



Commentary: On the joys of perceiving: Affect as feedback for perceptual predictions

Sabrina Trapp^{1,2*}

¹ Department of Psychology, Ludwig-Maximilians-University, Munich, Germany, ² Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat-Gan, Israel

Keywords: predictive coding, exploration-exploitation, novelty, affect, choice, uncertainty, reinforcement (psychology)

A commentary on

On the joys of perceiving: Affect as feedback for perceptual predictions

by Chetverikov, A., and Kristjánsson, Á. (2016). *Acta Psychol.* 169, 1–10. doi: 10.1016/j.actpsy.2016.05.005

OPEN ACCESS

Edited by:

Antonio Damasio,
University of Southern California,
United States

Reviewed by:

John Monterosso,
University of Southern California,
United States

*Correspondence:

Sabrina Trapp
sabrina.trapp@biu.ac.il

Specialty section:

This article was submitted to
Decision Neuroscience,
a section of the journal
Frontiers in Neuroscience

Received: 25 July 2017

Accepted: 22 September 2017

Published: 23 October 2017

Citation:

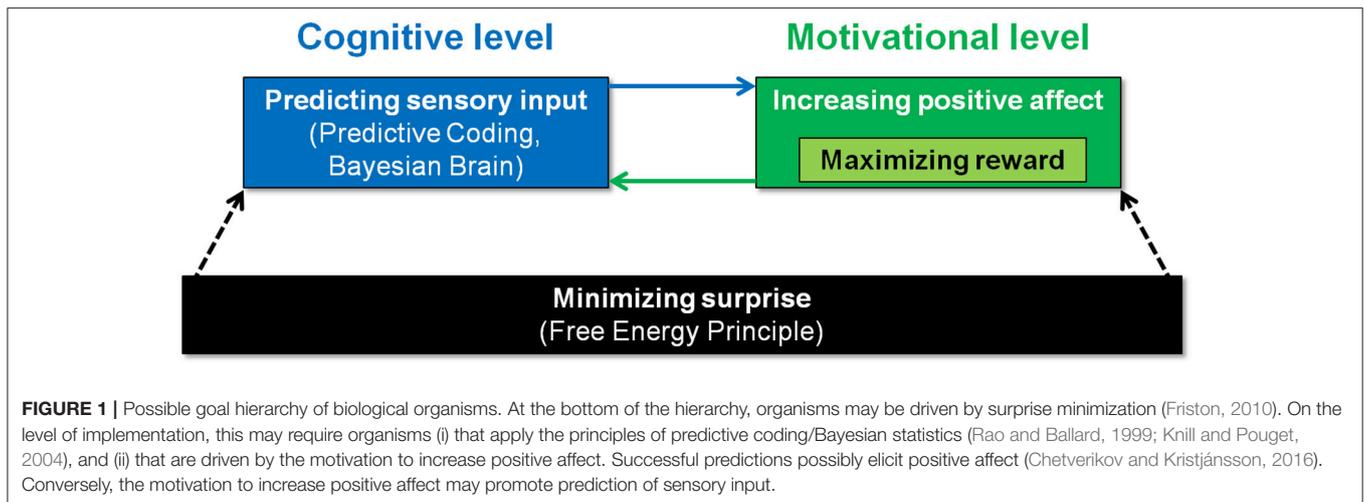
Trapp S (2017) Commentary: On the joys of perceiving: Affect as feedback for perceptual predictions. *Front. Neurosci.* 11:556. doi: 10.3389/fnins.2017.00556

The intuition that perception relies on prior information when inferring the causes of sensory input has received strong theoretical and empirical support (for a review see Clark, 2013). In the framework of *predictive coding*, it is assumed that cortico-cortical feedback connections provide predictions about sensory input, and only the residual errors (prediction errors) are fed forward in the visual hierarchy to be further processed (Rao and Ballard, 1999; Lee and Mumford, 2003). It has even been suggested that conceptualizing the brain as minimizing surprise can account for several neurophysiological and neuroanatomical observations (Friston, 2005, 2010). This raises the question of why organisms are not attracted to sensory vacuums where the prediction error is zero (referred to as the ‘dark room problem’). Friston et al. (2012) argued that organisms harbor models of the environment, in which such a scenario does not exist. Changes do occur in natural environments, and consequently, the cognitive systems of organisms do expect occasional prediction errors. Leaving a sensory vacuum, such as a dark room, may translate into an attempt to seek an environment that mirrors the degree of uncertainty the organisms’ mind expects. However, these frameworks are formulated on an abstract, mathematical level, leaving open the question of what motivates the organism at a *psychological* level of explanation.

A framework that has the potential to address this issue was suggested by Chetverikov and Kristjánsson (2016). Their core assumption appeals to an important psychological concept, namely affect. It is suggested that successful prediction elicits positive affect¹. This way, affect can foster increasingly accurate predictions. Importantly, it is conjectured that affective feedback is weighted with the inverse prior probabilities of events. In other words, highly predictable information tends to elicit no positive affect. Consequently, leaving a dark room would be required in order to increase positive affect.

I conjecture that this explanatory route has something to add to the field of reinforcement learning (RL). While previous proposals have emphasized the commonalities between the process of perceptual inference and RL (Rushworth et al., 2010), the exploration-exploitation dilemma in the latter has remained untouched by the connection.

¹The idea has already been formulated by Ramachandran and Hirstein (1999): The authors suggested that perception must be linked to limbic brain systems in order to provide incentive for discovering image correlations, binding, grouping, and identifying content.



The basic tenet in RL is that organisms strive to maximize their rewards. To this end, they capitalize upon efficient learning systems that attach values to cues in the environment which lead to good decisions in the future (Sutton and Barto, 1998). However, this conjecture introduces a conflict between exploitation and exploration. Given cues that promise rewards, why should organisms explore novel information for which there is no reward history? This dilemma in RL has fostered the development of a plethora of algorithms that address this issue. For example, the *shaping bonus* suggests initializing novel information with a higher value (Kakade and Dayan, 2002). More recently, it has been suggested that novel information receives high values via generalization of known stimuli in the same environment (Gershman and Niv, 2015). However, whilst these suggestions can explain that organisms choose novel stimuli rather than ignore them altogether, they are mute on the question why organisms occasionally even prefer novel to familiar stimuli². In addition, there is a puzzling effect of context. It seems that laboratory rats tend to be more neophilic, e.g., preferring novelty even to cocaine (Reichel and Bevins, 2008)³, whereas rats in the wild show more neophobic behavior (Barnett, 1958). As of yet, there is no overarching theory that can account both for a preference for novel to familiar stimuli and such contextual modulations.

The exploration-exploitation dilemma bears structural similarity to the dark room problem—how does novelty come into play when the organism is supposed to be driven by successful predictions? Possibly, this problem cannot be

accounted for by explanations rooted in RL, but requires the reference to affect and the domain of perception. If feedback for predictions is weighted with the inverse prior probabilities of events, then laboratory rats would experience no positive affect because of the high likelihood of surrounding stimuli, and consequently, seek out for novelty.

The idea aligns with data from functional imaging: Several studies have shown that the ventral striatum, a key structure in reward coding, is also activated by mere novelty of stimulus material (e.g., Bunzeck et al., 2010). These findings suggest that novelty may be intrinsically rewarding (Wittmann et al., 2008).

Figure 1 illustrates how surprise minimization, predictive coding, affect, and reward seeking could be orchestrated. At the highest level of abstraction, one might conceptualize organisms as driven by surprise minimization, as captured in the *Free Energy Principle* (FEP) (Friston, 2005). At the lowest level of processes, i.e., perception, the FEP posits that organisms actively construct hypotheses or predictions about sensory input. Thus, the Bayesian approach logically follows from the assumption that organisms must minimize their surprise via the proxy of the free-energy. This means the “Bayesian brain” hypothesis rests on the free-energy principle, and adds a functional level of explanation, i.e., what makes the system operate in a Bayesian manner (Friston, 2010). Where does affect may come into play? Biological organisms can be assumed to strive for positive, and avoidance of negative affect (Panksepp, 2008). This motivational force may ultimately serve surprise minimization, if one assumes that positive affect can be achieved by successfully using prior information (Chetverikov and Kristjánsson, 2016).

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

²There are a few algorithms that try to address how agents make choices preferably to uncertain stimuli, e.g., by Szita and Lőrincz (2008).

³In the predictable laboratory environment, the neophilic rats may *like* (have positive affect for) a novel stimulus enough to override their *wanting* of the highly predictable cocaine stimulus.

REFERENCES

- Barnett, S. (1958). Experiments on “neophobia” in wild and laboratory rats. *Br. J. Psychol.* 49, 195–201. doi: 10.1111/j.2044-8295.1958.tb00657.x
- Bunzeck, N., Dayan, P., Dolan, R. J., and Duzel, E. A. (2010). A common mechanism for adaptive scaling of reward and novelty. *Hum. Brain Mapp.* 31, 1380–1394. doi: 10.1002/hbm.20939
- Chetverikov, A., and Kristjánsson, Á. (2016). On the joys of perceiving: affect as feedback for perceptual predictions. *Acta Psychol.* 169, 1–10. doi: 10.1016/j.actpsy.2016.05.005
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Friston, K., Thornton, C., and Clark, A. (2012). Free-energy minimization and the dark-room problem. *Front. Psychol.* 3:130. doi: 10.3389/fpsyg.2012.00130
- Gershman, S. J., and Niv, Y. (2015). Novelty and inductive generalization in human reinforcement learning. *Top. Cogn. Sci.* 7, 391–415. doi: 10.1111/tops.12138
- Kakade, S., and Dayan, P. (2002). Dopamine: generalization and bonuses. *Neural Netw.* 15, 549–559. doi: 10.1016/S0893-6080(02)00048-5
- Knill, D. C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719. doi: 10.1016/j.tins.2004.10.007
- Lee, T. S., and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A.* 20, 1434–1448. doi: 10.1364/JOSAA.20.001434
- Panksepp, J. (2008). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York, NY: Oxford University Press.
- Ramachandran, V. S., and Hirstein, W. (1999). The science of art: A neurological theory of aesthetic experience. *J. Conscious. Stud.* 6, 15–51.
- Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580
- Reichel, C. M., and Bevens, R. A. (2008). Competition between the conditioned rewarding effects of cocaine and novelty. *Behav. Neurosci.* 122, 140–150. doi: 10.1037/0735-7044.122.1.140
- Rushworth, M. F., Mars, R. B., and Summerfield, C. (2010). General mechanisms for making decisions? *Curr. Opin. Neurobiol.* 19, 75–83. doi: 10.1016/j.conb.2009.02.005
- Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press. doi: 10.1109/TNN.1998.712192
- Szita, I., and Lőrincz, A. (2008). “The many faces of optimism: a unifying approach,” in *Proceedings of the 25th International Conference on Machine Learning (Helsinki)*, 1048–1055. doi: 10.1145/1390156.1390288
- Wittmann, B. C., Daw, N. D., Seymour, B., and Dolan, R. J. (2008). Striatal activity underlies novelty-based choice in humans. *Neuron* 58, 967–973. doi: 10.1016/j.neuron.2008.04.027

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer JM and handling Editor declared their shared affiliation.

Copyright © 2017 Trapp. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.