Predictive Processing and the Possibility of Evolutionary Differences in Precision-Weighting

Aleksandar Bahat

December 19, 2017

1. Introduction to the predictive brain hypothesis

1.1. A new theory of perception

A traditional explanation of perception might go something like this: Sensory organs collect raw information about the world. That information then gets processed by the brain, step by step, to create increasingly sophisticated internal representations of the world. This processing relies on prior knowledge of the world to make sense of the input, and attention can modify the details of how the input is handled; nevertheless, it is clear that the direction of information flow in this picture is fundamentally "bottom-up" (Clark, 2015). In short, we sense, then compute, then perceive.

As a prototypical example of this model, consider David Marr's theory of vision. According to Marr's theory, the process of visual perception begins with the eyes detecting light intensities throughout the visual field. Those light intensity data then get processed into basic geometric features (blobs, edges, lines, bars, and the like), which then get processed into surface orientations, which eventually get processed into a full three-dimensional representation of what is seen (Marr, 1982). The traditional view, then, portrays perception as fundamentally *data-driven* (more complex representations come from progressively synthesizing great quantities of unrefined sensory data) and *passive* (the brain is "on hold" until it receives external input to process).

The "predictive processing" (**PP**) approach to perception ("predictive coding," "the predictive brain hypothesis," "the Bayesian brain hypothesis," and "the free-energy principle" all refer to closely related, if not quite identical, ideas) inverts this thinking. In the PP framework, the

1

brain does not wait passively to receive input for processing; rather, it constantly tries to *actively predict* incoming input ahead of time. Furthermore, this process is hierarchically organized: each level of the system attempts to predict the input of the level below it, all the way down to the level of the senses. In addition to the top-down flow of predictions, PP also incorporates a bottom-up flow of *prediction error*, which serves to update the predictions to more accurately represent the sensory data (and therefore the external world) (Clark, 2012). Thus, when it comes to perception, sensory data are no longer valuable "for their own sake" (as the building blocks from which the final perceptive model is built), but only as a prompt to change inaccurate predictions.

To summarize, the PP picture is one of a highly layered cognitive architecture, in which each level is constantly performing two distinct types of information-processing tasks: 1) making predictions about the incoming future information at lower levels and 2) reporting errors in predictions made at higher levels. In contrast to the traditional theory of perception considered earlier, PP is *active* rather than passive, *data-guided* rather than data-driven. The idea is neatly summed up in the phrase "perception is controlled hallucination," attributed separately to Max Clowes, Rodolfo Llinas, and Ramesh Jain (Clark, 2014, p. 236).

Let us consider a concrete example of how one might perceive something according to PP. Suppose I am driving a car. The "first" step in perceiving something is the formation of a high-level prediction: "I will pass the car in the lane to my right in a second." (Of course, there is no actual "first" step, as perception, and thus the feedback between prediction and error, is a continuous, "rolling" process.) This prediction then implies (or rather, predicts) a new, lower-level prediction: "The large, vaguely rectangular, glossy object directly to my right, moving in the same direction as me, will be to my right and behind me in a second." This prediction, in turn, yields even lower-level predictions: "The collection of dark pixels (the car) on the right side of my field of vision will entirely be replaced by lighter pixels (the background) in a second." Meanwhile, the sensory data act as a

check on every level: Do the lowest-level predictions hold? That is, do the dark pixels get replaced by light pixels in the correct way? If so, all is well, the predictions need no modification, and the product of all this is: "I see myself passing the car." If not, then we must send the resulting error signals "up" to the next level, so that the prediction can be updated; but if that prediction changes, so must the one above it; and so on, until the predicted sensory input (the bottom level, derived from all the levels above it) matches the actual sensory input from the environment. When all the error signals have "gone quiet," and the system reaches a equilibrium of sorts (only in a loose sense; remember, perception is continuous, so this "equilibrium" is relative to that particular moment in time; it is fleeting and dynamic, because the environment is), we have again reached a state that can be associated with a perception—but this one is a bit different than the original, perhaps "I see myself passing the car, but not for another five seconds, because the car sped up unexpectedly."

1.2. Action as prediction-error minimization

In addition to offering an explanation of perception, PP also has a surprising and elegant explanation for *action* (motor control). Action, like perception, just serves to minimize prediction error. That is, if one *expects* the sensory input that results from a particular action, then one way of fulfilling that expectation is by *actually performing the movement.* This theory of "active inference" makes perception and action two sides of the same coin—both are part of the same predictive machine which tries to anticipate sensory input, whether visual, proprioceptive, or otherwise (Adams, Shipp, & Friston, 2013, p. 4). So if I *expect* to reach for the pen on my desk, two ways I can reduce the error signals from sensory input are to 1) expect something else or 2) actually reach for it, with the resulting cascade of predictions at all levels (with the corresponding levels of detail more or less: *gross movement pattern* at the top, *location of limbs in space* in the middle, and *precise sequence of muscular contractions* at the bottom). Note that this is not a conscious "decision" to be made; whether I reach for the pen or not depends on the (complex, non-linear) dynamics of how the

predictions are updated to account for the error signals. Some patterns of updating could lead to a loss of interest in the pen, while others could lead to actual motor output.

1.3. Evidence for and against a predictive framework

There are several different kinds of evidence in favor of a PP-like model of perception (and maybe action too). Some is "anecdotal," consisting of individual peculiarities of perception which are consistent with a predictive brain hypothesis. Many optical illusions can be (or even must be) perceived differently after the observer is given some cue or prior knowledge (a simple example: a flat trapezoid can also appear as a rectangle tilted away from the viewer). Likewise, the perception of distorted sensory input, like "sine-wave" speech or black-and-white versions of gray-scale images, can also vary dramatically depending on prior knowledge, that is, the predictive model (Clark, 2014, pp. 228-29). The "rubber hand illusion" also provides a startling demonstration of the importance of expectations on perception (Apps & Tsakiris, 2014). Finally, the phenomenon of "binocular rivalry" (when staring at a different image with each eye, one's vision oscillates back and forth between the two images, rather than settling on something in between) also has an attractive explanation in PP terms: the brain predicts it sees one of the images, but high error persists, so it switches the prediction to the other, but high error persists, and so on, never able to find a stable equilibrium (Clark, 2013c).

Taking a more theoretical view, PP still has much to offer. First of all, it must answer the same big question as any other theory of cognition: does it make evolutionary sense? In this case, the answer seems to be yes, at least at first glance. Organisms with a more efficient means of information processing and decision making (efficient both in a temporal and metabolic sense) will certainly have a selective advantage. From a computational perspective, prediction is quite effective, and importantly, computationally cheap—in fact, predictive coding has roots in techniques for visual and auditory data compression, developed at Bell Labs during the 1950s (Clark, 2013c). Additionally, "Bayesian" brain models (where the brain's hypotheses about the world are revised in

light of experience in a manner that at least approximates the statistician's ubiquitous *Bayes' theorem*) like PP are attractive for their ability to explain how brains learn so well from little evidence (Horgan, 2016). As Andy Clark explained, "The beauty of the bidirectional hierarchical structure is that it allows the system to infer its own priors (the prior beliefs essential to the guessing routines) as it goes along. It does this by using its best current model—at one level—as the source of the priors for the level below" (Clark, 2013c, p. 183).

Furthermore, it may be that PP is not only an adaptive way for organisms to process and respond to their environment, but the result of a more general principle of self-organizing complex systems, what Karl Friston calls the *free-energy principle*: "any self-organizing system that is at equilibrium with its environment must minimize its free energy" (Friston, 2010). Free energy for a brain, then, would just be what one would call *surprisal* or *prediction error* in PP. In summary, PP is a highly explanatory (there are yet more concepts to be explained by it in the next section!) and highly flexible theory that is consistent with and can explain many of the less intuitive features of human perception and action.

That being said, an immediate concern is that PP is *too* explanatory and flexible—is it even a proper, non-trivial scientific theory? Some think it is too broad to be falsifiable; others think it is too narrow, working too close to the implementational level (Clark, 2013a). Another common objection is the problem of novelty-seeking, or, phrased more dramatically: if the brain is trying to minimize prediction error, why not sit in a dark room and do nothing? The usual answer, which seems at least superficially satisfactory, is that animals *already* predict that they will move, search for food, acquire resources, and so on, so a deviation from these built-in priors (shaped by natural selection), like sitting in a dark room, would actually produce error, not minimize it (Clark, 2013c).

2. The idea of precision-weighting and some of its consequences

2.1. Flexible precision as a means of handling uncertainty

One important mechanism within PP theory that we have not yet mentioned is that of *precision-weighting*. Sensory data about the world is inherently statistically noisy, and it would be nonsensical to expect prediction error to ever be truly zero (for the simple reason that measurements of real quantities necessarily have finite precision). Thus, PP needs a way to decide what level of non-zero error is "good enough" to still stick with a prediction. A key feature of the PP model as it is usually described is that this error tolerance is *dynamic* and *context-dependent*: in other words, the system can change the relative importance ascribed to the predictions and the prediction errors at each level, so as to optimize the resulting predictive model (Clark, 2014). (It does no good to overturn a prediction that has worked thousands of times in the past based on one questionable error; by flexibly determining when to modify predictions, the entire system can converge on accurate representations of the world more rapidly.) For example, while driving along an unfamiliar, winding road, one expects that the system will turn the "volume" up for prediction error, as the unpredictability of the scene will render any internal prediction in need of frequent updating. On the other hand, navigating through a familiar room in the dark would be facilitated by decreasing sensitivity to prediction error—in a dark room, nominal errors are much more likely to be mere sensory noise, mistakes rather than evidence against a prediction.

2.2. Explaining higher-level cognition with precision-weighting

Incorporating dynamic precision into PP opens up the possibility of explaining a wide variety of cognitive tasks besides perception and action. The most obvious is *attention*: to pay attention to something now means to raise the precision applied to the perception of that thing and lower it elsewhere (that is, predictions about the object of attention are very error-sensitive, while predictions

about other things are less error-sensitive) (Clark, 2013c). Additionally, higher-level cognition like imagination, planning, abstract reasoning, and "theory of mind" can also be accounted for by PP with dynamic precision, as "off-line simulations" of a normal perception-action cycle. The way to take the simulation off-line, of course, is just to turn the relevant prediction errors "off" altogether (Clark, 2013b).

2.3. The effects of precision malfunctions

With precision-weighting, PP might be able to describe not only the above normal features of cognition, but also some of its occasional systematic abnormalities. In particular, consider first the case where prediction error is systematically undervalued (the precision is generally lowered compared to a typical brain). The top-down predictions will not be kept in check by prediction error to the degree that they are in a normally-functioning brain, and subsequent predictions will get worse (more inaccurate) without this feedback, until the error, perhaps quite large now, finally registers. The updated prediction will have little evidence to draw on (remember, all the input that did not cause an error signal was simply ignored by all higher levels of the system), so the new model will likely be another inaccurate representation of the world, triggering another large error signal, and so on. Thus the effect of insufficient error sensitivity leads to the predictions becoming severely decoupled with reality, the "controlled hallucination" of perception losing control; in other words, a condition strongly reminiscent of schizophrenia (Horga, Schatz, Abi-Dargham, & Peterson, 2014). Some experimental results also point towards such a conclusion: for example, when experiencing binocular rivalry, the "switching rate" (rate at which perception switches from one image to the other) for schizophrenics is roughly half that of non-schizophrenics, indicating that the threshold at which the error signal becomes intolerable is higher in the former (Heslop, 2012).

On the other hand, what if the prediction errors are systematically overvalued (the precision is generally raised compared to a typical brain)? In this case, even slight errors, extraneous details

that would normally be ignored, cause a reassessment of the predictive model. This inability to settle on an "explanation" of the sensory data makes it difficult to process the overwhelming amount of information being received, particularly in complex and dynamic situations (like social interaction), where reliance on flexible, reasonably accurate high-level concepts is essential. This scenario is consistent with many of the symptoms of autism (Van de Cruys, et al., 2014). In particular, people with autism are "less influenced by contextual information, and hence see the world more accurately (as it actually is)" (Van Boxtel & Lu, 2013, p. 1). Additionally, such a theory is consistent with the autistic tendency towards repetitive behaviors: they are a way of shielding oneself from the endless barrage of "surprise" caused by over-sensitivity to prediction errors (Van de Cruys, et al., 2014).

3. Are there significant evolutionary pressures on precision-weighting?

3.1. The effect of environmental predictability

If PP is correct, precision-weighting is a crucial feature of the "predictive engine." As we have seen, turning the dial too far in either direction (systematically under- or over-valuing prediction error) could have serious negative consequences for the effectiveness of the brain in helping an organism interact with its environment. Thus, one expects that less extreme deviations in the precision will have correspondingly less extreme, but still non-trivial, effects on perception and cognition. If appropriate for the environment, subtle adjustments of the average precision could presumably even *enhance*, rather than diminish, an organism's efficacy in responding to stimuli and taking action. And so long as average precision has any effect on reproductive fitness (whether positive or negative), then natural selection could act on that parameter to, over evolutionary time, make it better-suited to the local environmental conditions. This raises the question: If systematic differences in precision-weighting between species have arisen from natural selection, how would

those differences look? What sorts of environments would promote high or low precision in predictive processing?

For an organism with high prediction-error tolerance to be successful (in particular, more successful than a similar organism with low prediction-error tolerance), there must be something advantageous about ignoring a greater proportion of error signals. Ignoring more error signals, and thus creating novel predictions less frequently, is metabolically (computationally) cheaper—there's a benefit. The associated cost is that, with less prediction error to guide them, the predictions will be less "in sync" with the world. The cost of less accurate predictions is minimized when the environment is *predictable*. Therefore we expect that species in more predictable environments should have a higher tolerance for prediction error, relative to other species in their genus. In other words, predictable (simpler, more stable, e.g., "Arctic") environments should tend to promote "prediction-biased" cognition, while unpredictable (more complex, more unstable, e.g., "Amazon") environments should tend to promote "sensory-biased" cognition.

3.2. A reformulation in terms of r- and K- selection

To flesh out this hypothesis a bit more, let us connect it with the ecological theory of life histories. Naturally, different environments lend themselves to the evolution of different life-history strategies, which roughly form a continuum between "r-selected" species and "K-selected" species. (These names come from the logistic equation $dN/dt = rN(1-N/K)$ describing population growth, where N is the population size, r is growth rate when the population is small, and K is the carrying capacity.) In a nutshell, r-selected species follow a "fast life" strategy, while K-selected species follow a "slow life" strategy. Some of the salient characteristics of these two extremes are shown in the table below (Reznick, Bryant, & Bashey, 2002, p. 2).

r-selection	K-selection
Short life span	Long life span
Small body size	Large body size
Rapid development, early reproduction	Slow development, delayed reproduction
Reproduce once or few times over lifetime	Reproduce multiple times over lifetime
Many offspring	Relatively few offspring
Short gestation period (in mammals)	Long gestation period (in mammals)
Low paternal investment in offspring	High paternal investment in offspring
Variable, often lax, competition	Usually keen competition
Variable population size	Fairly constant population size (equilibrium)
Many individuals die before maximum age	Fewer individuals die before maximum age
Mortality often random, density-independent	Mortality more directed, density-dependent
"Productivity" / "quantity"	"Efficiency" / "quality"

Quintessential examples of r- and K-selected species are rabbits (along with many insects, fish, rodents, etc.) and wolves (along with humans, elephants, many birds, etc.), respectively.

The significance of this classification for us is in its connection to the species' environment: r-selection is associated with *unpredictable* or variable environments, while K-selection is associated with *predictable* or relatively constant environments. It does no good for an r-selected species to invest heavily in a few offspring, because they might die for unforeseeable and unavoidable (unrelated to quality) reasons before reaching reproductive maturity; the best strategy in that situation is to have many offspring quickly and hope some small fraction can survive and reproduce. Conversely, K-selected species, occupying a "settled" space, cannot afford the resource-cost of producing many offspring. And even if a K-selected organism could afford the cost of simply *producing* many offspring, those offspring would have poor chances of survival without a long, expensive "childhood," as they need to be able to compete for resources in a saturated niche.

Thus, the hypothesis (an unpredictable environment promotes greater precision, in the **PP** sense) can be adapted to a related hypothesis about r/K-selection: namely, that r-selected species are more sensitive to prediction error, and K-selected species less. (Of course, this is not quite the same as the original hypothesis, since r/K-selection and environment predictability are not identical

10

concepts. In fact, since "r" and "K" are merely convenient but fuzzy bundles of imperfectly correlated traits, this is really a *collection* of hypotheses about the relationship between average precision-weighting and any one of the qualities associated with r/K-selection in the table above. Some of these qualities might be highly correlated with precision, others less so.)

3.3. A counter-argument, a counter-counter-argument, and some ideas about testing

One can imagine an objection to the hypothesis along the lines of: "Unpredictable environments have more statistical noise, so they require quieter error signals to reach useful conclusions about the world. Conversely, predictable environments produce less frequent error signals, so that that *do* occur are more significant, and thus should be weighted more strongly."

I believe this counter-argument can be addressed as follows. The "noise" of an unpredictable environment (e.g., in the Amazon: semi-constant whooping from miles off, the persistent buzzing of insects in the background) gets integrated into the predictions, and thus the *remaining* errors (e.g., a sudden rustling in the brush next to me, the loud crack of a stick a few meters away) can be, and indeed *must* be, afforded greater significance. Likewise, in a predictable environment, the occasional important error signal (e.g., in the Arctic, from a polar bear's perspective: all of a sudden, the prediction of an all-white visual field is wrong, as a brown seal has become visible) does not need to be weighted especially strongly to kick off a prediction update, as it is "competing" with fewer other error signals for attention. Thus, everything remains consistent if we conceive of precision-weighting as a factor which works in tandem with the typical (temporal) density of errors.

Finally, let us conclude with a suggestion of how this idea could be studied further. For practical purposes, it is probably best to focus on one aspect of perception alone, say, visual perception. First, we would need a way of distinguishing between predictable and unpredictable environments—some information-theoretic measure of image complexity, applied to the visual background, ought to do the trick. Then, we would need to be able to measure any systematic

11

difference in "perceptive style" (that is, prediction-weighting) between species of the same genus (to reduce the confounding effect of relatedness) in predictable versus unpredictable environments. This is surely the tricky part! It would likely require a clever modification of one of the PP-consistent illusions used on humans, along with some way to consistently associate perceptions with specific actions in the animals being studied. Additionally, the illusion would have to be accompanied by a means of comparing error-sensitivity (like the switching rate for binocular rivalry). We can only give a few vague recommendations here, since any operational procedure would necessarily be constructed around the specific choices of species, environment, and perceptive phenomenon. In any case, undertaking the challenge of designing such a study could provide interesting new information about the survival value of different precision-weighting schemes for predictive brains.

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. doi:10.1007/s00429-012-0475-5

Apps, M. A., & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience and Behavioral Reviews, 0*, 85-97. doi:10.1016/j.neubiorev.2013.01.029

Clark, A. (2012). Do thrifty brains make better minds? *The New York Times*. Retrieved from https://opinionator.blogs.nytimes.com/2012/01/15/do-thrifty-brains-make-better-minds

Clark, A. (2013a). Are we predictive engines? Perils, prospects, and the puzzle of the porous perceiver. *Behavioral and Brain Sciences, 36*(3), 233-253. doi:10.1017/S0140525X12002440

Clark, A. (2013b). The many faces of precision (Replies to commentaries on "Whatever next? Neural prediction, situated agents, and the future of cognitive science"). *Frontiers in Psychology, 4*, 270. doi:10.3389/fpsyg.2013.00270

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

Clark, A. (2014). *Mindware: An introduction to the philosophy of cognitive science.* New York: Oxford University Press.

Clark, A. (2015). Embodied prediction. In T. K. Metzinger, & J. M. Windt (Eds.), *Open MIND.* Frankfurt am Main: MIND Group. doi:10.15502/9783958570115

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*, 127-138. doi:10.1038/nrn2787

Healy, K., McNally, L., Ruxton, G. D., Cooper, N., & Jackson, A. L. (2013). Metabolic rate and body size are linked with perception of temporal information. *Animal Behaviour, 86*(4), 685-696. doi:10.1016/j.anbehav.2013.06.018

Heslop, K. R. (2012). Binocular rivarly and visuospatial ability in individuals with schizophrenia. PhD thesis, Queensland University of Technology. Retrieved from https://eprints.qut.edu.au/59610/

Horga, G., Schatz, K. C., Abi-Dargham, A., & Peterson, B. S. (2014). Deficits in predictive coding underlie hallucinations in schizophrenia. *The Journal of Neuroscience, 34*(24), 8072-8082. doi:10.1523/JNEUROSCI.0200-14.2014

Horgan, J. (2016). Are brains Bayesian? *Scientific American.* Retrieved from https://blogs.scientificamerican.com/cross-check/are-brains-bayesian/

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: W. H. Freeman.

Ponge, J.-F. (2013). Disturbances, organisms and ecosystems: a global change perspective. *Ecology and Evolution, 3*(4), 1113-1124. doi:10.1002/ece3.505

Reznick, D., Bryant, M. J., & Bashey, F. (2002). r- and K-selection revisited: The role of population regulation in life-history evolution. *Ecology, 83*(6), 1509-1520. doi:10.1890/0012-9658(2002)083[1509:RAKSRT]2.0.CO;2

Van Boxtel, J. J., & Lu, H. (2013). A predictive coding perspective on autism spectrum disorders. *Frontiers in Psychology, 4*, 19. doi:10.3389/fpsyg.2013.00019

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review, 121*(4), 649-675. doi:10.1037/a0037665