance [22–24]. This aligns with the well-known role of increasing uncertainty in negative affect such as anxiety [25], and of decreasing uncertainty in pleasure, such as aesthetic appreciation [26].

This sets the stage for an understanding of distress as rooted in persistent breaches in predictive models, especially concerning high-level, high precision expected states (goals), because those predictions structure major parts of one's behavior, and are the basis for interpretation of much of the perceptual inputs the person encounters (past, present, and future). Persistent and increasing prediction errors signal the subversion of a person's constructed identity: the regularities from the world that they have absorbed but that no longer hold. It is here that we situate normal and pathological distress, be it grief, depression, or anxiety.

Usually, prediction errors (even with regard to core goals) can be neutralized by changing one's auxiliary assumptions, and so keeping the core of a person's models intact (sometimes called cognitive immunization). That is just run-of-the-mill model updating which will take place first for lower-level, less precise predictions, rather than for higher-level, precise predictions. However, persistent model disconfirmations can incite a number of different responses to reduce uncertainty (increase model evidence). A first approach might be to tighten up on one's constructs, and insist on their validity. This is often expressed as a more selective data gathering to file off the sharp edges of disconfirmations, in other words: avoiding particular (e.g., social) situations and retreat to more reliably predictable environments, as we often see in the rituals of anxious people. Another outcome could be hostility, to be seen as extreme efforts to force the circumstances to confirm to one's constructs again. Whether avoidant action or aggressive action is taken may depend on

perceived control, a well-known factor in emotion and psychopathology, which, in PP, can be conceptualized as uncertainty about (our prediction of) action consequences [27].

Crucially, persistent prediction errors in reaching wanted action outcomes can lead to model updates that increase the (meta-)expected uncertainty or, simply put, the perceived powerlessness of actions. The generalized loss of confidence (certainty) that one can attain the expected consequences of one's own actions (cf. learned helplessness) can then lead to withdrawal and other depressive symptoms. This resonates with the fact that unpredictable, unresponsive environments, especially early on in development, are a key risk factor for (depressive) psychopathology [28].

The 'system-wide' generalization of loss of confidence in the efficacy of all actions (as in depression) might seem unwarranted but it may actually be internally-rational, given one's models and the experiences that shaped them. Usually, the reverse generalization, a certain generalized overconfidence (known as the optimism bias or the overestimation of control) keeps us motivated and is often seen as a mark of good mental health [29]. However, a generalized *loss* of confidence in actions shifts one's mental model from trying to fit the world to their mental model (by actions) to fitting the mental model to the world and may therefore be a successful way to more easily achieve prediction error minimization. This might relate to depressive rumination as well as depressive realism, the finding that people with depression often have a more accurate perception of their capacities and control.

Finally, another effective strategy to deal with persistent prediction errors is to formulate predictions on a more global, less granular level [30]. Simply stated, categorizing something as merely 'fruit' or 'vegetable', rather than at a more specific level, allows one to be correct more often. Applied to psychopathologies such as anxiety disorder, over-general threat-related priors (e.g., "social situations are threatening") may start to dominate thought and behavior because they 'worked' (i.e., they were effective in reducing errors) and so become insensitive to new evidence [31*].

Toward more effective treatments

Current psychotherapeutic practice (of motivational or affective disorders) predominantly involves trying to uncover and modify a client's maladaptive beliefs (CBT). The PP account fits well in this tradition [32**] because it provides a mental process theory about how these beliefs are formed and how they can be changed (using scaffolded learning environments).

Correspondingly, most papers that cast psychopathology in PP terms emphasize belief updating [33*] and, because updating is thought to be mediated by neuromodulators, express the hope for better pharmacological treatments in the future. However, PP underscores at least three additional elements for psychotherapy.

Firstly, clinical practice should give attention to specifying the generative model of the client, to determine what the client construes as valuable (or ‘expected’) instead of relying on objective definitions or normative ideas of what is rewarding or unpleasurable. Failing to do so may make us blind to the attractiveness of “what should objectively be” aversive situations, as well as the avoidance of “what should objectively be” neutral or even positive situations. For instance, harmless situations are in our predictive mind sometimes construed as (near) catastrophes precluding the updating of maladaptive avoidance behavior in exposure therapy [32**].

Secondly, ‘model validation’ will be an important, recurrent step in psychotherapy. This means providing support (in dry PP terms: model evidence) for core constructs such as adaptive beliefs, values and goals (cf. acceptance and commitment therapy). Earlier, we alluded to the ‘protective belt’ that models build for themselves by explaining away prediction errors using auxiliary hypotheses or by downregulating the salience of conflicting evidence by lowering trust (expected confidence or certainty) in particular information sources. To alter these models, it is important to cultivate (renewed) trust in suppressed information sources and in one’s own actions, evoking a generalized sense of initiative beyond the therapeutic setting. In PP terms, this constitutes regaining confidence that one’s actions will minimize prediction errors (that one’s planning works, that future can be produced), in layman’s terms, it is about consolation.

Thirdly, the model validation provides a safe basis for the (sometimes threatening) introduction of new, more adaptive constructs and inferences in mental models. Eliciting a change in generative models requires us to set-up the environment to evoke prediction error and to aid accommodation of these errors in an adaptive manner. From this perspective, therapists should create learning environments that allow clients to confidently predict that they will perform alternative behavior (e.g. drinking water in case of alcohol addiction) in relevant contexts (e.g. after dinner) and infer that this alternative behavior aligns better with their goals [34].

Conclusion

The emerging description of mental distress in PP terms dovetails with insights from several psychotherapeutic traditions, as well as with dimensional and transdiagnostic views in psychopathology. The PP approach proposes plausible (inferential) mechanisms not only for pharmacological intervention, but also for the science of psychotherapy, showing that the ‘soft skills’ of psychotherapy lend themselves to a ‘dry’ Bayesian PP analysis. It can do so because it is decidedly not biologically reductionist, through its emphasis on how the environment shapes inferences and constructed values [35]. Much more than a PP gloss on old ideas, the framework continues to show great heuristic power for psychopathology.

Conflict of interest statement

Nothing declared.

Author note

This work is supported by the Flemish Government under Grant METH/14/02 awarded to Johan Wagemans and by the Scientific Research Foundation Flanders under Grant FWO19/PDS/041 awarded to Pieter van Dessel.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest, ** of outstanding interest.

- [1] Clark A: **Whatever next? Predictive brains, situated agents, and the future of cognitive science.** *Behav Brain Sci* 2013, **36**:181–204. <https://doi.org/10.1017/S0140525X12000477>.
- [2] Hohwy J (Ed): *The Predictive Mind*. Oxford University Press; 2013.
<http://books.google.be/books?id=3m8nAgAAQBAJ>
- [3] Friston K: **The free-energy principle: a unified brain theory?** *Nat Rev Neurosci* 2010, **11**:127–138. <https://doi.org/10.1038/nrn2787>
- [4] Bruner JS: **On perceptual readiness.** *Psychol Rev* 1957, **64**:123. <http://psycnet.apa.org/journals/rev/64/2/123>
- [5] Kelly GA: **A brief introduction to personal construct theory.** In *Perspectives in Personal Construct Theory*. Edited by Bannister D. London Academic Press; 1970: 29.
<https://www.aippc.it/wp-content/uploads/2019/04/2017.01.003.025.pdf>

- [6] von Glasersfeld E (Ed): *Radical Constructivism: A Way of Knowing and Learning. Studies in Mathematics Education Series: 6*, Falmer Press; 1995. <http://eric.ed.gov/?id=ED381352>
- [7] Hoffman DD, Singh M, Prakash C: **The Interface Theory of Perception**. *Psychon Bull Rev* 2015, **22**: 1480–1506. <https://doi.org/10.3758/s13423-015-0890-8>
- [8] Wiese W, Metzinger T: **Vanilla PP for Philosophers: A Primer on Predictive Processing**. In: *Philosophy and Predictive Processing*. Edited by Metzinger T, Wiese W. Philpapers.org; 2017. <https://philpapers.org/rec/WIEVPF>
- [9] Hohwy J: **The Self-Evidencing Brain**. *Noûs* 2014, **50**: 259–285. <https://doi.org/10.1111/nous.12062>
- [10] Bem DJ: **Self-perception theory**. *Adv Exp Soc Psychol* 1972, **6**:1–62. <http://healthyinfluence.com/wordpress/wp-content/uploads/2011/05/SP-Theory-Bem-Advances.pdf>
- [11**] Pulcu E, Browning M: **The Misestimation of Uncertainty in Affective Disorders**. *Trends Cogn Sci* 2019, **23**:865–875. <https://doi.org/10.1016/j.tics.2019.07.007>
The authors review evidence that estimation of uncertainty plays an important role in learning and explain how misestimation of uncertainty can lead to symptoms of anxiety and depression and might underlie affective disorders.
- [12] Moors A, Boddez Y, De Houwer J: **The Power of Goal-Directed Processes in the Causation of Emotional and Other Actions**. *Emot Rev* 2017, **9**: 310–318. <https://doi.org/10.1177/1754073916669595>
- [13] Van Dessel P, Hughes S, De Houwer J: **How Do Actions Influence Attitudes? An Inferential Account of the Impact of Action Performance on Stimulus Evaluation**. *Pers Soc Psychol Rev* 2019, **23**: 267–284. <https://doi.org/10.1177/1088868318795730>
- [14] Austin JT, Vancouver JB: **Goal constructs in psychology: Structure, process, and content**. *Psychol Bull* 1996, **120**:338. <https://psycnet.apa.org/doiLanding?doi=10.1037/0033-2909.120.3.338>
- [15] Pezzulo G, Rigoli F, Friston K: **Active Inference, homeostatic regulation and adaptive behavioural control**. *Prog Neurobiol* 2015, **134**:17–35. <https://doi.org/10.1016/j.pneurobio.2015.09.001>
- [16] FitzGerald THB, Dolan RJ, Friston KJ: **Model averaging, optimal inference, and habit formation**. *Front Hum Neurosci* 2014, **8**:457. <https://doi.org/10.3389/fnhum.2014.00457>
- [17] Van de Cruys S, Evers K, Van der Hallen R, Van Eylen L, Boets B, de-Wit L, Wagemans J: **Precise minds in uncertain worlds: predictive coding in autism**. *Psychol Rev* 2014, **121**:649–675. <https://doi.org/10.1037/a0037665>
- [18] Sterzer P, Adams RA, Fletcher P, Frith C, Lawrie SM, Muckli L, Petrovic P, Uhlhaas P, Voss M, Corlett PR: **The Predictive Coding Account of Psychosis**. *Biol Psychiatry* 2018, **84**:634–643. <https://doi.org/10.1016/j.biopsych.2018.05.015>
- [19] Fradkin I, Adams RA, Parr T, Roiser JP, Huppert JD: **Searching for an anchor in an unpredictable world: A computational model of obsessive compulsive disorder**. *Psychol*

- Rev 2020, 127:672–699. <https://doi.org/10.1037/rev0000188>
- [20] Duncan S, Barrett LF: **Affect is a form of cognition: A neurobiological analysis.** *Cogn Emot* 2007, 21:1184–1211. <https://doi.org/10.1080/02699930701437931>
- [21] Frijda NH (Ed), *The laws of emotion.* Psychology Press; 2007.
- [22] Hesp C, Smith R, Allen M, Friston K, Ramstead M: **Deeply felt affect: the emergence of valence in deep active inference.** PsyArxiv 2019. <https://psyarxiv.com/62pfd/download?format=pdf>.
- [23] Van de Cruys S: **Affective Value in the Predictive Mind.** In *Philosophy and Predictive Processing.* Edited by Metzinger TK, Wiese W. MIND Group; 2017. <https://doi.org/10.15502/9783958573253>
- [24] Joffily M, Coricelli G: **Emotional Valence and the Free-Energy Principle.** *PLoS Comput Biol* 2013; 9:e1003094. <https://doi.org/10.1371/journal.pcbi.1003094>.
- [25] Whalen PJ: The uncertainty of it all. *Trends Cogn Sci* 2007, 11:499–500. <https://doi.org/10.1016/j.tics.2007.08.016>
- [26] Muth C, Raab MH, Carbon C-C: **The stream of experience when watching artistic movies. Dynamic aesthetic effects revealed by the Continuous Evaluation Procedure (CEP).** *Front Psychol* 2015, 6:365. <https://doi.org/10.3389/fpsyg.2015.00365>
- [27] Huys QJM, Dayan P: **A Bayesian formulation of behavioral control.** *Cognition* 2009, 113:314–328. <https://doi.org/10.1016/j.cognition.2009.01.008>
- [28] Ellis BJ, Bianchi J, Griskevicius V, Frankenhuis WE: **Beyond Risk and Protective Factors: An Adaptation-Based Approach to Resilience.** *Perspect Psychol Sci* 2017, 12:561–587. <https://doi.org/10.1177/1745691617693054>
- [29] S.E. Taylor, J.D. Brown, Illusion and well-being: a social psychological perspective on mental health, *Psychol. Bull.* 103 (1988) 193–210. <https://www.ncbi.nlm.nih.gov/pubmed/3283814>.
- [30] Kwisthout J, Bekkering H, van Rooij I: **To be precise, the details don't matter: On predictive processing, precision, and level of detail of predictions.** *Brain Cogn* 2017, 112:84–91. <https://doi.org/10.1016/j.bandc.2016.02.008>
- [31*] Van den Bergh O, Brosschot J, Critchley H, Thayer JF, Ottaviani C: **Better Safe Than Sorry: A Common Signature of General Vulnerability for Psychopathology.** *Perspect Psychol Sci* 2020. <https://doi.org/10.1177/1745691620950690>
The authors explain how a person's general strategy to process prediction error in low detail can give rise to weakly informed prior beliefs that might underlie (affective) mental disorders.
- [32**] Moutoussis M, Shahar N, Hauser TU, Dolan RJ: **Computation in Psychotherapy, or How Computational Psychiatry Can Aid Learning-Based Psychological Therapies.** *Comput Psychiatr* 2018, 2:50–73. https://doi.org/10.1162/CPSY_a_00014.
The authors describe how recent advances in PP and computational modeling can help formalize and test hypotheses regarding the maladaptive inferences that underlie mental disorder and can improve learning-based therapies such as CBT.
- [33*] Kube T, Rozenkrantz L: **When Beliefs Face Reality: An Integrative Review of Belief**

Updating in Mental Health and Illness. *Perspect Psychol Sci* 2020.

<https://doi.org/10.1177/1745691620931496>.

The authors review evidence about how healthy people and people with mental disorders update their beliefs after receiving new information that supports or challenges their views. They argue that the strategy to afford very high or very low precision to prior beliefs may underlie psychopathology.

- [34] Wiers RW, Van Dessel P, Köpetz C: **ABC Training: A New Theory-Based Form of Cognitive-Bias Modification to Foster Automatization of Alternative Choices in the Treatment of Addiction and Related Disorders.** *Curr Dir Psychol Sci* 2020, **29**:499–505.
<https://journals.sagepub.com/doi/abs/10.1177/0963721420949500>.

- [35] Adams RA, Huys QJM, Roiser JP: **Computational Psychiatry: towards a mathematically informed understanding of mental illness.** *J Neurol Neurosurg Psychiatry* 2016, **87**:53–63.
<https://doi.org/10.1136/jnnp-2015-310737>.