Enhancing the Scope of Predictive Processing via Nonlinearity

Luis H. Favela
Department of Philosophy and Cognitive Sciences Program
University of Central Florida

Mary Jean Amon
School of Modeling, Simulation, and Training
University of Central Florida
Abstract

Predictive processing (PP) is an increasingly popular general theoretical framework for understanding behavior and cognition; particularly perception. However, PP is limited by an over-reliance on linear statistics common to Bayesian approaches. Consequently, in order to more fully account for behavioral, cognitive, and neural activity, PP approaches should incorporate, and make more central, nonlinear theory and methods. We proceed as follows: First, we discuss the linearity inherent to traditional Bayesian analyses and the consequences such methodological commitments have for understanding the systems being analyzed. Second, we discuss what nonlinear theory and methods means in the current context. Third, we present a recent attempt at incorporating nonlinearity in PP, and demonstrate how it falls short due its conception of “nonlinearity” as being merely a matter of noise or random fluctuations in a system. We conclude by recommending ways nonlinearity can be incorporated in PP frameworks.

Keywords: Bayesian analysis; linearity; nonlinearity; predictive processing
“The elegant body of mathematical theory pertaining to linear systems [...] tends to dominate even moderately advanced University courses [...]. The mathematical intuition so developed ill equips the student to confront the bizarre behavior exhibited by the simplest of discrete nonlinear systems [...]. Yet such nonlinear systems are surely the rule, not the exception, outside the physical science. [...] Not only in research, but also in the everyday world.” (May 1976, 467)

1. Introduction

Perhaps the one thing cognitive scientists, neuroscientists, philosophers of mind, and psychologists can agree on is that there is no clear frontrunner for the title, “Grand Unified Theory of the Mind.” Contenders vying for the title include coordination dynamics (Bressler and Kelso 2016), free-energy principle (Friston 2010), network theory (McIntosh 2000), neural Darwinism (Edelman 1987), and neural reuse (Anderson 2010), just to name a few. Although it is unlikely that any of those will be crowned as the single unified theory of the whole mind any time soon, there are strong contenders for theories of particular aspects of the mind.

In recent years, predictive processing has garnered much interest as a possible unified theory of perception and action. There is no single theory of “predictive processing” that is accepted by all who use the term (Wiese and Metzinger 2017). However, predictive processing (PP) is generally understood as a framework centering on the claim that the brain/mind is in many ways a hypothesis-testing system that attempts to minimize predictive errors based on sensory input from the body and world (Clark 2013, 2015; Friston 2010; Hohwy 2013; Wiese and Metzinger 2017). Put differently, the brain/mind is in the business of making predictions about states of the body and world, and then correcting those predictions when something is
incorrect. An illustrative and common example of PP is when somebody hears a sound and believes it is a song they know, but it turns out that the sound is not the song they thought it was or any kind of music altogether (Clark 2019; Merckelbach and van de Ven 2001). In such cases, when people realize the sound is not what they thought it was, the next time they hear it they do not make the same mistake of believing it to be a familiar song. This process of making predictions and updating based on errors is typically accounted for in terms of Bayes’ rule, which states that explanations of a set of data (e.g., sensory input) are based on a set of candidate explanations that have prior credibility. When new data is presented, credibility is shifted towards the better candidate explanations and away from those that do not explain as well (Kruschke and Liddell 2018). This conception of hypothesis testing is aligned with Bayes’ rule, or the “mathematical relation between the prior allocation of credibility and the posterior reallocation of credibility conditional on data” (Kruschke 2015, 100).

Although the exact connection between PP and Bayesian approaches can vary quite broadly (e.g., Harkness and Keshava 2017; Sanborn and Chater 2016; Thornton 2017), it is the prevailing interpretation. This is in no small part due to the fact that there is no single interpretation of “Bayesian.” On the one hand, there is a stronger interpretation in which the brain is literally “Bayesian” (Friston 2012; Knill and Pouget 2004). That is to say, various features of the brain, such as computations and neuronal firing, are Bayesian in nature (Clark 2013, 105; Kording 2014). On the other hand, a weaker interpretation is one in which Bayesian methods are fruitful when analyzing and modeling behavioral, cognitive, and neural phenomena related to PP (Colombo, Elkin, and Hartmann 2018; Doya, Ishii, Pouget, and Rao 2007; Harkness and Keshava 2017). Here, we focus on a more intermediate interpretation, specifically, we are concerned with inferences made about targets of inquiry based on Bayesian data analyses.
At least insofar as PP is concerned, we claim that an over-reliance on linear statistics common to Bayesian approaches runs the risk of leading investigators to transfer the assumptions of the analyses used to assess a system to be assumptions about the system itself. These assumptions may come at the cost of overlooking other system characteristics, for example, dynamic fluctuations, qualitative shifts in behavior, and other nonlinear patterns. In short, we are concerned that the following inference is not always justified: If Bayesian analyses are useful for quantifying and modeling the brain/mind, then the brain/mind is a Bayesian system. In order to more fully account for behavioral, cognitive, and neural activity, PP approaches should incorporate, and make more central, nonlinear theory and methods, which capture behavior intrinsic to a variety of brain/mind systems.

In the next section, we discuss the linearity inherent to Bayesian analyses and the consequences such methodological commitments have for understanding the systems being analyzed. Second, we discuss what nonlinear theory and methods means in the current context. Third, we present a recent attempt at incorporating nonlinearity in PP, and demonstrate how it falls short due its conception of “nonlinearity” as being merely a matter of noise or random fluctuations in a system. We conclude with an example and set of recommendations concerning ways nonlinearity can be incorporated in PP frameworks.

2. Bayesian analysis as linear analysis

There are many excellent introductions to Bayesian analysis and theory (e.g., Bernardo and Smith 2000; Easwaran 2011a, 2011b; Kruschke 2015; Kruschke and Liddell 2018; Spiegelhalter and Rice 2009). Our aim here is not to provide another comprehensive introduction, but to provide enough of an overview to support our claim that traditional Bayesian
analyses are typically types of linear analysis. At its most basic, “linearity” refers to a cause-effect relationship in which outputs are proportional to inputs (Lam 1998, 7).

\[ f(x + y) = f(x) + f(y) \]

Here, an event, \( f \), is characterized as composed of \( x \) and \( y \), or \( f \) can be characterized as \( x \) and \( y \) added together (Ivancevic and Ivancevic 2007, 52). The event \( f \) is equally conceptualized as being an event comprised of \( x \) and \( y \) or as an event comprised of \( x \) added to \( y \). Explaining Bayesianism is not as straightforward.

Like PP, there is no single “Bayesianism” (e.g., default, frequentist, and subjectivist; Mayo 2018; Easwaran 2011a). Nevertheless, there are some core ideas common to its various forms, namely, Bayes’ theorem (Bernardo and Smith 2000):

\[ P(H|D) = P(D|H) \times P(H)/P(D) \]

\( H \) refers to a hypothesis and \( D \) to data. \( P(H) \) refers to the prior probability of belief in \( H \) before data \( D \) is obtained. \( P(D) \) refers to the probability that the data \( D \) is true. Thus, \( P(H|D) \) captures the probability of hypothesis \( H \) being true after data \( D \) is obtained.

With respect to PP, Bayes’ theorem underlies the claim that various cognitive and perceptual capacities are the result of a probabilistic computational process. Think back to the example above of somebody hearing a sound and believing it is a song they know, such as Bing Crosby’s “White Christmas” (Clark 2019; Merckelbach and van de Ven 2001). In this case, \( H \) refers to the hypothesis that the sound heard is the song “White Christmas,” and \( D \) refers to data such as the sound. Thus, \( P(H|D) \) captures the probability of hypothesis being true—“I believe I am hearing Bing Crosby’s ‘White Christmas’”—after data (i.e., sound emitted from a car stereo speaker) is obtained. As it turns out, in this case the sound was not “White Christmas.” It was Los del Rio’s “Macarena Christmas,” which sampled a part of “White Christmas.” Next time you
turn the radio on, since your prior hypothesis has been updated, when you hear “Macarena Christmas,” you are correct to hypothesize that the sound could be either Bing Crosby or Los del Rio. The question now is, what makes Bayes’ theorem, and analyses based on it, linear?

Bayesian analyses can be considered linear analyses primarily due to their overlap with frequentist statistics, in particular, a number of assumptions underlying both types of statistics. Target practice serves as a metaphor to help to understand these assumptions, especially in terms of naturally occurring variation during repeated trials (Amon and Holden 2019; Klein 1997). While firing a gun at a shooting range, each bullet strike varies in how close it lands to the target’s bullseye, with strikes clustering closer together around the center and becoming progressively more dispersed away from the bullseye. Even with the most experienced shooter, the bullet’s trajectory will vary as a function of many influences, including the bullet’s shape, amount of gunpowder, etc. A bullet’s deviation from the target can therefore arise from a variety of independent perturbations that have an additive effect on the overall trajectory. The metaphor of target practice underscores the basic logic of the *normal distribution* developed by Gauss (1809; Fischer 2011) and Laplace ([1812] 1820; Fischer 2011). Much like the concentration of bullet strikes around a target, Laplace reasoned that measurement error is often the result of many random, weak, and independent sources of variation that have an additive effect on measured behavior. The sum effect of these numerous perturbations is a bell-shaped or “Gaussian” distribution of measurements that are relatively symmetrical around a central tendency. This distribution and variability of events is known as the *central limit theorem*, which is a statistical theorem that assumes that the sum of numerous independent influences will have a normal, or upside down ‘U’-like shaped, distribution (Cumming 2012). Normal distributions are also the result of a *frequentist* view of events, that is, each event (e.g., bullet fired at a target) is
considered isolated such that each event outcome is equally probable (de Finetti in Bernardo and Smith 2000, 3). Consequently, events falling along normal distributions are conceived of as resulting from stationary processes. Processes, or events, are stationary when temporal ordering does not affect the probability of their distribution, such that mean and variance do not change with time (Everitt 2006).

The logic of the Gaussian distribution is rooted in systems where the probability of observed outcomes varies based largely on chance. Scholars traditionally framed variation during coin tosses, dice rolls, measurement accuracy of star positions, and target practice as being randomly influenced with no common thread of history linking observations (Klein 1997). As a result, linear statistics have been situated in a context that emphasizes logical variation. Logical variation is the difference between one discreet measurement and another, where variation is not influenced by history. This contrasts with historical variation, or fluctuations that occur as a process unfolds over time (Klein 1997). Thus, while the Gaussian distribution has been the basis of numerous statistical advancements, its development left statisticians with the formidable task of applying principles derived from logical variation to natural systems whose behavior is influenced by history. While debates concerning the advantages and disadvantages of commitments to assumptions of logical and historical variation occurred, a number of linear statistics were developed, including correlation, simple regression, and multiple regression. All the while, the “perpetual flux” of natural systems was largely ignored as worthwhile targets of investigation in favor of evaluating discrete measurements (Klein 1997). Consequently, each of these statistics relied on the central limit theorem, carrying with them assumptions inherent to the Gaussian distribution.
Due to the fact that linear statistics only account for logical variation in independent and dependent variables, additional fluctuations within biological or social systems (i.e., historical variation) is commonly relegated into the category of unsystematic noise. Unsystematic perturbations are assumed to be weak and occur randomly, such that their effects will cancel each other out. The variation accounted for by the primary variables of interest (i.e., independent and dependent variables) must exceed the variation produced by “unsystematic noise” by approximately two standard deviations in order for the relationship between these variables to be considered statistically significant using null hypothesis significance testing. The ability to uncover statistically significant results rests on the idea that the noise within a system is unsystematic.

In this way, the general linear model and linear analysis served as the default organizing theory for the sciences of mind, especially psychology and then cognitive science and neuroscience. That is, the bulk of the quantitative theories in behavioral and social sub-disciplines can be reframed as a version of a regression equation: Factors, processes, latent variables, and other constructs can be present, absent, or combined in weighted sums to approximate empirical measurements. Consider stage-wise developmental theories, feature-based perceptual theories, additive factors theory, the double dissociation and dual-process theories of psychology and cognitive science: they are all frequentist and stationary at their core.

All statistical procedures require non-trivial, a priori assumptions. In the case of linear statistics, they are situated within a context that has largely emphasized logical variation and ignored historical variation. The widespread generalization of the Gaussian distribution to biological systems (e.g., human action, perception, and cognition) has left many statistical procedures to rely on the following four underlying assumptions: First, independent
perturbations have *additive* effect on global behavior. Second, experimental measurement will result in a *Gaussian* distribution. Third, systems are best characterized by *logical variation* and not historical variation. Fourth, variation not accounted for by an experimental manipulation is weak, *random noise* that will cancel out over repeated observations. Note that the third and fourth assumptions focus on the concept of stationarity. The message implicit in these assumptions is that the systems we study in the mind sciences, and those that we characterize with linear statistics, are not influenced by history and, save the effect of the manipulation, are relatively isolated events that are static over time.

A general lack of knowledge about basic assumptions of linearity (e.g., normality and homogeneity of variance) means that they are rarely checked and are altogether violated, leading some researchers to describe common statistical procedures as “opportunistic” (Hoekstra, Kiers, and Johnson 2012). For the minority of researchers that do attempt to satisfy underlying assumptions of their linear analyses, they often rely on techniques such as averaging, outlier censorships, transformations, and the robustness of *t*- and *F*-tests. There are strategies to deal with a number of statistical violations common to linear statistics, but other assumptions of linearity are fundamentally mismatched with the perpetually fluctuating systems often studied in the brain/mind sciences. It is rarely called into question that linear statistics treat changes in a system outside the effect of the independent variables as unsystematic noise that will cancel out over repeated observations. Given that Bayesian analyses typically adhere to the above four assumptions of linear statistics, so too do they fall victim to the limitations facing other linear statistics. For a clear depiction of this, let us examine Bayesianism in terms of both a data analysis and as the theoretical foundation for PP.
In terms of a data analysis, Bayesianism typically adheres to many of the assumptions of frequentist statistics, including stationarity and linearity (Bernardo and Smith 2000, 19). The following are a few examples that demonstrate such commitments: When utilizing Bayes’ factor as a means to assess competing hypotheses, a normal distribution of data is usually assumed (e.g., Dienes 2014). Frequentist statistics are often needed to evaluate inferences made based on Bayesian analyses, that is, the conclusions of Bayesian analyses can require further validation by frequentist methods (Gelman et al. 2009). Additionally, before a Bayesian approach is applied to certain data sets (e.g., large sample sizes), traditional frequentist statistical assumptions are often appealed to in order to make approximations about the data, for example, the assumption of normal distributions (Gelman et al. 2009). The need to supplement Bayesian methods with traditional statistics is particularly evident in the mind sciences, for example, regarding perception where experimental work has usually considered simple tasks based on static variables (e.g., Rao and Ballard 1999) and underlying neuronal dynamics are treated as linear (e.g., Makin, Dichter, and Sabes 2015). Consequently, simple perceptual tasks based on static variables and linear neuronal dynamic models run the risk of being neither ecologically valid nor biologically realistic. These examples draw attention to limitations of Bayesian data analyses brought about by assumptions shared with linear statistics. Still, those shared assumptions may bring about even more limitations for Bayesianism as the theoretical foundation of PP.

PP is essentially a form of Bayesian inference (Clark 2013; Harkness and Keshava 2017; Hohwy 2013; Wiese and Metzinger 2017), which is itself an instantiation of Bayes’ theorem. As an account of certain features of mind—especially perception—PP is most clearly understood as appealing to a linear Bayes approach. Remember, along the lines of Bayes’ theorem, PP conceives of perception, for example, as being about the probability of a hypothesis $H$ being
supported by the probability of the data $D$ supporting that hypothesis being true, $P(H|D)$. Take visual perception, for example: $H$ could be the hypothesis that a person I see approaching from the distance wearing a red shirt and black pants is my friend Silvia, which is based on the prior information that I believe that Silvia usually wears a red shirt and black pants. Here, $D$ is sensory data of a person in the distance wearing a red shirt and black pants. Thus, $P(H|D)$ is the probability of $H$ being supported by $D$. This example is straightforwardly captured by Bayes’ theorem and is conceptualized as a Bayes linear approach (Goldstein 2015). The Bayes linear approach is applied to problems where we want to combine prior judgments of uncertainty ($H$; e.g., belief that Silvia usually wears a red shirt and black pants) with observational data ($D$; e.g., seeing someone in the distance wearing a red shirt and black pants), and we use expected value (i.e., the person in the distance is Silvia) to express the perceptual judgment instead of probabilities (Goldstein 2015, 1). In this way, the Bayes linear approach further demonstrates the four assumptions of linear statistics (Goldstein 2015, 2): First, the independent perturbations are additive in that they are treated as uncorrelated sums. Second, the distributions of the calculations will be Gaussian. Third, variation is accounted for in terms of other measurements and not previous states. Fourth, variation in beliefs (i.e., hypotheses and priors) are reduced or cancelled out by variation in the data, as is variation in data reduced or cancelled by variation in said beliefs (Goldstein 2015 makes explicit reference to similarities with de Finetti; Bernardo and Smith 2000, 3).

In spite of the apparent limitations of Bayesianism, our aim is not to deride PP. Though it remains to be seen if PP is the “Grand Unified Theory of the Mind,” we agree with many of its proponents that it is a compelling theory of particular aspects of the mind, especially perception. For instance, PP could subsume various, often opposing, theories of perception such as
computational-representational (Wiese and Metzinger 2017) and ecological, embodied, and enactive (e.g., Anderson 2017; Clark 2015; Lara et al. 2018). However, in order to more fully account for behavioral, cognitive, and neural activity, PP approaches should incorporate, and make more central, nonlinear theory and methods.

3. Nonlinearity is the rule, not the exception

Truly linear systems are rare in nature and tend to be limited to artifacts (e.g., combustion engines, computers, dishwashers, etc.). Despite that fact, the sciences of the brain/mind have primarily focused on conceptions of targets of inquiry as linear and have utilized linear methodologies. From Donders’ development of a subtractive method to attempt to decompose response time performance into components ([1869] 1969), to the general linear model central to analyses of functional magnetic resonance imaging data (fMRI; Huettel, Song, and McCarthy 2009), linearity has been the rule and not the exception in scientific investigations of the brain/mind. Theoretically, the issue is whether or not the assumption of linear spatial and temporal structures in neural physiology and behavioral and cognitive activities is justified or not.

Nonlinearity is the rule and not the exception across the physical and social world (May 1976). At its most basic, a system is nonlinear if the causal relationship between its current state, former state, and future state are not linear. Rather than assuming additivity and independence of components as in linear statistics, nonlinear systems may be composed of interdependent components that exhibit exponential or multiplicative dynamics (Lam 1998, 6). Examples of nonlinear dynamics include qualitative shifts in behavior, various forms of structured noise (e.g., fractal scaling) or fluctuations over time, as well as chaos and self-organization.
There is ever more empirical data to support the claim that the brain/mind is a nonlinear system. Here is but a small sample: In regard to spatial organization, there is ample data to suggest that across micro- (e.g., molecular networks), meso- (e.g., synaptic clusters), and macro-scales (e.g., neuronal networks), the brain is fractal and multifractal in structure (Di Ieva 2016). In regard to temporal characteristics, there is much research indicating nonlinear dynamics in single-neuron activity (Favela et al. 2016; Izhikevich 2007), neocortical circuits (Beggs and Plenz 2003), and neuronal networks (Sporns 2011). At larger scales exhibiting behavioral and cognitive activity, there is data to suggest nonlinearity as well, for example, decision-making (van Rooij et al. 2013), postural control (Riley et al. 1998), speech categorization (Tuller et al. 1994), temporal estimation (Amon et al. 2018), and visual search (Aks et al. 2002), just to name a few.

Notably, the analysis of nonlinear systems like those described above have been facilitated by a wide range of nonlinear analytic tools. Although these categories have much overlap, for the sake of simplicity, we distinguish two groups of nonlinear methods: modeling and time-series analysis (Richardson, Paxton, and Kuznetsov 2017; Riley and Van Orden 2005). Typical modeling techniques when investigating nonlinear systems include agent-based (e.g., cellular automata), computational (e.g., simulations), dynamical (e.g., differential equations), and network (e.g., neuronal networks). Time-series analyses utilized when investigating nonlinear systems include dynamical correlation, entropy (e.g., sample entropy), fractal and multifractal (e.g., detrended fluctuation analysis), phase space reconstruction (e.g., attractor reconstruction, e.g., Lorenz attractor), recurrence quantification analysis (e.g., cross-recurrence quantification analysis), and wavelet (e.g., wavelet cross-spectrum).
A number of lessons can be taken from the theory and methods that comprise nonlinearity: The presence of nonlinearity in many natural systems—including those related to the brain/mind—challenges the basic assumptions of linear statistics, including additivity (i.e., linearity), independence of perturbations (i.e., modularity), and logical, unsystematic variability. In contrast, the ubiquity of nonlinearity in natural systems suggests the following are more appropriate assumptions: First, nonlinear methods take into account that interdependent perturbations have multiplicative feedback effects on global behavior. Second, nonlinearity acknowledges that variation within a system is often systematic and reflects underlying processes that interact to support behavior. Third, natural systems are in a state of perpetual flux and are often best characterized by historical variation. Fourth, experimental measurement of natural systems that entails discrete, cross-sectional measurements will inevitably capture changes in a system that are due to complex, interaction-dominant dynamics (i.e., and not component-dominant dynamics).

The fifth lesson has not been discussed yet but is important: Nonlinear approaches have three complementary and interrelated features that drive experimental design, inform hypotheses, and facilitate explanation and understanding: qualitative explanations, quantitative explanations, and theoretical explanations or generalizable principles of how the biological system works. For example, when attempting to provide an explanation and understanding of phenomenon X, dynamical methods such as differential equations could be used to model the temporal evolution of the system as well as plotting the behavior in a phase space reconstruction. In this way, a quantitative and qualitative account of X is given. In order to understand the nonlinear features of the system and to explain why it has those features, the quantitative and qualitative methods can be supplemented with appeal to “principles,” or rules that underlie the spatial and temporal
structure of many nonlinear systems. Examples of principles include catastrophe flags and universality classes. *Catastrophe flags* are a set of dynamical patterns that anticipate or accompany phase transitions in nonlinear dynamical systems (Gilmore 1993; Isnard and Zeeman 1976; Thom 1975). Hysteresis, multistability, and sudden change are three of the eight known catastrophe flags (Gilmore 1993). Observing hysteresis, or one of the other flags, strongly suggests that a system is nonlinear and undergoes phase transitions. *Universality classes* are distinct system behaviors that are determined by a few characteristics, take place across multiple spatial and temporal scales, and are substrate neutral (Batterman 2000; Thouless 1993). Self-organized criticality (SOC) is one universality class that is observed across a variety of natural systems (Jensen 1998), including various brain/mind systems (Beggs 2007; Cocchi et al. 2017; Hesse and Gross 2014; Plenz and Niebur 2014). SOC is posited as an explanation for the ability of some nonlinear systems to exhibit multiple spatial and temporal scales that organize near critical states, which is a balance between two qualitatively different behaviors. Some argue that SOC explains why nonlinear systems are often dictated by power law features such as fractal self-similarity (e.g., Bak et al. 1987). Accordingly, if a system exhibits characteristics such as fractal structured behavior, then it may be appropriate to use nonlinear methods and provide explanations via principles derived from, for example, catastrophe theory and universality classes. Thus, if systems that have exhibited nonlinear characteristics in other research are only assessed via Bayesian approaches, then it is quite likely that key features of those systems are being overlooked.

The application of Bayesian paradigms has undoubtedly led to many successes in the brain/mind sciences. However, there remains a fundamental contradiction between the common statistical assumptions of additivity, linearity, and straightforward cause-effect relationships that
underlie traditional Bayesian approaches and the nonlinearity inherent to the systems such assumptions are applied to. Here, we provided a very small sample of the large amount of literature that points to brain/mind phenomena as inherently nonlinear, as well as the various methods used to properly characterize nonlinear dynamics. In doing so, we have attempted to motivate the point that nonlinearity is the rule and not the exception (cf. May 1976), especially in regard to phenomena in the brain/mind sciences. In the next section, unsuccessful attempts at incorporating nonlinearity in Bayesian approaches are presented. We conclude by presenting an example and recommendations for properly integrating nonlinearity in PP approaches.

4. Nonlinear predictive processing

In our overview of Bayesian analysis, we noted that there are multiple forms of Bayesianism (e.g., default, frequentist, and subjectivist; Mayo 2018). Our explication of Bayes’ theorem as demonstrating frequentist and stationary views underlies the assessment of Bayesianism as centrally being a type of linear analysis. With that said, there has been work attempting to incorporate Bayesian and nonlinear approaches, for example, Bayesian model comparison of neural activity (Chen et al. 2010; Jacobsen et al. 2008) and natural image reconstruction (Naselaris et al. 2009). One attempt to incorporate nonlinear features in a Bayesian approach is by way of predictive coding, which provides the process by which PP occurs:

Predictive coding postulates that neural networks learn the statistical regularities inherent in the natural world and reduce redundancy by removing the predictable components of the input, transmitting only what is not predictable (the residual errors in prediction). (Huang and Rao 2011, 580)
This postulate is formalized via Bayes’ theorem (Huang and Rao 2011, 581):

\[ P(r|I) = P(I|r)P(r)/P(I) \]

\( P(r|I) \) is the model of the visual system, for example, whereby an internal representation of the external world is predicted based on sensory data. Here \( I \) is the probability of an image given hidden internal model parameters \( r \) (e.g., underlying neural network activity). Thus, for visual input \( I \), the system selects parameters \( r \) to maximize the posterior probability that the system’s representation of the world is accurate. This occurs via a straightforwardly linear process (Huang and Rao 2011, 581-583, 585, 588, 590). In addition, the process is treated as being top-down, whereby feedback occurs from higher-order visual areas passing predictions from lower-level neural activity via hierarchical neural networks (Rao and Ballard 1999).

Friston and colleagues have attempted to incorporate nonlinearity into their models of predictive coding (e.g. Friston 2010; Friston and Kiebel 2009). Friston (2010) provides the following generative model of predictive coding (note that explaining all of the variables is unnecessary for our current purposes):

\[ p(\tilde{s}, \theta) = p(\tilde{s}|\theta)p(\theta) \]

where

\[ s = g(x^{(1)}, v^{(1)}, \theta^{(1)}) + z^{(1)} \]
\[ x^{(1)} = f(x^{(1)}, v^{(1)}, \theta^{(1)}) + w^{(1)} \]

Here \( g^{(i)} \) and \( f^{(i)} \) refer to nonlinear functions with causal and hidden states with parameters \( \theta^{(i)} \) (Friston 2010, 130). Although Friston does not specify values for the variables, he states that \( z \) and \( w \) contribute “random fluctuations” and “noise” (2010, 130), which is intended to capture nonlinear characteristics. In other work, Friston and colleagues again describe such nonlinear functions as “hidden states” that contribute “observation noise” and “random fluctuations,”
among other characteristics (Friston and Kiebel 2009, 1211). If we take these models as indicative of current thinking regarding the relationship of Bayesianism, nonlinearity, predictive coding, and PP, then we must conclude that PP has yet to properly incorporate nonlinearity.

By treating nonlinear variables as random fluctuations and noise, Friston and colleagues are revealing that their Bayesianism remains committed to the assumptions of linear statistics. In particular, they demonstrate the fourth underlying assumption of linear statistics: Variation that is not accounted for by an experimental manipulation is assumed to be unsystematic noise that is not influenced by historical variation. While Friston and colleagues incorporate nonlinear variables, their status as random fluctuations and noise relegate their contributions as carrying no great weight in a Bayesian model that remains committed to the assumptions of linear statistics—take notice of the fact that this work is explicitly committed to Gaussian distributions of data involved in predictive coding (e.g., Friston 2010, 130; Friston and Kiebel 2009, 1212).

Other researchers have drawn attention to the challenges of integrating Bayesianism and nonlinearity. In work on populations of neuronal activity from a Bayesian approach, Ma and colleagues point out that restricting neuronal models to Bayesian methods will be limited because Bayesian inference is additive (i.e., linear) and real neurons are nonlinear (Ma et al. 2006, 1436). Similarly, in a review of Bayesian network analysis of fMRI, Mumford and Ramsey (2014) emphasize that “most [Bayesian network] approaches assume linearity and Gaussianity” (2014, 576). The question now is, what would it take for PP to properly integrate Bayesianism and nonlinearity?

In order for PP to incorporate nonlinearity, we recommend that such an approach do the following: First, models must incorporate variables that are truly nonlinear and not merely random or unsystematic noise. For example, parameters could be set to experimentally-validated
nonlinear structure, such as 1/f signals. Second, if appropriate for a particular model, then nonlinear functions ought to be exponential or multiplicative and not just additive, as is the case for \( g^{(i)} \) and \( f^{(i)} \) above. Third, models ought to be capable of exhibiting qualitative shifts in behavior or phase transitions and, thereby, exhibit catastrophe flags such as hysteresis and multistability. Fourth, principles such as those derived from universality classes ought to be able to contribute to prior probabilities (i.e., \( P(H) \)). For example, prior probabilities can be set to be in line with parameters that facilitate critical state behavior (e.g., SOC), such as those found across spatial and temporal scales from underlying neuronal dynamics (e.g., Beggs and Plenz 2003) to gross-level behavior (e.g., Ramos et al. 2011).

In addition to these four recommendations, a nonlinear-Bayesian-PP should explicitly appeal to the strengths of nonlinear approaches, which includes a focus on historical variation, as well as an appreciation that dynamic structure is essential to understanding nonlinear phenomena. Also, PP can increasingly incorporate explanations that include both quantitative (e.g., differential equations) and qualitative (e.g., phase space reconstruction) components. Lastly, PP can leverage theoretical justification provided by nonlinear dynamics in the form of principles such as those found in catastrophe flags and universality classes.

We offer bistable visual percepts as an example of real-world phenomena that put our recommendations to work and demonstrate the necessity of incorporating nonlinearity in PP accounts. A bistable percept is one that presents two or more stable states. Formally speaking, a dynamical system \( \dot{x} = f(x) \) is bistable (or multistable) when it possess two or more stable equilibrium points (Noori 2014). Examples of bistable visual percepts include the Rubin vase, Schroeder stairs, and—the philosopher’s favorite—duck-rabbit (Figure 1A-C). Here, we focus on the Necker cube (Figure 1D).
If PP is correct in its current form, namely, with its linear foundations, then the transition between bistable states ought to occur on regular intervals and, thus, exhibit logical variation based on additive effects. Yet, in our phenomenological experience, it does not seem true that transitions in how the Necker cube’s orientation is perceived—i.e., from the front facing down and left to the front facing up and right, or vice versa—occur at regular intervals. Instead, the state transitions seem to occur with spontaneous irregularity, for example, requiring more or less time and effort to change orientations. In addition, it also seems that the more experience we have viewing the Necker cube, the more variation there is in the occurrence of those transitions. Since PP does not account for the irregularity and historically-based variation of such perceptual experiences, then it must either be the case that our sense of timing regarding the state transitions of bistable visual percepts is radically inaccurate (a very real possibility; cf. Schwitzgebel 2008) or PP is inadequate to account for those features. Although evidence based on introspection can be methodologically problematic, since PP is purported to account for phenomenology as well, we believe that the Necker cube and other bistable visual percepts are cases of the PP framework falling short.

In order to account for bistable visual percepts such as the Necker cube, nonlinear theory and methods are needed to supplement PP. As we discussed above, nonlinear methods account for, among other things, features such as historical variation. Hysteresis is one characteristic of some nonlinear phenomena that exhibit historical variation. Although there are various kinds of hysteresis, the basic idea is that a system’s current state is dependent on its history (Haken 1983). Ferromagnetic materials provide a clear example of the history-dependence resulting from the hysteresis effect. Once a material, such as iron, becomes magnetized, if it is demagnetized, then it will take a different magnetic field to magnetize it again. Thus, the current state of the system
depends on its previous state such that magnetization does not occur at context-free values.

Ta’eed and colleagues (1988) investigated hysteresis in the perception of the Necker cube. They demonstrated that nonlinear dynamical models of bistable perception were superior to linear models, primarily because the latter do not effectively account for multivalued functional relationships such as those involved in visual perception. Specifically, Ta’eed et al. demonstrated that bias (i.e., history), among other variables, affected the perception of the Necker cube upon subsequent presentations, thereby indicating a nonlinear process. Along those lines, Haken (1983; 2006) has also conducted experimental work on hysteresis exhibited during visual perception of bistable percepts. Although some kinds of visual perception may be amenable to explanatory accounts based on linear assumptions (e.g., Fischer and Whitney 2014), the work of Ta’eed and colleagues demonstrate that others require a nonlinear-based framework.

Consequently, for PP to be a full account of perception, it must be able to account for nonlinearity.

PP and nonlinearity have both facilitated understanding of various aspects of behavioral, cognitive, and neural activity. While they may be framed as competing perspectives, attempts—though unsuccessful—have been made to combine them. The above recommendations and example from bistable visual percepts provide a preliminary attempt at guiding the successful integration of PP and nonlinearity. Such guidelines can maintain core features of PP (e.g., predictive control, statistical estimation, and top-down processing; Wiese and Metzinger 2017) and capture a greater degree of biological realism via nonlinear dynamical concepts (e.g., historical variation, nonlinear patterns of behavior, and qualitative shifts). Understood in this way, PP would join much of the everyday natural world in following the rule of nonlinearity and not the exception of linearity (cf. May 1976).
References


doi:10.3389/fpsyg.2014.00781


Figure 1. Examples of bistable visual percepts. (A) Rubin vase; (B) Schroeder stairs; (C) duck-rabbit; and (D) Necker cube (Image rights: CC BY-SA 3.0 or public domain).