Check for updates

Received: 4 October 2021 | Revised: 15 January 2023 | Accepted: 18 January 2023

DOI: 10.1111/jopy.12815

ORIGINAL ARTICLE

WILEY

Behavioral regulation relies on interacting forces and predictive models

Stephen J. Read[1] | **Lynn C. Miller**[2]

[1]Department of Psychology, University of Southern California, Los Angeles, California, USA

[2]School of Communication, USC Annenberg School for Communication and Journalism, University of Southern California, Los Angeles, California, USA

Correspondence
Stephen J. Read, Department of Psychology, University of Southern California, Los Angeles, CA, USA.
Email: read@usc.edu

Abstract

Objective: We discuss how our recent neural network model of personality and motivation can explain many aspects of the regulation of behavior.

Method: Contrary to approaches that focus on a goal-corrected, set-point, and discrepancy-reducing mechanism, we argue that many aspects of regulation can be understood in terms of two other mechanisms. First, many aspects of the stability and coherence of personality, as well as the dynamics of personality, can be understood in terms of the interaction of forces within organized motivational systems, and their interaction with features of the environment and interoceptive states, that identify an individual's current needs. This has been described as a *settling* point or equilibrium of forces model, rather than a *set*-point architecture. Second, regulation has been shown to depend also on the use of predictive models of the world, either learned or innate. Such models can be thought of as feedforward models, in contrast to the feedback models characteristic of set-point, goal-corrected systems.

Results and Conclusions: We describe a neural network model of these processes that simulates the behavior over time and situations of an individual and shows how important regulatory processes can operate through a process of interactive forces and predictive models of the world.

KEYWORDS

behavioral regulation, computational model, motivation, motivational dynamics, neural network, personality, personality dynamics

1 | INTRODUCTION

We have recently developed a neural network model of personality and motivation (Read et al., 2010, 2017, 2021) that can explain many aspects of the regulation of behavior. In our model, the stability and coherence of personality, as well as the dynamics of personality, can be understood in terms of the interaction of forces within organized motivational systems (settling point or equilibrium of forces), and the use of predictive models of the world[1]; stability and regulation in our model result from these processes, rather than from a discrepancy-reducing process, as is found in models that rely on goal-corrected systems. As we argue in more detail later, this focus on settling point and predictive models is consistent with a considerable amount of recent research that strongly argues

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2023 The Authors. *Journal of Personality* published by Wiley Periodicals LLC.

that goal-corrected, discrepancy-reducing processes do not provide sufficient explanations for many aspects of physiological or motivational regulation.

Researchers and theorists have identified various different mechanisms for the regulation of physiological systems and behavior. Probably the best-known mechanism is a set-point, discrepancy-reducing mechanism, borrowed from cybernetic control theory, where the organism acts to reduce the discrepancy between their current state and the set point (or goal). A typical example of this would be a thermostat in a combined heating–cooling system. If the current temperature is lower than the set point, then the heat would come on until the air temperature matches the set point, whereas if the current temperature is higher than the set point, the air conditioning would come on until the air temperature reaches the set point.

Various researchers and theorists (e.g., Carver & Scheier, 1982, 1998, 2012; DeYoung, 2015; Quirin et al., 2020) have presented models of personality in which such set-point, negative feedback loop mechanisms are responsible for the regulation of behavior. For example, in a series of papers and books, Carver and Scheier have presented a detailed account of the self-regulation of behavior and personality, based heavily on ideas from cybernetic control theory, especially goal-corrected, discrepancy-reducing feedback systems. DeYoung (2015) has presented his Cybernetic Big Five model in which he argues that the various Big Five traits map onto various aspects of such a cybernetic control system. More recently, Quirin and a number of colleagues (2020) have presented 20 tenets for uncovering the causal mechanisms of personality in which a central tenet is the idea that discrepancy-reducing feedback control loops are responsible for the regulation of behavior. Various other models could also be cited. In fact, such set-point, discrepancy-reducing models are probably the most prevalent mechanism used in models of self-regulation. However, in addition to our work, several other researchers have proposed models of personality that rely on settling point (equilibrium of forces) mechanisms (Kuhl, 2000; Quirin et al., 2019; Quirin & Kuhl, 2022; Revelle & Condon, 2015).

As noted above, many researchers studying physiological regulation and the control of motivated behavior have concluded that negative feedback loop, set-point models are not sufficient as models of behavioral regulation (for reviews see Berridge, 2004; Ramsay & Woods, 2014). They have argued that two other processes are more important: (1) settling point or balance of forces and (2) predictive models. There are at least two major reasons for this conclusion: (1) It has been difficult, if not impossible, to identify specific set-point representations, or biological mechanisms that regulate behavior by using a set point, and (2) there is considerable evidence for settling point and predictive mechanisms in the regulation of behavior. So, even if one does not accept the conclusion that set-point mechanisms are nonexistent, as some researchers do (Berridge, 2004), there is considerable evidence that settling point or balance of forces mechanisms and predictive mechanisms play a central role in the regulation of behavior.

A useful way to think about the mechanisms that underlie both settling point and predictive processes is that they consist of sensor–effector systems (Ramsay & Woods, 2014), in which there is a direct link between a "detector" or sensor that responds to its target input and an effector system, such as a behavioral system or a physiological system, that takes some action in response to the signal from the sensor. There is no intervening discrepancy detection mechanism, rather the effector system responds directly to the sensor signal. One example of this would be a fire alarm, where if the sensor detects sufficient levels of smoke, it sends a signal that triggers the alarm. Settling point or equilibrium of forces models operate when two or more of these systems interact to control behavior, for example, consider eating behavior. Hormones in the bloodstream that signal hunger will, through intermediate systems, such as the hypothalamus or brain stem, elicit eating, whereas different hormones that signal satiation will act to inhibit eating. The point at which the organism stops eating is the result of the specific balance among these systems. So, settling point behavior can be seen as the result of such interacting systems, sometimes through competition and sometimes through joint action.

Predictive modeling can also be seen as being based on this kind of sensor-effector system, where the sensor generates a behavior based on a predicted future state. This prediction may be hard-wired or it may be learned. For example, there is considerable evidence (e.g., Berridge, 2004; Ramsay & Woods, 2014) that the motor acts of eating and swallowing food will reduce feelings of hunger well before there is a discernible change in the bloodstream. Presumably, this is to prevent the organism from eating too much before it receives a satiety signal from the bloodstream. Or in our own lives, we typically prepare and eat meals at set times, rather than waiting until we are hungry, because we have developed models of when we will be hungry.

Such sensor–effector systems are central to both settling point mechanisms and predictive mechanisms. Rather than having behavior be driven by a discrepancy from a set point in a goal-corrected system, the systems or behavior are directly triggered by a signal from the environment or from some internal bodily state.

2 | A NEURAL NETWORK MODEL OF PERSONALITY AND MOTIVATION

Our recent model of personality (e.g., Read et al., 2010, 2017, 2021) has been implemented as a neural network

14676494, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jopy.12815 by SEA ORCHID (Thailand), Wiley Online Library on [09/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

model that provides an explicit, instantiated account of how such settling point and predictive processes can account for many aspects of the regulation of personality-related behavior. This model has been implemented in the *emergent*/Leabra neural network architecture (O'Reilly et al., 2020). We first describe our model and then describe how different mechanisms can account for these two mechanisms of regulation.

In our model, we argue that personality-related behavior (and much, if not all behavior) can be viewed as the result of a process in which there are organized motivational systems, consisting of multiple motives, where the motives compete for the control of behavior. A visual depiction of the structure of the model is in Figures 1 and 2. Figure 1 is a schematic depiction of the model, and Figure 2 is one implementation that we have done to model the everyday behavior of a college student (Read et al., 2017, 2021). There are two primary motivational systems, an Approach system that governs sensitivity to reward and an Avoidance system that governs sensitivity to threat and punishment. Within each system, are a number of more specific motive systems, that govern both social motives (e.g., affiliation and dominance) and nonsocial motives (e.g., hunger and thirst).

Following Berridge et al.'s (2009) and Berridge and Robinson (1998) work, the degree of activation of a motive represents the degree of "wanting" for that reward. (This is also sometimes referred to as the incentive salience of the corresponding reward). For example, the strength of activation of the WFood node represents how much you "want" Food, how strongly motivated you are, and the strength of the Waff node represents how much you want social affiliation.

Berridge et al. (2009) and Zhang et al. (2009) proposed that "wanting" or incentive salience is a multiplicative function of the strength of a reward cue (e.g., ice cream) and the strength of the corresponding interoceptive state (e.g., hunger). (See Smith & Read, 2022 for an explicit mathematical model for calculating "wanting" or incentive salience). For example, "wanting" for a particular food object is a multiplicative function of how attractive that food is (e.g., broccoli versus ice cream) and how hungry you are.

So, the primary factors that influence the activations of the motives, where degree of activation governs which motives come to control behavior at any moment, are:

- Situational cues to affordances for the pursuit of motives/goals, where the strength of the relationship between the cues and the motives, are a function of both hard-wired biological factors, and learning from experience about the relationship between cues and the probability of the availability of motive relevant outcomes. The strength of the weights between situational cues and motives represents the degree of influence of the situational cue on the activation of the motive and can be modified by learning.
- Interoceptive cues about current need states, such as hunger, need for affiliation, need for achievement. The strength of the weights between interoceptive cues and motives represents the degree of influence of the interoceptive cue on the activation of the motive, and can also be modified by learning.

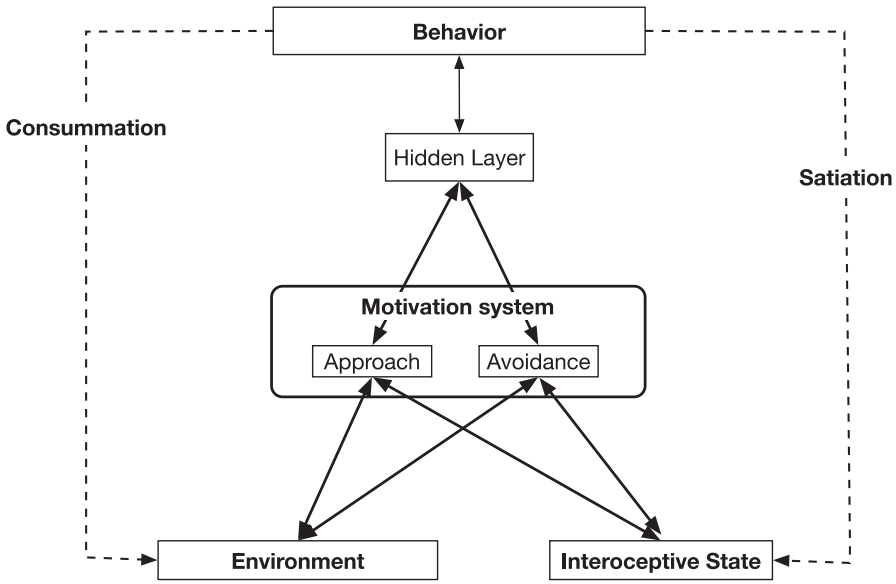

FIGURE 1 Structure of the neural network model.

14676494, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jopy.12815 by SEA ORCHID (Thailand), Wiley Online Library on [09/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

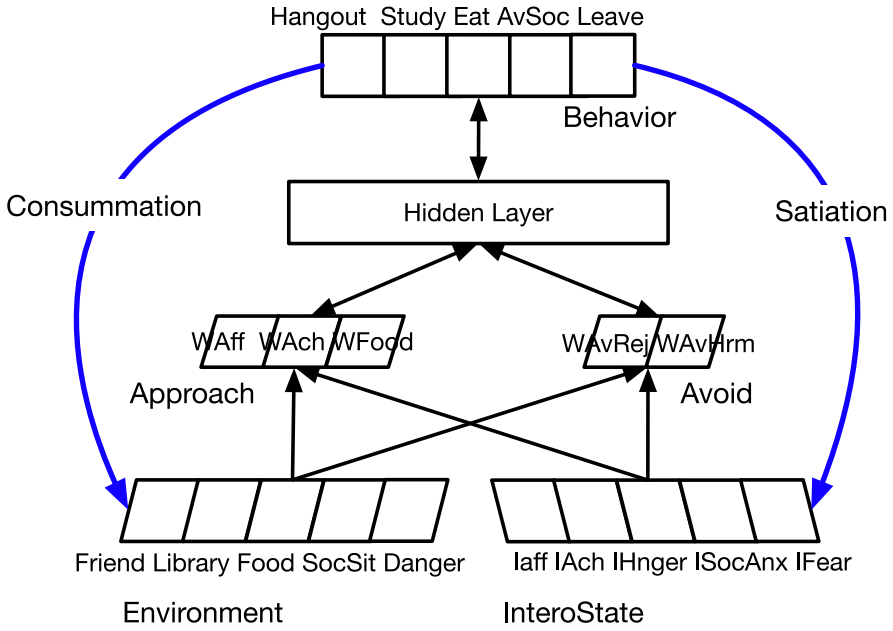

FIGURE 2 Model of a college student.

As noted above, the activation of the relevant node (which indicates strength of "wanting") is a multiplicative function of the relevant situational and interoceptive cues.

In addition, the degree of activation of a motive, its "wanting," is also influenced by:

- Sensitivity of the Approach and Avoidance systems to situational and interoceptive cues. Differences in sensitivity can be a result of both biological factors and experience. This sensitivity is essentially a gain factor that represents how much each unit of input influences the strength of the output activation. In other work (e.g., Read et al., 2010), we have noted how this differential sensitivity of the Approach and Avoidance systems can be viewed as capturing the broad personality traits of Extraversion and Introversion.
- Baseline activation of the different motives, which represents individual differences in the importance of different motives. This essentially represents the default or chronic importance of the motive.

Once motives are activated, motives within each of the Approach and Avoidance layers compete with each other, and a subset of the nodes in each layer "win" the competition. So, it would be possible to have both Approach and Avoidance nodes active at the same time and sending activation to behaviors. The nodes in the Approach and Avoid systems then send activation to the behavior layer, where the behavior nodes sum the incoming activation and compete to see which behavior is enacted. The most activated behavior will be enacted. The behavior can then change situational affordances in the world (e.g., eating can reduce the amount of food available) and the interoceptive state of the organism (e.g., eating can lead to eventual satiety). The new state of the world *and* the new internal state feed back into the environment and InteroState layers, so that any changes in the situational affordances and the interoceptive state will lead to changes in the activation of motives. As in the real world, this cycle continues.

In this model, multiple regulatory processes play out. The motives/goals in this model play a central role in regulatory processes, and they provide the consistency, stability, or coherence of behavior/personality over time. For example, individual differences in the chronic importance of motives, represented in the model by chronic differences in the baseline activity of individual motives in the Approach/Avoidance layers, provides a degree of stability in the models. Similarly, stable individual differences in the sensitivity/gain of Approach and Avoidance systems play a central role in the behavior of the motives and thus in the stability, consistency, and coherence of the model.

The interaction of people's goals/motives with the situations in which they act also play a role in stability and coherence. People's goals/motives lead them to choose situations that have affordances that enable the pursuit of those goals and motives. And once people are in a situation, the affordances of that situation make the situation "sticky." First, the situation is "sticky" because the availability of affordances will increase their wanting for and pursuit of the rewards in that situation. Second, once one is in a situation, one will pursue the goals afforded in that situation until they have made sufficient progress. Third, a situation is "sticky" because once in a situation it often takes at least some effort to move to

14676494, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jopy.12815 by SEA ORCHID (Thailand), Wiley Online Library on [09/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

another situation. To the extent that people tend to stay in the same or similar environments, their behavior will be more stable, as long as the situation itself is relatively stable.

2.1 | Key regulatory mechanisms in our model

Two of the self-regulatory mechanisms we discussed earlier are important in this model. One mechanism is a settling point or balance of forces system, in which regulation is the result of the interaction of multiple forces. A second mechanism is predictive modeling, in which organisms use models of the world to predict future states and modify their behavior accordingly. The models of the world can either be innate or learned. In our model, set-point, discrepancy-reducing mechanisms do not play a role.

2.1.1 | Settling point, equilibrium of forces

We first discuss how the regulatory processes in our model function through a process of interacting forces. Motives in our model compete for the control of behavior, and the winner is a function of the *relative* strengths of the different motives.[2] Each of the motives can be viewed as a different interacting force with a particular strength or level of activation. One implication of this conceptualization is that which motive controls behavior is not just a function of its own activation, and it is also a function of the activation of its competitors. So, under one set of circumstances, a motive with a moderate level of activation may control behavior, whereas in a different set of circumstances that same motive with the same level of activation may be outcompeted by other motives. There is no simple correspondence between the performance of a behavior and the strength of a corresponding motive. For example, if an organism eats, it may be that they are moderately hungry with weak competing motives or very hungry with more strongly competing motives (or very hungry with weakly competing forces). Note that the relative strength of these two forces often changes over time. For example, how long an individual talks to a friend may be a function of the current strength of the affiliative motive, how interesting or boring the conversation is or becomes, and the need to study for an exam tomorrow. When the individual stops interacting with their friend and goes to study is a function of the interaction among these different forces.

Two other examples of these kinds of interacting forces come from work on eating behavior. First, considerable research shows that rats will eat less when obtaining and eating the food takes more effort (a countervailing force) and will eat more when the food is easier to get. Second, food intake is greater for attractive foods; the forces toward eating are stronger for attractive foods.

Something similar to an interacting forces analysis has been provided for the trait of impulsivity. Various theorists (Depue & Collins, 1999; Zuckerman, 2005) have argued that impulsivity is not a unidimensional construct, but is rather the result of the interaction of at least two competing forces: appetitive or approach behaviors, which move the individual toward a goal, and aversive or self-control forces, which inhibit the expression of such approach behaviors. Thus, impulsivity may be the result of a number of different combinations of these factors. For example, we can get impulsive behavior in an individual if they have excessively strong approach behavior in the face of a strong countervailing inhibitory force, or they have modest approach behavior with a weak countervailing inhibitory force. Other possible combinations of levels of the two forces can also result in impulsive behavior. A similar analysis could be provided for a nonimpulsive or controlled person. Thus, it is the balance of forces that is important, not their absolute levels.

This model does not need an explicit discrepancy-reducing, goal-corrected process to get consistency, stability, or coherence in personality. We can think of the system as moving toward a settling point, which is the result of the interaction of multiple forces. That is, competing motives and competing cues will determine a settling point, toward which the organism moves as the result of the competing forces. And if the individual moves away from this settling point, the competing forces will push it back.

Another important factor in our model that can drive settling point behavior is interoceptive state information. Interoceptive state can be thought of as information from sensors that provide signals about the state of the body. Many, if not all, interoceptive state signals are not driven by discrepancy, but instead are driven by sensors that are sensitive to specific biological signals. For example, for hunger or thirst, there are specific hormones that signal these and neural systems in the brain that are sensitive to these hormones (Berridge, 2004; Ramsay & Woods, 2014). There are no explicit discrepancy signals. Think about the smoke detector we mentioned earlier. As the amount of smoke goes up a sensor sends out an increasingly strong signal and when it crosses threshold, it sets off the alarm. But it does not operate in terms of reducing a discrepancy. Or think about the retina. The cells in the retina send out a signal that corresponds to the strength of the light.

Interoceptive state may also be signaled by the actions of motor systems. For example, the act of drinking and swallowing water helps regulate feelings of thirst (Berridge, 2004). Similarly, the act of chewing

14676494, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jopy.12815 by SEA ORCHID (Thailand), Wiley Online Library on [09/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and swallowing food helps regulate feelings of satiety (Berridge, 2004).

Important examples of effects of shifts in equilibrium of forces

Dopamine activity plays a central role in reward-seeking behavior and shifts in dopamine activity can lead to shifts in behavior. For example, degree of Extraversion (for which reward seeking is a major component) is strongly related to baseline or tonic dopamine activity in the ventral striatum (nucleus accumbens) (Depue & Collins, 1999; DeYoung, 2015). Depression (which involves reduced reward seeking) and drug addiction can both be understood, at least partially, in terms of downregulation of dopamine activity in the nucleus accumbens, which governs reward seeking or "wanting" behavior (Berridge, 2004). Below, we discuss the role of shifting of forces in both.

Stress and depression. Short-term stress leads to temporarily increased dopaminergic activity (Grace, 2016). However, long-term, chronic stress leads to overactivation of the dopamine system, which activates a countervailing system that seeks to downregulate dopamine activity. When chronic stress no longer exists, the overactivation should dissipate and the countervailing force should diminish. However, there is evidence that if the stress occurs for a long enough time, there is a permanent shift in the strength of the countervailing force so that activity in this dopaminergic system is continually inhibited below tonic or baseline activity (Grace, 2016). There is evidence that this is one reason the aftermath of chronic stress often leads to depression. This pattern could be explained in terms of a balance of forces. The basic idea is that in normal conditions, there are countervailing forces, one of which seeks to increase dopamine activity and another to downregulate it to keep dopamine activity within a reasonable range. If the countervailing force becomes overactivated for long enough because of chronic excessive dopamine activity and then may permanently strengthen so that when the excessive dopamine activity dissipates, the countervailing force will inhibit tonic dopamine activity below its previous baseline. Essentially, the system has developed a new settling point.

Drug addiction. Koob and Le Moal (2001, 2008) provide an analysis of drug addiction that draws on this idea of an equilibrium between opposing forces. They suggest that taking a drug, such as methamphetamine (METH), will strongly activate the reward system and that there are brain anti-reward systems that will act to downregulate activity in the reward system. However, with long-term drug use this anti-reward system may become chronically overactivated so that in the absence of drug taking, the reward system will be sufficiently inhibited to produce a negative hedonic state. Unfortunately, this may lead to further drug taking, in an attempt to increase the activation of the reward system, and this activation of the reward system will then further activate the anti-reward system. They argue (Ramsay & Woods, 2014) that this can lead to a negative spiral of drug addiction.

This model can be mapped onto the behavior of the dopaminergic system in the nucleus accumbens (ventral striatum), which is central to reward processing. Reward seeking partially depends on tonic dopamine activity in the nucleus accumbens. Taking a drug like METH, which is both a dopamine agonist and a reuptake inhibitor (Fleckenstein et al., 2007; Panenka et al., 2013; Sulzer et al., 2005), increases tonic dopamine activity (reward system activity in Koob and Le Moal's account). Occasional use of METH only temporarily increases tonic dopamine activity and when METH is not being taken, dopamine activity goes back to baseline or its regular level. However, chronic taking of METH can over-activate the "anti-reward" system, which chronically inhibits or downregulates tonic dopamine activity. In an attempt to return dopamine activity to its normal tonic or baseline level, an individual may take more METH, which then further activates the countervailing system and leads to drug addiction.

It is worth noting that withdrawal from chronic METH use typically leads to at least short-term depression. This is consistent with the argument above that chronic METH use leads to chronic downregulation of tonic dopamine activity. Since tonic dopamine activity plays an important role in motivation and "wanting," then downregulation of dopamine activity would lead to decreases in motivation and "wanting," the central characteristic of serious depression.

2.1.2 | Predictive models

A second major mechanism for regulation is predictive models.[3] The basic idea is that the brain learns models of the world and uses them to predict future outcomes. For example, most people eat meals at a particular time because they have learned that this behavior allows them to proactively deal with low blood sugar or hunger before any real discrepancy develops. As Ramsay and Woods (2014, p. 229) note, "Meals are initiated at times dictated by habit, time of day, convenience, opportunity, and other factors as opposed to being reactions to deficits of available energy ...".

For example, I know that if I don't eat lunch by around Noon, then by 1 PM, I start to get irritable, so to proactively keep me from entering that state and to keep my

14676494, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jopy.12815 by SEA ORCHID (Thailand), Wiley Online Library on [09/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

blood glucose at a reasonable level I have learned to eat around Noon. Moreover, I don't simply eat until I am full; I make or buy a set amount of food, based on past experience, and then eat that amount of food. Other examples of such predictive mechanisms can be seen in the regulation of drinking and eating behavior. Considerable evidence shows that the experience of swallowing water helps to downregulate feelings of thirst before there are actual changes in the hormonal signaling and the experience of chewing and swallowing food downregulates hunger before there are any changes in the hormones in the bloodstream that trigger feelings of hunger (Berridge, 2004; Ramsay & Woods, 2014).

The neural network architecture we use, *emergent*/Leabra (O'Reilly et al., 2020), has both Hebbian (associative or correlational learning) and error-correcting learning mechanisms in the architecture. Thus, our model, as is true of any neural network architecture with learning, can learn to make predictions. We use this learning mechanism to teach our model about characteristics of its environment and its desired behavior.

Using these learning mechanisms, we have implemented two kinds of learning as a core part of the model. First, the model has Pavlovian learning. The current view of Pavlovian learning (Cartoni et al., 2015; Dayan & Berridge, 2014; Dickinson & Balleine, 2002; Kornoski, 1967; Smith & Read, 2022) is that when an animal is presented with a Neutral stimulus, like a bell, that predicts the appearance of a reward, like food, that the animal learns to associate the cue with both an unconditioned response (e.g., salivating) and a representation of the features of the reward. So, after learning, presentation of the cue predicts the likelihood of the appearance of the reward and activates a representation of the reward. In our model, we train the model to associate cues in the environment (e.g., friend) and a particular motive (reward representation) (e.g., affiliation). (It would be relatively straightforward to add the capability for the model to learn much more sophisticated models of the environment and more sophisticated representations of possible rewards, but it would not change the basic point we are making here). Thus, the model learns that cues in the environment about affordances of rewards, plus information about current bodily or interoceptive state, can jointly trigger a representation of that reward. Essentially, the system is learning to predict the possible occurrence of the reward (or punishment).

Second, the model has instrumental learning. We specifically train our model that specific motives (e.g., affiliation) are associated with specific behaviors (e.g., hang out) that can satisfy that motive (achieve a reward). Since our model has bi-directional connections between the motive layer and the behavior layer, it learns that if it has a particular motive, then it should enact an associated behavior and it also learns that instrumental actions achieve a reward. Another way to think of this is that the system is learning to predict that if they enact a particular instrumental action, then that will result in the receipt of a reward.

Together Pavlovian conditioning and instrumental conditioning form a process called Pavlovian to instrumental transfer (PIT). The key idea is that Pavlovian conditioning or learning enables us to use cues in the environment to predict the likelihood that rewards of specific types will be available. And once we predict the likelihood of those rewards, the representation of the reward will activate an instrumental action that allows us to pursue or attain that reward. This is a key type of predictive modeling that results in models of the world that allow us to use cues in the environment to trigger (activate) instrumental actions for pursuing a reward.

Predictive modeling in other models of personality

Safron and DeYoung (2021) have recently argued that Friston's (2010) free energy principle and related predictive coding approaches can be used to provide an account of the mechanisms in DeYoung's (2015) Cybernetic Big Five model. DeYoung (2015) argued that the structure and dynamics of the Big Five could be explained in terms of the operation of various components of a cybernetic control system. However, it is not clear whether Safron and DeYoung are arguing that the FEP is a way to instantiate the set-point, discrepancy-reducing control systems in the Cybernetic Big Five model or whether they see predictive coding as an additional regulatory mechanism. (As we noted in Footnote 1, predictive coding approaches go beyond the idea that the brain learns predictions based on the statistical structure of the environment).

Predictive coding and set-point, discrepancy-reducing models are very different mechanisms. They differ in at least two important ways. First, predictive models drive proactive responses to an input. Their role is to allow the organism to predict the future and to engage in either internal or external actions that proactively address potential future outcomes or events. Predictive coding mechanisms are trying to address a potential or experienced discrepancy (surprisal) before it happens or happens again (Friston, 2013). By contrast, set-point, discrepancy-reducing processes are reactive, they are responding to something that has already happened, and they are trying to address any discrepancy *after* it happens. Second, although both use error signals, they use them for very different purposes. Predictive coding mechanisms use error signals to drive learning, to learn new models of the world and update old models. It is

14676494, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jopy.12815 by SEA ORCHID (Thailand), Wiley Online Library on [09/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

all about building more accurate models of the world, so that the brain (organism) can make more accurate predictions about the future to avoid discrepancy or surprisal. It is the representations and the predictions they enable that guide behavior. By contrast, a set-point mechanism directly uses error signals to guide current behavior. (Although it is possible to have a separate learning mechanism that could use error signals to update a set-point value).

As a reviewer of this paper noted, Friston has explicitly argued that his variational free energy principle could be used to understand both predictive mechanisms and set-point, discrepancy-reducing mechanisms. However, Safron and DeYoung (2021) have not been clear about whether their focus is on how Friston's free energy principle can account for the cybernetic control mechanisms in DeYoung's cybernetic Big Five model or whether they are interested in how the free energy principle can also be used for predictive modeling. Perhaps they are interested in both, but which is not clear.

2.2 | Explaining why satiety often results in an aversive or avoidant response

One last question we would like to address is how a settling point or predictive modeling mechanism can account for the aversive or avoidance response that frequently occurs when an organism becomes completely satiated. We identify at least four different ways in which this might occur. Some of these ways are specific to a particular kind of stimulus, such as food or water, whereas others are more general.

First, for certain bodily needs, such as for food and water, there are specific signals, such as feelings of overfullness or stomach distention, that provide an aversive or avoidance signal that tells the organism that it has ingested more than enough of the stimulus. Thus, we have two competing signals: an approach signal that leads to approaching and ingesting the food or water, and one which is inherently aversive and which is activated by sufficient ingestion of the food or water. Here, the strength of the aversive signal increases as a direct result of the ingestion of the food or water, while at the same time, the approach response weakens(satiation) with the ingestion of food or water.

A second situation occurs when there is no increase in the aversive response because of the approach behavior (such as eating a lot of food), but there is instead a constant aversive signal, which is initially overweighed by the approach signal. For example, many activities have both positive and negative attributes. When we start working on the activity, it may be that its positive attributes activate an Approach "wanting" signal that initially is far stronger than the Avoidance signal activated by the negative attributes. However, as the needs related to the positive attributes are satiated, the impact of the negative attributes and the resulting avoidance motive will be far stronger than the impact of the positive attributes.

A third situation might occur when performing a rewarding activity requires a reasonable amount of effort, which is typically inherently aversive. This has the same form as the previous example, although it is not identical. As the organism becomes satiated and the Approach signal (or "wanting") decreases, while the degree of effort required is constant, the aversiveness of the effort required will overweigh the attractiveness of the activity.

A fourth possibility is that the lack of incentive salience or "wanting" is inherently aversive. Satiation results in strongly reduced or even minimal levels of incentive salience or "wanting." This could be perceived as boredom or a lack of interest. (Mom! I'm bored! There's nothing to do!). Our claim is that many (most?) organisms have a need for some level of incentive salience or feelings of interest and that lack of incentive salience/interest is aversive and motivates behavior to increase interest. One possible source of this aversive signal might be reduced activity in the dopaminergic reward system. The basic idea is that there is a neural system, perhaps in the insula, that is sensitive to the level of incentive salience or activation of the dopamine system in the Nucleus accumbens. There is considerable evidence (Berridge, 2004; Berridge et al., 2009; Berridge & Robinson, 1998) that incentive salience, "wanting," or interest in a stimulus are tied to dopamine activity in the ventral striatum/nucleus accumbens. Lower levels of dopaminergic activity cause lower "wanting."[4]

Some researchers have argued that the boredom associated with lower degrees of motivation or interest can be viewed as a negative emotion that signals or motivates the individual to engage in the pursuit of new goals. For example, Bench and Lench (2013), in line with much emotion theory, note that emotions often tell us about the state of our goals and provide information about what to do next. Boredom tells us that a goal has been satisfied or is no longer interesting, and it is time to move to something else. They argue that boredom is a negative emotion that motivates movement to new goals. There is some evidence from work on drug abuse and depression (e.g., Brankovi, 2015; Krotava & Todman, 2014; Todman, 2003, 2007) that is consistent with the idea that reduced dopaminergic activity in the NACC is related to feelings of boredom and that these feelings of boredom predict substance use and depression.

14676494, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jopy.12815 by SEA ORCHID (Thailand), Wiley Online Library on [09/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3 | SUMMARY AND CONCLUSION

As we noted above, many personality theorists have proposed that personality and behavioral regulation can be understood in terms of goal-corrected, discrepancy-reducing systems. However, we argue that two other mechanisms play a key role in personality regulation: a settling point, equilibrium of forces mechanism, and predictive modeling. Over the last 15 years, we have developed an explicit computational model of personality and motivated behavior in which these two regulatory processes play a central role. The current paper has outlined how these two mechanisms function.

Furthermore, this model demonstrates how behavior can be successfully regulated without the use of goal-corrected mechanisms. As we discussed earlier, Ramsay and Woods (2014) have argued that current models of physiological regulation are less likely to draw on set-point, discrepancy-reducing mechanisms due to a failure by researchers to identify biological instantiations of such systems. Although we are not arguing against a role for goal-corrected, discrepancy-reducing systems in behavioral regulation, we think theorists should take seriously the possibility that, in line with the considerable amount of work on the regulation of physiological systems (e.g., Berridge, 2004; Ramsay & Woods, 2014) that goal-corrected discrepancy-reducing systems have, at best, a limited role in behavioral regulation. Other mechanisms, such as equilibrium of forces and predictive modeling seem to play a far greater role, and we have presented a detailed model of how that would operate in the regulation of motivation and personality-related behavior.

ACKNOWLEDGEMENTS

None.

AUTHOR CONTRIBUTIONS

The conceptual framework and argument was jointly developed by SJR and LCM. The writing was primarily done by SJR.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest.

ETHICS STATEMENT

The paper does not involve human subjects.

PERMISSIONS

No material is reproduced from another source.

ORCID

Stephen J. Read https://orcid.org/0000-0003-4562-7248

ENDNOTES

[1] It is important to distinguish between what we mean here by predictive models and the general idea of predictive coding (Brown et al., 2011; Clark, 2013; Friston, 2010; Rao & Ballard, 1999). We are making much more limited claims than are made in predictive coding models: We are simply claiming that the brain and the kinds of neural networks we are presenting can learn the structure of the environment and learn to make predictions from their inputs. Predictive coding models make further claims that the brain uses top-down models of the world to guide learning about the structure of the environment. These further assumptions are not necessary for the claims we make here.

[2] Of course, things can get more complicated. Imagine 3 goals (A, B, and C). A has the highest activation among all three goals, but A activates behavior p and goals B and C both activate behavior q leading q to have a higher level of activation than p. Therefore, in this example, the most strongly activated goal does not outcompete the combined effects of two other goals, where those two other goals activate the same behavior.

[3] As we noted earlier in the paper, we are distinguishing between predictive models and predictive coding models, and are only focusing on the ability to make predictions about the future, which is not unique to predictive coding models.

[4] One way this might work is whether there is a neural system that is positively driven by the average activation of the Approach system in the NACC, and this neural system has an inhibitory influence on a chronically active system that signals the negative emotion of boredom, perhaps in the insula. When the Approach system is strongly activated, this neural system will inhibit the negative emotion signal. However, when the average activation of the Approach system is low enough, the negative emotion signal will be released from inhibition and will signal the negative emotional state.

REFERENCES

Bench, S., & Lench, H. (2013). On the function of boredom. *Behavioral Sciences*, *3*(3), 459–472. https://doi.org/10.3390/bs3030459

Berridge, K. C. (2004). Motivation concepts in behavioral neuroscience. *Physiology & Behavior*, *81*(2), 179–209. https://doi.org/10.1016/j.physbeh.2004.02.004

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, *28*, 309–369. https://doi.org/10.1016/S0165-0173(98)00019-8

Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward: 'Liking', 'wanting', and learning. *Current Opinion in Pharmacology*, *9*(1), 65–73.

Brankovi, S. (2015). Boredom, dopamine, and the thrill of psychosis: Psychiatry in a new key. *Psychiatria Danubina*, *27*(2), 126–137.

Brown, H., Friston, K., & Bestmann, S. (2011). Active inference, attention, and motor preparation. *Frontiers in Psychology*, *2*, 218. https://doi.org/10.3389/fpsyg.2011.00218

Cartoni, E., Moretta, T., Puglisi-Allegra, S., Cabib, S., & Baldassarre, G. (2015). The relationship between specific Pavlovian instrumental transfer and instrumental reward probability. *Frontiers in Psychology*, *6*, 1697. https://doi.org/10.3389/fpsyg.2015.01697

14676494, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jopy.12815 by SEA ORCHID (Thailand), Wiley Online Library on [09/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological Bulletin*, *92*(1), 111–135. https://doi.org/10.1037/0033-2909.92.1.111

Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. Cambridge University Press.

Carver, C. S., & Scheier, M. F. (2012). Cybernetic control processes and the self-regulation of behavior. In R. M. Ryan (Ed.), *The Oxford handbook of human motivation*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399820.013.0003

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

Dayan, P., & Berridge, K. C. (2014). Model-based and model-free Pavlovian reward learning: Revaluation, revision, and revelation. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(2), 473–492. https://doi.org/10.3758/s13415-014-0277-8

Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and Brain Sciences*, *22*(3), 491–517. https://doi.org/10.1017/S0140525X99002046

DeYoung, C. G. (2015). Cybernetic big five theory. *Journal of Research in Personality*, *56*, 33–58. https://doi.org/10.1016/j.jrp.2014.07.004

Dickinson, A., & Balleine, B. (2002). The role of learning in the operation of motivational systems. In H. Pashler (Ed.), *Stevens' handbook of experimental psychology* (pas0312). John Wiley & Sons, Inc. https://doi.org/10.1002/0471214426.pas0312

Fleckenstein, A. E., Volz, T. J., Riddle, E. L., Gibb, J. W., & Hanson, G. R. (2007). New insights into the mechanism of action of amphetamines. *Annual Review of Pharmacology and Toxicology*, *47*, 681–698.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. (2013). Active inference and free energy. *Behavioral and Brain Sciences*, *36*(3), 212–213. https://doi.org/10.1017/S0140525X12002142

Grace, A. A. (2016). Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. *Nature Reviews Neuroscience*, *17*(8), 524–532. https://doi.org/10.1038/nrn.2016.57

Koob, G. F., & Le Moal, M. (2001). Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology*, *24*(278), 97–129.

Koob, G. F., & Le Moal, M. (2008). Addiction and the brain antireward system. *Annual Review of Psychology*, *59*(1), 29–53. https://doi.org/10.1146/annurev.psych.59.103006.093548

Kornoski, J. (1967). *Integrative activity of the brain; an interdisciplinary approach*. University of Chicago Press.

Krotava, I., & Todman, M. (2014). Boredom severity, depression and alcohol consumption in Belarus. *Journal of Psychology and Behavioral Science*, *2*(1), 73–83.

Kuhl, J. (2000). A functional-design approach to motivation and self-regulation: The dynamics of personality systems and interactions. In M. Boekaerts & P. R. Pintrich (Eds.), *Handbook of self-regulation* (pp. 111–169). Academic Press.

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., & Contributors. (2020). *Computational cognitive neuroscience* (4th ed.). Online Book. https://github.com/CompCogNeuro/ed4

Panenka, W. J., Procyshyn, R. M., Lecomte, T., MacEwan, G. W., Flynn, S. W., Honer, W. G., & Barr, A. M. (2013). Methamphetamine use: A comprehensive review of molecular, preclinical and clinical findings. *Drug and Alcohol Dependence*, *129*(3), 167–179. https://doi.org/10.1016/j.drugalcdep.2012.11.016

Quirin, M., & Kuhl, J. (2022). The concert of personality: Explaining personality functioning and coherence by personality systems interactions. *European Journal of Personality*, *36*(3), 274–292. https://doi.org/10.1177/08902070221078478

Quirin, M., Robinson, M. D., Rauthmann, J. F., Kuhl, J., Read, S. J., Tops, M., & DeYoung, C. G. (2020). The dynamics of personality approach (DPA): 20 tenets for uncovering the causal mechanisms of personality. *European Journal of Personality*, *34*(6), 947–968. https://doi.org/10.1002/per.2295

Quirin, M., Tops, M., & Kuhl, J. (2019). Autonomous motivation, internalization, and the integrative self: A self-regulation framework of interacting neuropsychological systems. In R. M. Ryan (Ed.), *The Oxford handbook of human motivation* (pp. 393–413). Oxford University Press.

Ramsay, D. S., & Woods, S. C. (2014). Clarifying the roles of homeostasis and allostasis in physiological regulation. *Psychological Review*, *121*(2), 225–247. https://doi.org/10.1037/a0035942

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Read, S. J., Brown, A. D., Wang, P., & Miller, L. C. (2021). Neural networks and virtual personalities: Capturing the structure and dynamics of personality. In J. F. Rauthmann (Ed.), *The handbook of personality dynamics and processes*. Elsevier.

Read, S. J., Monroe, B. M., Brownstein, A. L., Yang, Y., Chopra, G., & Miller, L. C. (2010). A neural network model of the structure and dynamics of human personality. *Psychological Review*, *117*(1), 61–92. https://doi.org/10.1037/a0018131

Read, S. J., Smith, B., Droutman, V., & Miller, L. C. (2017). Virtual personalities: Using computational modeling to understand within-person variability. *Journal of Research in Personality*, *69*, 237–249. https://doi.org/10.1016/j.jrp.2016.10.005

Revelle, W., & Condon, D. M. (2015). A model for personality at three levels. *Journal of Research in Personality*, *56*, 70–81. https://doi.org/10.1016/j.jrp.2014.12.006

Safron, A., & DeYoung, C. G. (2021). Integrating cybernetic Big Five theory with the free energy principle: A new strategy for modeling personalities as complex systems. In D. Wood, S. J. Read, P. D. Harms, & A. Slaughter (Eds.), *Measuring and modeling persons and situations* (pp. 617–649). Academic Press.

Smith, B. J., & Read, S. J. (2022). Modeling incentive salience in Pavlovian learning more parsimoniously using a multiple attribute model. *Cognitive, Affective, & Behavioral Neuroscience*, *22*(2), 244–257. https://doi.org/10.3758/s13415-021-00953-2

Sulzer, D., Sonders, M. S., Poulsen, N. W., & Galli, A. (2005). Mechanisms of neurotransmitter release by amphetamines: A review. *Progress in Neurobiology*, *75*(6), 406–433. https://doi.org/10.1016/j.pneurobio.2005.04.003

Todman, M. (2003). Boredom and psychotic disorders: Cognitive and motivational issues. *Psychiatry: Interpersonal and Biological Processes*, *66*(2), 146–167. https://doi.org/10.1521/psyc.66.2.146.20623

Todman, M. (2007). Psychopathology and boredom: A neglected association. In K. A. Fanti (Ed.), *Psychological science: Research,*

14676494, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jopy.12815 by SEA ORCHID (Thailand), Wiley Online Library on [09/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

theory and future directions (pp. 145–170). Athens Institute for Education and Research.

Zhang, J., Berridge, K. C., Tindell, A. J., Smith, K. S., & Aldridge, J. W. (2009). A neural computational model of incentive salience. *PLoS Computational Biology*, *5*(7), e1000437. https://doi.org/10.1371/journal.pcbi.1000437

Zuckerman, M. (2005). *Psychobiology of personality* (2nd ed.). Cambridge University Press.

How to cite this article: Read, S. J., & Miller, L. C. (2023). Behavioral regulation relies on interacting forces and predictive models. *Journal of Personality*, *91*, 917–927. https://doi.org/10.1111/jopy.12815

14676494, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jopy.12815 by SEA ORCHID (Thailand), Wiley Online Library on [09/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License