Personality Neuroscience

cambridge.org/pen

The Virtual Personalities Neural Network Model: Neurobiological Underpinnings

Stephen J. Read, Ashley D. Brown, Peter Wang and Lynn C. Miller

Department of Psychology, University of Southern California, Los Angeles, CA, USA

Review Paper

Cite this article: Read SJ, Brown AD, Wang P, Miller LC. (2018) The Virtual Personalities Neural Network Model: Neurobiological Underpinnings. *Personality Neuroscience*. Vol **1**: e10, 1–11. doi:10.1017/pen.2018.6

Inaugural Invited Paper
Accepted: 11 February 2018

Key words:
neural network models; personality dynamics; personality structure; Motivation

Author for correspondence:
Stephen J. Read, E-mail: read@usc.edu

Abstract

The Virtual Personalities Model is a motive-based neural network model that provides both a psychological model and a computational implementation that explicates the dynamics and often large within-person variability in behavior that arises over time. At the same time the same model can produce—across many virtual personalities—between-subject variability in behavior that when factor analyzed yields familiar personality structure (e.g., the Big Five). First, we describe our personality model and its implementation as a neural network model. Second, we focus on detailing the neurobiological underpinnings of this model. Third, we examine the learning mechanisms, and their biological substrates, as ways that the model gets "wired up," discussing Pavlovian and Instrumental conditioning, Pavlovian to Instrumental transfer, and habits. Finally, we describe the dynamics of how initial differences in propensities (e.g., dopamine functioning), wiring differences due to experience, and other factors could operate together to develop and change personality over time, and how this might be empirically examined. Thus, our goal is to contribute to the rising chorus of voices seeking a more precise neurobiologically based science of the complex dynamics underlying personality.

1. The virtual personalities neural network model

1.1. Neurobiological underpinnings

The science of human personality has many enigmas. How can we understand the dynamics of persons and situations and how they operate together over time to produce emergent behavior? How could the same underlying system that results in between-subject stability in differences in behavior over time and reliable personality structure across individuals (e.g., the Big Five), plausibly explain broad within-person variability in behavior across situations? How do underlying biological mechanisms and different learning histories produce enduring individual differences in response to situational cues? Earlier, we (Read, Droutman, & Miller, 2017; Read et al., 2010; Read, Smith, Droutman, & Miller, 2017) argued that a computational model (here implemented as a neural network model) is needed that allows us to begin to construct plausible models of such dynamics that could begin to address these questions. In the current work, after an introduction that describes our personality model and its implementation as a neural network model, we focus on what is known about the underlying neurobiological and learning mechanisms underpinning the Virtual Personalities Model. We then discuss the implications of these processes for understanding individual differences.

The Virtual Personalities Model (Read, Droutman, & Miller, 2017; Read et al., 2010; Read et al., 2017) is a motive-based neural network model of personality that is both a psychological model and a computational neural network implementation of that model. At a psychological level, the Virtual Personalities Model (Read & Miller, 2002; Read et al., 2010), grew out of a focus on the dynamics of motivational and cognitive structures (Miller & Read, 1991; Read, Jones, & Miller, 1990; Read & Miller, 1989). Thirty years ago, we argued that traits could be viewed as goal-based structures, where the goals of the individual were the central part of a structure consisting of goals, plans, resources, and beliefs (Miller & Read, 1987). Thus, a major basis of individual differences was differences between people in the chronic activation of their goals. The Virtual Personalities Model drew upon that earlier work to examine in more detail how personality could be understood in terms of the behavior of structured motivational systems. It draws on diverse literatures, summarized elsewhere (e.g., Read, Brown, Wang, & Miller, in press; Read et al., 2010), including those involving the factor structure of personality measures (e.g., Eysenck, 1983, 1994; Lee & Ashton, 2004; McCrae & Costa, 1999; Tellegen, & Waller, 2008; Wiggins & Trapnell, 1996; Zuckerman, 2005), the lexical analysis of trait language (e.g., Digman, 1997; Goldberg, 1981), temperament and neurobiological bases of personality (e.g., Clark & Watson, 2008; Gray, 1987a, 1987b; Gray & McNaughton, 2000; Pickering & Gray, 1999; Rothbart & Bates, 1998; Zuckerman, 2005), an evolutionary analysis of social tasks (e.g., Bugental, 2000; Fiske, 1992; Kenrick & Trost, 1997), taxonomies of human motives (Chulef, Read, & Walsh, 2001; Talevich, Read, Walsh, Iyer, & Chopra, 2017), and our

CAMBRIDGE UNIVERSITY PRESS

earlier work (e.g., Miller & Read, 1987, 1991; Read, Jones & Miller, 1990; Read & Miller, 1989) on traits as goal-based structures.

Central to the Virtual Personalities psychological model are two broad motivational systems: a Behavioral Approach System (BAS) and an avoidance system, the latter originally referred to as the Behavioral Inhibition System (BIS) by Gray (e.g., Gray, 1987a, 1991; see also Tellegen, Watson, & Clark, 1999). Although there is very good agreement among researchers regarding the BAS, Gray and others reconceptualized the BIS. They now refer to various fear responses (e.g., flight, fight, freeze) as the Fight Flight Freeze System, or FFFS (not the BIS), and now refer to a separate system —one associated more with anxiety and goal conflict—as the BIS (see Gray & McNaughton, 2000; Smillie, Pickering, & Jackson, 2006). To avoid confusion, we do not use the BAS/BIS specification here, and instead refer to the Approach and Avoidance systems associated with reward (opportunity) and punishment (threat).

These two broad motivational systems have been argued to map most closely to Extraversion (Approach system) and Neuroticism (Avoidance system) and they also are related to the two broad metatraits, Plasticity and Stability (DeYoung, 2015; Digman, 1997) that are found when one investigates higher order factors of the Big Five. Plasticity consists of Extraversion and Openness to Experience/Intellect, whereas Stability consists of Neuroticism, Agreeableness, and Conscientiousness.

Separate specific motives are nested in the Approach (e.g., dominance, social affiliation, mating) and Avoidance (e.g., avoid physical harm, avoid social rejection) systems and are part of the basis of more specific traits. Individuals differ in the baseline sensitivities of the two broad systems, as well as in the baseline activation of the specific goals that are "nested" within the broader motivational systems. Goals in the Approach and Avoidance systems are jointly activated by cues from the Environment that identify the goal affordances of the situation, and by internal cues that indicate the current Interoceptive state of a variety of bodily systems (see Figure 1). General activation of the Approach and Avoidance systems combined with activation of the specific motives then results in the transmission of activity to motor systems that guide behavior to satisfy the active motives. These behaviors result in changes in Interoceptive bodily state (Satiation) and alter the organism's Environment (Consummation).

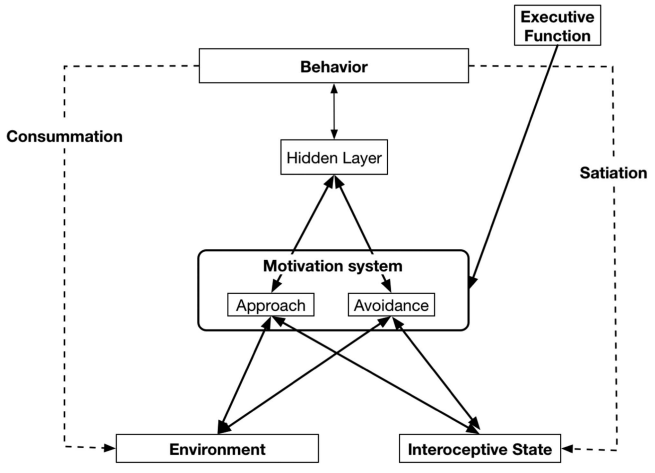

Figure 1. Basic Virtual Personalities Model: Abstract systems that make up the model and the flow of activation between them.

Figure 1 highlights this basic model, without the details of the specific neural network implementation of our Approach and Avoidance motivational systems—but elsewhere see our Virtual Personalities Model (Read et al., in press; Read, Droutman, & Miller, 2017; Read et al., 2010; Read et al., 2017), as well as a detailed tutorial regarding the implementation process for a neural network model in general, and the Virtual Personalities Model, in particular (Read et al., 2017).

We will shortly describe how these systems are implemented in neurobiological systems.

2. Neural networks: Brief introduction

Our theory of personality ties together the psychological assumptions of our model with what is increasingly known about the neurobiology of Motivation and decision making. We implement our theory as a neural network model. Implementing the theory as a computational model has the benefit of allowing us to actually simulate and run the model to test the implications of the various assumptions we make in the theory. Moreover, it has the added advantage that it makes it easier to map between the features of the model and the underlying neurobiology of the brain in which personality is embodied.

Neural network models are biologically inspired models. They are designed to capture what we know of the abstract properties of how real brains process information. This approach to modeling can capture central aspects of cognitive processing, such as image recognition and language comprehension (O'Reilly, Munakata, Frank, Hazy, & Contributors, 2012).

Neural network models are constructed of nodes and the weighted links between them. Nodes sum activation they receive over the weighted links from other nodes. Links between nodes can be either excitatory (positive) or inhibitory (negative). The input to a node is the sum of the activations of the different nodes sending a link to the target node, times the weight on the links to the nodes. The strength of the activation that a receiving node then sends is a function of this summed input. The output function can be linear, but is usually nonlinear, typically binary or sigmoidal (S-shaped). Neural networks with nonlinear activation functions have more powerful learning and processing capabilities.

In the architecture we use, nodes are typically arranged in layers and there is competition among the nodes within a layer. The degree of competition within a layer can be tuned to control the average activation of the layer and the number of nodes that become active in the layer.

Processing in a neural network model proceeds by sending activation through a hierarchy of nodes. One central feature of neural networks is what are called Hidden layers. Hidden layers provide powerful learning and representational abilities to neural network models, as they enable such models to learn a hierarchy of increasingly abstracted features (O'Reilly et al., 2012). (One way to think of this is that they enable the system to learn representations for interactions among input features.) For example, in the visual system they allow humans and other animals to start with "points" of light on the retina and build up increasingly abstract representations so that they are ultimately able to recognize that the object in front of them is a friend, or a cat sitting on a desk.

Our neural network models are constructed in a specific neural network architecture called Leabra (Aisa, Mingus, & O'Reilly, 2008; O'Reilly et al., 2012), which is a biologically inspired architecture. Several important aspects of Leabra, for our

https://doi.org/10.1017/pen.2018.6

purposes, are the following. First, competition between nodes in a layer is an inherent feature of the architecture, and can be easily tuned, which makes it easy to capture competitive dynamics within a layer. Second, the activation function for nodes is biologically inspired, modeling the independent contributions of excitatory and inhibitory conductances (corresponding to the number of synaptic channels) into the node. These can be tuned as desired. Third, the architecture integrates Hebbian learning (correlational) and error-correcting learning into a single learning rule. The error-correcting component of the rule is functionally similar to the back-propagation rule widely used in error-correcting learning, but it implements such learning in a more biologically plausible way. Fourth, the architecture assumes that neurobiological systems are massively bidirectionally connected, consistent with what we know of the architecture of real brains.

Building a neural network model of our psychological theory has several advantages. First, because it is "runnable" we can explicitly vary different aspects of the model, such as the sensitivity of the Approach and Avoidance systems, the baseline importance of different goals, and learning history for associations between different cues and goals. Second, we can use the model to demonstrate things such as how structured motivational systems can result in something like the Big Five. In other work, we have shown how this neural network model can capture a number of different aspects of personality. Read, Droutman, and Miller (2017) have shown how such a neural network model, composed of structured motivational systems, can capture the between individual structure of personality (e.g., the Big Five). Read et al. (2017) have shown how the same kind of model can also capture the high level of within subject variability in trait-related behavior over time and across situations. Third, because the architecture is biologically inspired it allows us to more easily identify conceptual links to possible neural substrates for the different processes and representations we propose in our theory. For example, nodes or artificial neurons in Leabra have parameters for the number of excitatory and inhibitory conductances, which would allow one to do things like vary individual differences in the sensitivity of dopamine receptors in the nucleus accumbens.

However, we want to make it clear that simply because Leabra has some biological inspiration, we are not claiming that use of the Leabra architecture automatically allows us to say something about the detailed neurobiological organization of the different systems we are discussing. In order to have something to say about that we would have to do the detailed work of creating systems in which the neural organization was explicitly represented. Our current model building focuses more on modeling the functions of different systems and their functional organization.

3. Virtual Personalities Model: A neural network implementation

Figure 1 outlines the general structure of the neural network model. Features in the Environment activate nodes in the Environment layer, and information about bodily state activate nodes in the Interoceptive state layer. Weighted links from these nodes then transmit activation to the relevant nodes in the Approach and Avoidance layers where the resulting degree of activation is calculated independently for each type of reward. To be clear, the Approach and Avoidance layers function as two separate processing systems.

Following work by Berridge (Berridge, 2012; Zhang, Berridge, Tindell, Smith, & Aldridge, 2009) we characterize the degree of activation of a goal/reward in these two motivational systems as representing the degree to which the individual desires or WANTS that goal. According to Berridge (Berridge, 2007, 2012; Zhang et al., 2009) the degree of WANTING is a multiplicative function of the relevant Environment cues and Interoceptive cues. In an impressive body of work, Berridge and his colleagues (e.g., Berridge, 2007; Berridge & O'Doherty, 2013; Berridge & Robinson, 1998) have made a strong case for the distinction between Wanting a reward and Liking it. He has provided evidence that these are phenomenologically distinct and they arise in different neural circuits. WANTING is the strength of the need or desire for a reward, whereas LIKING is the pleasure received from consuming the reward (or sometimes from imagining its consumption).

As the process is multiplicative, the Environment and Interoceptive features can serve a gating influence for each other. That is, if Environmental cue strength is high, but Interoceptive state is low there will be little resultant activation (or WANTING). Moreover, the reverse would also be true. Nodes representing the different motives compete for activation within each layer or motivational system (Approach, Avoidance). The Approach and Avoidance layers process the two types of motives independently; they do not compete with each other. An Executive Function system can provide top-down biasing on the strength of activation of different motives by maintaining an active goal representation and sending excitatory activation to the corresponding representation in either the Approach or Avoidance systems. Motives that win the competition within each layer then send activation to the Behavior layer, where the different potential behaviors that are activated by the strength of Wanting from different motives compete with each other for activation. The enacted behavior can then change the Environment (Consummation) and/or it can change the Interoceptive state (Satiation), which then changes the inputs on the next step in the behavioral sequence.

4. Neurobiological underpinnings

Here we describe the mapping from our neural network model to its neurobiological underpinnings (see Figure 2).

4.1. Environment layer: Sensory/Cortical systems

The Environment layer represents inputs from sensory and cortical systems about cues to different objects or events in the Environment. Some of the cues may be unconditioned stimuli or primary rewards, such as food, water, other people, whereas other cues will be conditioned stimuli (CSs), which are learned cues, such as visual features of the Environment that are associated with and can activate the representation of a particular reward, such as food.

Through what is alternatively called Classical or Pavlovian Conditioning the organism learns that certain CSs (such as an arbitrary visual or auditory cue) predict the occurrence of rewards in the Environment. This occurs through a process of associative learning; when a CS and a reward are presented closely in time, the CS and reward become associated, so that presentation of the CS activates a representation of the reward. This process of associative learning is thought to occur in the amygdala (Hazy, Frank, & O'Reilly, 2010). In the neural network model, when the CS and reward are concurrently activated, the weight of the link

https://doi.org/10.1017/pen.2018.6

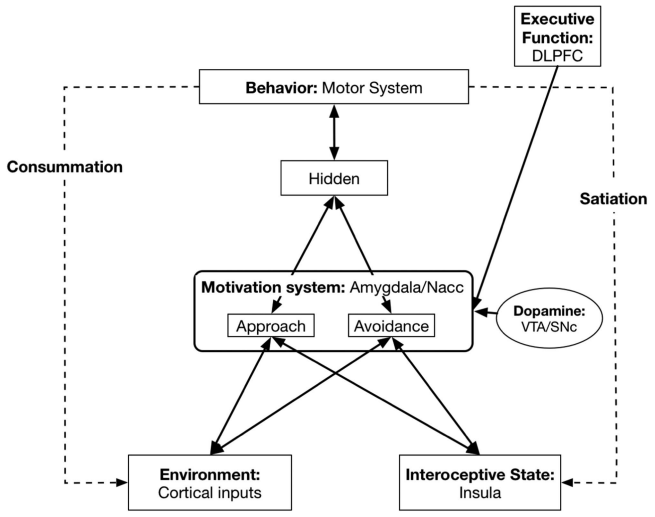

Figure 2. Basic model with neurobiological Approach/Avoidance system neurobiological underpinnings. DLPFC = dorsolateral prefrontal cortex; NAcc = nucleus accumbens; VTA = ventral tegmental area; SNc = substantia nigra pars compacta.

between the CS and reward representations are strengthened, eventually allowing the CS to activate Motivations.

One way to think about this process is that it represents how individuals learn what features of the Environment come to predict the presence of goal affordances in the Environment. Through experience, individuals learn what cues in the Environment predict the goal relevant aspects of the Environment. This learning process can capture one aspect of individual differences. As a result of different learning histories or experiences, different individuals will develop different cue-affordance associations. For example, some people may learn that social situations are likely to be rewarding, whereas others may learn that social situations are an opportunity for social rejection.

4.2. Interoceptive state: Posterior insula

In our current model, we focus on the role of the posterior insula in representing information about the current Interoceptive state of the organism. For example, are they hungry, thirsty, lonely, horny, fatigued? However, the insula is a relatively large neural region that is involved in a number of different aspects of processing. Recent reviews of the insula (Droutman, Bechara, & Read, 2015; Droutman, Read, & Bechara, 2015) suggest that subregions of the insular cortex play significant roles in four phases of the decision-making process. The first phase involves Salience processing and refocusing attention (dorsal anterior insula cortex [dAIC]). The second phase involves evaluation that involves all three subregions of the anterior insula cortex including: (1) the ventral anterior insula cortex (vAIC) that appears to track arousal variance, skew, and risk-prediction error, (2) the posterior insula cortex that is involved in urge processing and the signaling of homeostatic imbalance, and (3) the dAIC that plays a role in tracking arousal, and the magnitude of that arousal and its variance and associated risk and urge generation. The third phase of the insular cortex's role in decision making involves action. The dAIC plays a role in the "what," "when," "whether to act" aspects of action while the vAIC plays a role in action inhibition. In the final phase of decision making the IC is involved in outcome processing, with the vAIC and dAIC both playing a role in error awareness and social outcomes and the dAIC also playing a role in posterror correction and harm prevention. An additional review of work on the insula and fatigue (Dantzer, Heijnen, Kavelaars, Laye, & Capuron, 2014) suggests that peripheral inflammation may also register in the anterior insula, affecting a subjective experience of fatigue or uncertainty about an action's usefulness (suggesting that the latter phases of decision processes in the insula may be impacted). Interestingly insula activation may also activate the fronto-striatal network resulting in diminished switching capacity from goal-directed to habit-like behavior and/or decreased Incentive Motivation (Dantzer et al., 2014).

In the current model, the insula tracks different bodily states (e.g., related to hunger, sexual excitement, loneliness, etc.), which are typically thought to be represented in posterior insula. That information about Interoceptive state is multiplicatively combined with outcome predictions from the Sensory/Cortical systems to produce the WANTING (or Incentive Salience) for a reward.

4.3. Motivation system: Amygdala/nucleus accumbens (NAcc)

In Figure 2 we specify the underlying neural underpinnings for the Approach and Avoid systems. Central to the model is a Motivation system that represents the degree to which a particular motive is wanted. Both Approach and Avoidance type motives are represented within the Motivation system, although the two types of motives are processed independently, represented by two separate layers in our neural network model. Following Berridge and colleagues (Mahler & Berridge, 2009, 2012; Peciña & Berridge, 2013) this Motivation system is composed of a circuit involving the amygdala and the NAcc. (The NAcc is also referred to as the ventral striatum.). Different regions of neurons in NAcc (Reynolds & Berridge, 2008) represent different types of rewards. This system multiplicatively combines input information about the strength of cues to reward from the Sensory/Cortical systems with Interoceptive information from the insula about the organism's current Interoceptive or bodily state. The result is the current level of WANTING for the relevant motive. The impact of conditioned cues on WANTING, as represented in the Amygdala/NAcc circuit, has been termed by Berridge (Berridge, 2012; Zhang et al., 2009) as cue-triggered Wanting.

Because of the multiplicative relationship, each of the two types of cues essentially plays a gating role on the influence of the other cue. For example, if there is a strong cue to a reward, such as food or attractive people, but there is no need, then the Motivation system will not be activated. And if there is a strong need, but no cues to rewards then there will not be an activation of the Motivation system. However, to capture cases where there is strong need but no cues to the reward, we have implemented a direct connection from the Interoceptive state to possible seeking behaviors. For example, if an individual is very hungry or very lonely, but their current Environment does not contain the relevant affordances, the individual would be motivated to seek an Environment that does have those affordances.

As discussed earlier, Berridge and others (Berridge, 2012; Dayan & Berridge, 2014; Zhang et al., 2009) have made a convincing case that Wanting something and Liking something are different both psychologically and neurobiologically. Neurobiologically, Wanting seems to strongly depend on the neurotransmitter dopamine and the dopaminergic circuitry in the brain, whereas Liking depends on brain opioids. Wanting is implemented in our current model, but Liking is not.

https://doi.org/10.1017/pen.2018.6

4.4. Dopamine levels: Ventral tegmental area/substantia nigra pars compacta

The NAcc is strongly innervated by dopaminergic neurons from the ventral tegmental area (Collins & Frank, 2014). Higher dopamine inputs increase the sensitivity of the Approach system to the Environmental and Interoceptive inputs (through D_1 receptors) (Collins & Frank, 2014), whereas they decrease the sensitivity of the Avoidance system to threats (through D_2 receptors) (Collins & Frank, 2014). Thus, individual differences in Tonic levels of dopamine will influence the strength of Approach Motivations. Higher levels lead to greater Approach (and reduced Avoidance). Considerable evidence (e.g., Depue & Collins, 1999; DeYoung & Allen, in press) suggests that individual differences in Tonic dopamine level is one major factor that underlies Extraversion, as well as Openness to Experience/Intellect.

There is a key distinction between Tonic levels of dopamine and Phasic levels(Schultz, 2015). Tonic levels of dopamine are the consistent baseline level and have a chronic influence on the strength of Wanting (Berridge & O'Doherty, 2013; Berridge & Robinson, 1998; Collins & Frank, 2014). Phasic changes in dopamine level are momentary shifts in dopamine firing and are typically a response to reward prediction errors, and play an important role in learning a cue-reward association (Collins & Frank, 2014; Schultz, 2015). We have modeled the impact of Tonic dopamine levels, but have not explicitly included Phasic dopamine changes.

4.5. Behavior: Motor systems

Activation from the Motivation system, as well as direct activation from the Environment system that conveys information about the availability of various resources, are fed into a Hidden layer, where they may be combined in a conjunctive representation, and go from there to the behavioral system. In the behavioral system, different possible actions compete for activation and the most strongly activated will be enacted by the Motor system.

4.6. Satiation and Consummation

An organism's behavior has consequences for the next step in the behavioral sequence. Behaviors can change the nature of the Environment to which the organism is responding (reduce amount of food, leave a situation with people). Changes in the Environment can then influence the strength of the cues that feed into the Motivation system as the organism is considering subsequent behaviors. Behaviors can also lead to Satiation of internal needs: eating reduces hunger, hanging out with friends reduces loneliness. These changes in bodily states are then input into the network on the next time step. Thus, the behavior of the organism and its impact on internal and external Environment are major contributors to variability in trait-related behavior over time and situations. For example, an extraverted individual might end up seeking privacy after hours of partying.

4.7. Executive Function, self-regulation: Dorsolateral prefrontal cortex (DLPFC) and related systems

Our model captures the top-down influence of Executive Function on Wanting by sending activation from a sustained goal representation, in a layer that represents the function of DLPFC, to its corresponding representation in the Motivation system. This top-down influence (activation) helps maintain the activation of certain Motives in the face of competition from alternative Motives that might be activated, thus providing a degree of consistent goal focus. The implementation of Executive Function in our model is consistent with recent work in this area (Wiecki & Frank, 2013).

Individual differences in self-regulation or Executive Function can be modeled in terms of differences in the ability to maintain a goal representation when faced with competing influences, and in terms of differences in the strength of top-down influence on the Motivation system. This would capture some of the more controlled aspects of Conscientiousness.

More automatic aspects of inhibitory control could be modeled in terms of general levels of inhibition within early processing layers, as discussed by Read et al. (2010). In that paper, we showed that greater inhibition in the Motivation layer led to fewer switches to different goals, indicating that this can capture some aspects of automatic inhibitory control (Clark & Watson, 1999, 2008; Rothbart & Bates, 1998).

5. Model instantiates classic learning phenomena

This model and its underlying neurobiology allow us to capture basic learning phenomena and show how they can be related to personality and individual differences. In particular, the model captures Pavlovian or Classical Conditioning, Instrumental Conditioning, and their combination, called Pavlovian to Instrumental Transfer (PIT) (see Figure 3). In addition, it can also model the development and representation of habitual behavior.

5.1. Pavlovian (or Classical) Conditioning

The pathway from Environmental cues (both CS and unconditioned stimuli) to the amygdala in the Motivation system models Pavlovian Conditioning. This part of the model can capture individual differences in Conditioning history: the extent to which different features in the Environment activate different reward representations.

This pathway is also a critical part of what Berridge calls cue-triggered Wanting. As the amygdala and NAcc are tied together in a Motivation system, Pavlovian Conditioning is responsible for the development of connections between CS and WANTING for a particular reward.

5.2. Instrumental Conditioning

In Instrumental Conditioning an organism learns that a particular instrumental behavior leads to a particular reward or avoids a punishment. This has been shown to require involvement of the dorsal medial striatum (also called the Caudate) (Balleine & O'Doherty, 2010; Dolan & Dayan, 2013). Lesioning of the Caudate eliminates the sensitivity of behavior to the current level of Wanting (Balleine & O'Doherty, 2010; Dolan & Dayan, 2013). A number of researchers have argued and demonstrated that in Instrumental Conditioning the organism learns bidirectional connections between action and reward/punishment (Balleine & O'Doherty, 2010), so that activation of the representation of a reward/punishment can also activate the corresponding behavior.

Again, individual differences in experience can result in individual differences in the strength of connections between behaviors and potential goals. As a result, different individuals may learn different behaviors in response to the activation of different rewards/punishments. For example, some individuals will learn that social interaction is rewarding and will actively seek it out, whereas others will learn that it is often punishing and will learn to avoid it.

https://doi.org/10.1017/pen.2018.6

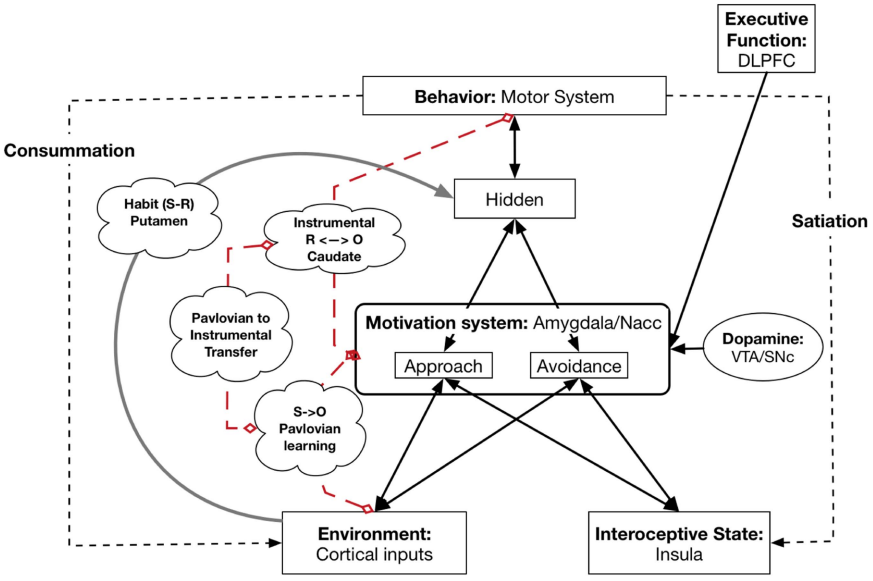

Figure 3. Neural underpinnings of learning. Broken red lines indicate which systems are involved in Pavlovian learning, Instrumental learning, and Pavlovian to Instrumental Transfer; they are not pathways. The gray line from Environment to the Hidden layer represents the Habit system (instantiated in the Putamen or dorsal lateral striatum that develops with high levels of learning of the associations for S to R. DLPFC = dorsolateral prefrontal cortex; NAcc = nucleus accumbens; VTA = ventral tegmental area; SNc = substantia nigra pars compacta.

5.3. Pavlovian to Instrumental Transfer (PIT)

A central part of the Virtual Personalities Model can be viewed as an instantiation of a well-known phenomenon in reward learning called PIT (Corbit & Balleine, 2011; Talmi, Seymour, Dayan, & Dolan, 2008). In PIT, an organism learns an association between a CS and a reward/punishment, so that presentation of the CS leads to activation of a representation of the reward/punishment (Talmi et al., 2008). (Although PIT applies to both reward and punishment, the preponderance of work has focused on reward.) Separately, the organism learns the association between an Instrumental response (e.g., lever pressing) and the attainment of the same outcome (Talmi et al., 2008). This association between Instrumental response and outcome has been shown to be bidirectional, such that activation of the reward or punishment representation can trigger the Instrumental response (Holland, 2004). Once both of these sets of associations have been learned, if the CS is presented concurrently with the opportunity to perform the instrumental behavior, the presence of the CS increases the vigor of responding with the instrumental behavior (Talmi et al., 2008).

We propose that this PIT process underlies a tremendous amount of human behavior. As we have proposed for a long time in our personality model (Miller & Read, 1987, 1991; Read & Miller, 1989), we can think of situations as containing sets of cues that predict the availability of affordances for the pursuit of various goals. Features of a classroom give cues to the affordances for academic achievement in that context. Features of a dance party give cues to the affordances for pursuing affiliation and mating. (Although for some a dance party affords opportunities for social rejection and potential humiliation.) Affordances can then specify behavior (e.g., completing classwork, approaching other partygoers, staying home and reading) to be implemented in pursuit of these motives, completing the link from cues to goal activation to behavior.

In our current neural network model of personality, learned and unlearned cues to reward activate Wanting for those rewards, which then bias choice of actions that can achieve those rewards. This process is essentially PIT. Consistent with our model, we suspect that PIT may help explain how personality traits emerge from an individual's learning history. The chronic manifestation of a trait may result when a large portion of an individual's chronic Environments cue the same set of goals. On the other hand, the trait will not be observed when individuals encounter stimuli that have not been conditioned with those goals, accounting for within-individual variability.

Interestingly, Pavlovian cues can also invigorate responses associated with a different reward, in a process known as general PIT (Corbit & Balleine, 2011). When trained to associate a keyboard button with popcorn, human subjects increase a button response not only when a Pavlovian cue for popcorn is present, but also when a cue for cashew nuts is present (Watson, Wiers, Hommel, & de Wit, 2014). This suggests that a given Pavlovian cue can motivate responses generally, invigorating response for unrelated rewards. One possible way this might work is that the general Pavlovian cue could activate the dopamine system in the Motivational system, which would then potentiate Approach to all rewards.

Although research on this effect is limited, we suspect that general PIT may account for positive correlations between Approach-related traits (Elliot & Thrash, 2002) If a given cue can increase instrumental behavior in a variety of domains, then we would expect that a cue for positive social experiences (i.e., Extraversion cue) can increase Motivation for novelty (Openness to Experience), resulting in a correlation between Extraversion and Openness to Experience. While we have not explicitly included this effect in our Virtual Personalities Model, we are currently developing a model to incorporate general PIT.

5.4. Habitual Behavior

Habitual behavior is directly triggered by a cue and typically develops after extensive practice with an Instrumental response

https://doi.org/10.1017/pen.2018.6

(Balleine & O'Doherty, 2010). Once a behavior becomes habitual its enactment is insensitive to the current goal state of the organism (Balleine & O'Doherty, 2010). For example, once lever pressing for food pellets has become Habitual, a rat fed to satiety may continue to press the lever, even though they are no longer hungry. The learning and performance of habitual behavior depends on the Putamen (dorsolateral striatum) (Balleine & O'Doherty, 2010) and lesioning of the Putamen eliminates Habitual responding and reinstates goal-based responding (Balleine & O'Doherty, 2010). These lesion findings strongly argue that habitual behavior does not depend on the Motivation system, but instead depends on a separate circuit involving the Putamen. In line with this, in the Virtual Personalities Model, there is a link between the Environmental input and the behavior layer that bypasses the Motivation system. Individual differences in experience can lead to individual differences in whether a particular cue-behavior link becomes Habitual or remains sensitive to current goals.

6. How Individual differences can be captured in this model

6.1. Neurobiology

This model allows us to examine how underlying mechanisms and their interactions with one another may produce various individual differences. First, the model allows us to capture various aspects of the neurobiology of individual differences. One major factor, which has been identified by a number of theorists, is the role of the Approach or reward system and the role of dopamine in Approach behavior. Depue and others (Depue & Collins, 1999; DeYoung, 2015) have argued and provided evidence that Tonic dopamine levels are central to Extraversion, as well as to Openness to Experience/Intellect, with higher levels of Tonic dopamine resulting in stronger Approach-related behaviors (Extraversion). Tonic dopamine levels are explicitly modeled in our network and can be manipulated to examine the impact of this parameter on different behaviors.

More speculatively, individual differences in Avoidance behaviors may be related to Tonic levels of serotonin (Corr, DeYoung, & McNaughton, 2013) and the role that they play in the Avoidance system. Somewhat controversially, higher levels of serotonin are argued to lead to greater activity of the Avoidance system and thus to increased Avoidance behaviors (Neuroticism). There is some indication that this might occur through down-regulation of dopaminergic receptors (Corr, DeYoung, & McNaughton, 2013), so that the organism would be less sensitive to Tonic levels of dopamine.

We can also simulate the impact of individual differences in the importance of different kinds of rewards by manipulating their baseline activations in the Approach and Avoidance systems. Research by Berridge and others (Reynolds & Berridge, 2008) suggests that there are specific groups of neurons in NAcc that represent different kinds of rewards. The sensitivities of these systems to their inputs may vary.

Individual differences in some aspects of Conscientiousness can be captured by individual differences in Executive Function, as captured by the DLPFC. As we noted earlier, DLPFC can maintain a representation of desired goals and through top-down connections can influence the degree of activation of WANTs in the amygdala–NAcc circuit.

As all of the parameters in the model can be systematically varied simultaneously, we can also examine interactions among different factors. For example, we can explicitly model some types of impulsivity as the result of the interaction between the strength of the Approach system and the Control system. A number of different researchers have argued that Impulsivity is not just the result of single cause or system, but is the result of the interaction of several different systems, such as the reward and the self-control system (e.g., Depue & Collins, 1999; Zuckerman, 2005). The suggestion is that impulsive people are those whose Approach or reward systems can override self-control systems. Our model can capture this form of impulsivity. However, we recognize that there may be other forms of impulsivity (Revelle, 1997).

6.2. Learning

Because learning is central to our neural network architecture (and indeed to almost any neural network architecture), we can examine the impact of different learning experiences on the development of individual differences. Different individuals may learn different sets of cue-reward/punishment contingencies depending upon the kinds of contingencies to which they are exposed and whether those contingencies are rewarding or punishing. For instance, an individual who experiences consistent patterns of rejection at home and at school may develop strong rejection sensitivity and social anxiety, especially in those situations, because they develop a strong expectancy of being rejected in social situations.

Our model encompasses three different kinds of learning that might underlie individual differences. First, Pavlovian or Classical Conditioning plays a central role in the development of expectancies about the kinds of rewards and punishments that occur in different contexts. For different people, similar cues may become associated with very different kinds of rewards and punishments. For some people, cues to social contexts lead to expectancies of positive social outcomes, whereas for other people it leads to expectancies of negative outcomes. Second, Instrumental Conditioning is responsible for learning what is likely to happen in response to different actions, or conversely what actions to take if one desires a particular reward. Different individuals may learn different Instrumental responses to pursue the same reward. Or some individuals may learn the Instrumental responses needed to attain a particular reward or avoid a particular aversive outcome, whereas other individuals may not learn such associations or may not learn them as well.

Finally, Habits develop when there is long-term exposure to action-reward/punishment contingencies. After enacting the same action to get the same reward or avoid the same punishment over a large number of instances, the likelihood of enacting the action is no longer sensitive to the current goal state of the organism (Balleine & O'Doherty, 2010; Dayan & Berridge, 2014). Instead, it is directly triggered by the cue. Obviously, with different patterns of experience, different individuals can develop quite different Habitual responses in the same context.

7. Dynamics of personality

Above, we have laid out a neural network model of personality at an experiential and neurobiological level of scale. As we can model both individual differences in experience and individual differences in neurobiology, we can also study the interaction between learning history and underlying neurobiological differences. For example, as a number of researchers have pointed out, personality characteristics may influence what kinds of

https://doi.org/10.1017/pen.2018.6

situations individuals seek out (Ickes, Snyder, & Garcia, 1997; Snyder, 1983), which would potentially lead to different patterns of outcomes. For instance, Extraverts may be more likely to seek out social situations and as a result of this increased exposure they may be more likely to learn about cues to social reward. In addition, other factors (e.g., chronic fatigue due to chronic inflammation) that personality psychologists normally do not attend to, might play surprising roles in personality expression (or changes over time in apparent personality) because of how they (e.g., Interoceptive signaling) affect action propensity, Incentive Motivation, and ability to switch from more goal-directed to more habit-like behavior over time (Dantzer et al., 2014).

Examining the dynamics of persons and situations has historically been challenging. In the current work, we suggest that a neural network model of personality affords tremendous promise to allow us to build increasingly more refined models of these dynamics, leveraging emerging work on the neurobiological underpinnings that support them. For example, initial differences in propensities (e.g., dopamine functioning), wiring differences due to experiential learning, and other factors could operate together to develop and change personality over time: now, we are beginning to have tools to examine "what if" assumptions and the implications of those assumptions on behavioral outcomes. We can use such models and simulations not only to explain what we currently know (and to update them as new emerging findings suggest state of the art "plausible links"), but to suggest novel hypotheses—whose dynamics would have been hard to otherwise envision—that might be empirically examined. In this way, such models could become rich tools for the field to build a cumulative, ever more detailed and precise, model of personality dynamics.

8. Relationship to other neurobiological approaches to personality

Other neurobiologically inspired personality theories like reinforcement sensitivity theory (RST; Gray & McNaughton, 2000) and Cybernetic Big Five Theory (DeYoung, 2015) also draw upon the research described elsewhere in this article that pertains to Extraversion or Neuroticism, respectively. However, there are several important distinctions, especially with regard to RST's handling of different components of Neuroticism and to its more nuanced speculations with regard to brain structures and processes that others have ascribed to Extraversion or Neuroticism, but not both.

Corr, DeYoung, and McNaughton's figure 2 (2013, p. 163; adapted from McNaughton & Corr, 2004; see also Gray & McNaughton, 2000, p. 276, for an earlier version of this diagram) is particularly helpful in summarizing the manner in which RST maps onto brain structures. To begin with the frontal lobes, Corr, DeYoung, and McNaughton's (2013) figure suggests that the PFC's ventral stream is associated with RST's BIS, which governs "defensive Approach" behaviors, creates anxious affect, and may be likened to DeYoung, Quilty, and Peterson' (2007) "Withdrawal" aspect of Neuroticism, whereas the PFC's dorsal stream is associated with RST's FFFS, which governs "defensive Avoidance" behaviors, creates fearful affect, and may be likened to DeYoung, Quilty, and Peterson's (2007) "Volatility" aspect of Neuroticism. The PFC's role in Neuroticism (and in the BIS and FFFS) may also differ by hemisphere; greater activation in the right prefrontal regions (as well as damage to left prefrontal regions) is positively correlated with Withdrawal, whereas greater activation in the left prefrontal regions is positively correlated with Volatility. Like the dorsal and ventral PFC, respectively, the cingulate cortex may be divided into two regions, anterior cingulate cortex (ACC) and posterior cingulate cortex, which Corr, DeYoung, and McNaughton (2013) claim are differentially associated with RST's FFFS and BIS; these authors link the ACC to obsessive–compulsive disorder and the posterior cingulate cortex to ruminative anxiety. Given DeYoung and Allen's (in press) statement that frontal structures are more likely to be related to Neuroticism via their effect on affect regulation than on affect generation, it is interesting to note that they also report that Neuroticism, as well as Conscientiousness, has been negatively related to ACC thickness.

DeYoung and Allen (in press) attribute the role of affect generation in Neuroticism primarily to medial temporal lobe structures, especially the septo-hippocampal system and amygdala (which ostensibly comprise the Withdrawal, or BIS-like, aspect of Neuroticism) and the midbrain's hypothalamic-pituitary-adrenal axis (which controls Neuroticism's Volatility, or FFFS-like, aspect).

The RST literature suggests that the amygdala is involved in both defensive Avoidance (FFFS) and defensive Approach (BIS). Specifically, Corr, DeYoung, and McNaughton (2013) cite Cunningham, Arbuckle, Jahn, Mowrer, and Abduljalil's (2010) functional magnetic resonance imaging (fMRI) study, in which Volatility predicted amygdala activity in response to stimulus valence (negative over positive), whereas Withdrawal predicted amygdala activity in response to changes in distance (Approach over Withdrawal) from stimuli, independent of valence.

Research in the RST tradition has also painted a somewhat different picture of the relationship between the hypothalamic-pituitary-adrenal axis and Neuroticism's BIS- and FFFS-like aspects. Corr, DeYoung, and McNaughton (2013) maintain that the medial hypothalamus and the periaqueductal gray are capable of producing either defensive Approach or defensive Avoidance behaviors (e.g., risk assessment versus escape, defensive quiescence versus explosive panic); however, they admit that these more inferior structures are more likely to be activated in response to nearby, immediate threats than to threats that are more spatiotemporally distant. As such, they are more likely to create Avoidant behaviors characteristic of Volatility than defensive Approach behaviors characteristic of Withdrawal.

With regard to the role of various neurotransmitters, there is also much agreement between the RST literature and the basic neuroscience literature. For instance, the central role that dopamine plays in Extraversion (a positive correlate of the Behavioral Activation/Approach system, or BAS, which governs sensitivity to rewarding cues) is as undisputed among reinforcement sensitivity theorists (Pickering & Gray, 1999; Pickering & Pesola, 2014) as it is among other neurobiologically inclined personality theorists (e.g., Depue & Collins, 1999).

Where the RST literature differs from other lines of research is its strong emphasis on differentiating between defensive Approach and defensive Avoidance behaviors. Corr, DeYoung, and McNaughton (2013) suggest, for example, that higher tonic levels of serotonin do not increase one's Stability metatrait score simply by decreasing general Neuroticism, but rather decreases scores on some facets of Neuroticism (i.e., those linked to Volatility or FFFS) while increasing others (i.e., those linked to Withdrawal or BIS). Similarly, these authors draw a distinction between cortisol reactivity and various types of aggression (reactive or defensive aggression is linked to the FFFS and to cortisol, whereas proactive or offensive aggression is not).

https://doi.org/10.1017/pen.2018.6

9. Conclusion, limitations, and future directions

A rising chorus of voices seeks a more precise neurobiologically based science of the complex dynamics underlying personality. Our goal in the current work is to contribute to that rising chorus of voices. In doing so we suggest that we can model personality at a number of different levels of scale using our best guesses based on the available literatures in personality and neuroscience to address and gain insight into what have seemed like intractable enigmas. These are exciting times. The tools available today suggest that we can dig in and really get traction on cumulatively understanding these complex personality dynamics. How can we understand the complex dynamics of persons and situations and how they operate together over time to produce emergent behavior? How could the same underlying system that results in between-subject Stability in behavior over time and reliable personality structure across individuals (e.g., the Big Five), plausibly explain broad within-person variability in behavior across situations? How do underlying biological mechanisms and different learning histories produce enduring individual differences in response to situational cues? To date we have provided computational models to explore how such enigmas might plausibly be explored. But, what we have provided thus far is more akin to plausible computational models.

It's apparent to us that the road ahead involves a very steep slope. We must move beyond plausible computational solutions to enigmas, and use computational approaches to predict emerging patterns of individuals that we can measure (and then use those to refine our computational models). There are many challenges. First, ultimately the precision of our models of the dynamics of personality depend, in part, upon the precision of our measurement tools. For example, fMRI at present does not allow us to know the precise path of the neural circuitry signaling *over time* that results in various decisions, learning, and personality-related outcomes of interest. The cause-effect inferences needed for our computational models of the underlying neurobiological pathways depend on converging evidence (e.g., from animal and human studies where there can be systematic manipulation and fine-grained electrode measurements).

Second, understanding personality dynamics requires understanding how people respond in situations, including those involving others over time, while concurrently acquiring neurobiological (e.g., fMRI) data. fMRI at present is available only in the lab, so that studying more complex real-world situations and concurrently examining neurobiological circuits may require fMRI studies using more realistic gaming environments designed to be representative of real-life challenges (using intelligent agents, for example, who are modeled to be similar in their response patterns to other human partners). We have developed and used such a game in our lab, examining neural circuit patterns for men who have sex with men who are playing a virtual dating game that was designed to be representative of men's real-life scenarios involving sexual risks. That work is quite exciting. However, the next step is a Step 1. Can we create an iterative way to computationally model men's behavioral choices in the game and test those individual computational models against actual individual differences in men's actual game choices and corresponding model-generated expectations regarding their changing neural patterns while playing the game? In any event, cumulatively better understanding the situations in which within and between person differences in personality dynamics over time can be more precisely assessed both Behaviorally and neurobiologically is a major challenge. So too is developing cumulatively more precise measures of those dynamics, computationally and in "real" or "virtual" time.

Financial Support: This work was supported by the National Institute of General Medical Sciences (grant number R01GM109996).

Conflicts of Interest: All authors have nothing to disclose.

References

Aisa, B., Mingus, B., & O'Reilly, R. (2008). The emergent neural modeling system. *Neural Networks*, *21*, 1146–1152. https://doi.org/10.1016/j.neunet.2008.06.016

Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, *35*, 48–69. https://doi.org/10.1038/npp.2009.131

Berridge, K. C. (2007). The debate over dopamine's role in reward: The case for incentive salience. *Psychopharmacology*, *191*, 391–431. https://doi.org/10.1007/s00213-006-0578-x

Berridge, K. C. (2012). From prediction error to incentive salience: Mesolimbic computation of reward motivation. *European Journal of Neuroscience*, *35*, 1124–1143. https://doi.org/10.1111/j.1460-9568.2012.07990.x

Berridge, K. C., & O'Doherty, J. P. (2013). From experienced utility to decision utility. In P. W. Glimcher, & E. Fehr (Eds.), *Neuroeconomics, second edition. Decision making and the brain* (pp. 325–341). San Diego, CA: Academic Press.

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

Bugental, D. B. (2000). Acquisition of the algorithms of social life: A domain-based approach. *Psychological Bulletin*, *126*, 187–219. https://doi.org/10.1037/0033-2909.126.2.187

Chulef, A. S., Read, S. J., & Walsh, D. A. (2001). A hierarchical taxonomy of human goals. *Motivation and Emotion*, *25*, 191–232. https://doi.org/10.1023/A:1012225223418

Clark, L. A., & Watson, D. (1999). Temperament: A new paradigm for trait psychology. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research*, 2nd ed. (pp. 399–423). New York, NY: Guilford Press.

Clark, L. A., & Watson, D. (2008). Temperament: An organizing paradigm for trait psychology. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research*, 3rd ed. (pp. 265–286). New York, NY: Guilford Press.

Collins, A. G. E., & Frank, M. J. (2014). Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychological Review*, *121*, 337–366. https://doi.org/10.1037/a0037015

Corbit, L. H., & Balleine, B. W. (2011). The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. *Journal of Neuroscience*, *31*, 11786–11794. https://doi.org/10.1523/JNEUROSCI.2711-11.2011

Corr, P. J., DeYoung, C. G., & McNaughton, N. (2013). Motivation and personality: A neuropsychological perspective. *Social and Personality Psychology Compass*, *7*, 158–175. https://doi.org/10.1111/spc3.12016

Cunningham, W. A., Arbuckle, N. L., Jahn, A., Mowrer, S. M., & Abduljalil, A. M. (2010). Aspects of neuroticism and the amygdala: Chronic tuning from motivational styles. *Neuropsychologia*, *48*, 3399–3404. https://doi.org/10.1016/j.neuropsychologia.2010.06.026

Dantzer, R., Heijnen, C. J., Kavelaars, A., Laye, S., & Capuron, L. (2014). The neuroimmune basis of fatigue. *Trends in Neurosciences*, *37*, 39–46. https://doi.org/10.1016/j.tins.2013.10.003

Dayan, P., & Berridge, K. C. (2014). Model-based and model-free Pavlovian reward learning: Revaluation, revision, and revelation. *Cognitive, Affective, & Behavioral Neuroscience*, *14*, 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

DeYoung, C. G., & Allen, T. A. (in press). Personality neuroscience: A developmental perspective. In D. P. McAdams, R. L. Shiner, & J. L. Tackett (Eds.), *The handbook of personality development*. New York, NY: Guilford Press.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*, 880–896. https://doi.org/10.1037/0022-3514.93.5.880

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, *73*, 1246–1256. https://doi.org/10.1037/0022-3514.73.6.1246

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*, 312–325. https://doi.org/10.1016/j.neuron.2013.09.007

Droutman, V., Bechara, A., & Read, S. J. (2015). Roles of the different subregions of the insular cortex in various phases of the decision-making process. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00309

Droutman, V., Read, S. J., & Bechara, A. (2015). Revisiting the role of the insula in addiction. *Trends in Cognitive Sciences*, *19*, 414–420. https://doi.org/10.1016/j.tics.2015.05.005

Elliot, A. J., & Thrash, T. M. (2002). Approach-avoidance motivation in personality: Approach and avoidance temperaments and goals. *Journal of Personality and Social Psychology*, *82*, 804–818. https://doi.org/10.1037/0022-3514.82.5.804

Eysenck, H. J. (1983). Psychophysiology and personality: Extraversion, neuroticism and psychoticism. In A. Gale & J. A. Edwards (Eds.), *Psychological correlates of human behavior: Vol. 3. Individual differences and psychopathology* (pp. 13–30). San Diego, CA: Academic Press.

Eysenck, H. J. (1994). Personality: Biological foundations. In P. A. Vernon (Ed.), *The neuropsychology of individual differences* (pp. 151–207). San Diego, CA: Academic Press.

Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, *99*, 689–723. https://doi.org/10.1037/0033-295X.99.4.689

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology*, Vol. 2 (pp. 141–165). Beverly Hills, CA: Sage.

Gray, J. A. (1987a). The neuropsychology of emotion and personality. In A. Gray, S. M. Stahl, S. D. Iverson, & E. C. Goodman (Eds.), *Cognitive neurochemistry* (pp. 171–190). New York, NY: Oxford University Press.

Gray, J. A. (1987b). *The psychology of fear and stress (Vol. 5)*. New York, NY: Cambridge University Press.

Gray, J. A. (1991). The neuropsychology of temperament. In J. Strelau & A. Angleitner (Eds.), *Explorations in temperament: International perspectives on theory and measurement. Perspectives on individual differences* (pp. 105–128). New York, NY: Plenum Press.

Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*, 2nd ed. New York, NY: Oxford University Press.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2010). Neural mechanisms of acquired phasic dopamine responses in learning. *Neuroscience and Biobehavioral Reviews*, *34*, 701–720. https://doi.org/10.1016/j.neubiorev.2009.11.019

Holland, P. C. (2004). Relations between Pavlovian-instrumental transfer and reinforcer devaluation. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*, 104–117. https://doi.org/10.1037/0097-7403.30.2.104

Ickes, W., Snyder, M., & Garcia, S. (1997). Personality influences on the choice of situations. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of Personality Psychology* (pp. 165–195). San Diego, CA: Academic Press.

Kenrick, D. T., & Trost, M. R. (1997). Evolutionary approaches to relationships. In S. Duck (Ed.), *Handbook of personal relationships: Theory, research, and interventions* (pp. 151–177). Chichester, England: Wiley.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, *39*, 329–358. https://doi.org/10.1207/s15327906mbr3902_8

Mahler, S. V., & Berridge, K. C. (2009). Which cue to "want?" Central amygdala opioid activation enhances and focuses incentive salience on a prepotent reward cue. *Journal of Neuroscience*, *29*, 6500–6513. https://doi.org/10.1523/JNEUROSCI.3875-08.2009

Mahler, S. V., & Berridge, K. C. (2012). What and when to "want"? Amygdala-based focusing of incentive salience upon sugar and sex. *Psychopharmacology*, *221*, 407–426. https://doi.org/10.1007/s00213-011-2588-6

McCrae, R. R., & Costa, P. T. Jr. (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research*, 2nd ed. (pp. 139–153). New York, NY: Guilford Press.

Miller, L. C., & Read, S. J. (1987). Why am I telling you this?: Self-disclosure in a goal-based model of personality. In V. J. Derlega & J. Berg (Eds.), *Self-disclosure: Theory, research, and therapy* (pp. 35–58). New York, NY: Plenum Press.

Miller, L. C., & Read, S. J. (1991). On the coherence of mental models of persons and relationships: A knowledge structure approach. In G. J. O. Fletcher & F. Fincham (Eds.), *Cognition in close relationships* (pp. 69–99). Hillsdale, NJ: Erlbaum.

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., & Contributors (2012). *Computational cognitive neuroscience*, 1st ed. Retrieved from http://ccnbook.colorado.edu.

Peciña, S., & Berridge, K. C. (2013). Dopamine or opioid stimulation of nucleus accumbens similarly amplify cue-triggered "wanting" for reward: Entire core and medial shell mapped as substrates for PIT enhancement. *European Journal of Neuroscience*, *37*, 1529–1540. https://doi.org/10.1111/ejn.12174

Pickering, A. D., & Gray, J. A. (1999). The neuroscience of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research*, Vol. 2 (pp. 277–299). New York, NY: Guilford Press.

Pickering, A. D., & Pesola, F. (2014). Modeling dopaminergic and other processes involved in learning from reward prediction error: Contributions from an individual differences perspective. *Frontiers in Human Neuroscience*, *8*, 1–20. https://doi.org/10.3389/fnhum.2014.00740

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

Read, S. J., Droutman, V., & Miller, L. C. (2017). Virtual personalities: A neural network model of the structure and dynamics of personality. In R. R. Vallacher, S. J. Read, & A. Nowak (Eds.), *Computational social psychology* (pp. 15–37). New York, NY: Routledge.

Read, S. J., Jones, D. K., & Miller, L. C. (1990). Traits as goal-based categories: The importance of goals in the coherence of dispositional categories. *Journal of Personality and Social Psychology*, *58*, 1048–1061. https://doi.org/10.1037/0022-3514.58.6.1048

Read, S. J., & Miller, L. C. (1989). Inter-personalism: Toward a goal-based theory of persons in relationships. In L. A. Pervin (Ed.), *Goal concepts in personality and social psychology* (pp. 413–472). Hillsdale, NJ: Erlbaum.

Read, S. J., & Miller, L. C. (2002). Virtual personalities: A neural network model of personality. *Personality and Social Psychology Review*, *6*, 357–369. https://doi.org/10.1207/S15327957PSPR0604_10

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*, 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. http://dx.doi.org/10.1016/j.jrp.2016.10.005

Revelle, W. (1997). Extraversion and impulsivity: The lost dimension? In H. Nyborg (Ed.), *The scientific study of human nature: Tribute to Hans J. Eysenck at eighty* (pp. 189–212). Amsterdam, The Netherlands: Pergamon/Elsevier Science.

Reynolds, S. M., & Berridge, K. C. (2008). Emotional environments retune the valence of appetitive versus fearful functions in nucleus accumbens. *Nature Neuroscience*, *11*, 423–425. https://doi.org/10.1038/nn2061

Rothbart, M. K., & Bates, J. E. (1998). Temperament. In N. Eisenberg (Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development*, 5th ed. (pp. 105–176). New York, NY: Wiley.

Schultz, W. (2015). Neuronal reward and decision signals: From theories to data. *Physiological Reviews*, *95*, 853–951. https://doi.org/10.1152/physrev.00023.2014

Smillie, L. D., Pickering, A. D., & Jackson, C. J. (2006). The new reinforcement sensitivity theory: Implications for personality measurement. *Personality and Social Psychology Review*, *10*, 320–335. https://doi.org/10.1207/s15327957pspr1004_3

Snyder, M. (1983). The influence of individuals on situations: Implications for understanding the links between personality and social behavior. *Journal of Personality*, *51*, 497–516. https://doi.org/10.1111/j.1467-6494.1983.tb00342.x

Talevich, J. R., Read, S. J., Walsh, D. A., Iyer, R., & Chopra, G. (2017). Toward a comprehensive taxonomy of human motives. *PLOS ONE*, *12*, e0172279. https://doi.org/10.1371/journal.pone.0172279

Talmi, D., Seymour, B., Dayan, P., & Dolan, R. J. (2008). Human Pavlovian–instrumental transfer. *Journal of Neuroscience*, *28*, 360–368. https://doi.org/10.1523/JNEUROSCI.4028-07.2008

Tellegen, A., & Waller, N. G. (2008). Exploring personality through test construction: Development of the multidimensional personality questionnaire. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and testing: Vol. II. Personality measurement and assessment* (pp. 261–292). London, England: Sage.

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science*, *10*, 297–303. https://doi.org/10.1111/1467-9280.00157

Watson, P., Wiers, R. W., Hommel, B., & de Wit, S. (2014). Working for food you don't desire. Cues interfere with goal-directed food-seeking. *Appetite*, *79*, 139–148. https://doi.org/10.1016/j.appet.2014.04.005

Wiecki, T. V., & Frank, M. J. (2013). A computational model of inhibitory control in frontal cortex and basal ganglia. *Psychological Review*, *120*, 329–355. https://doi.org/10.1037/a0031542

Wiggins, J. S., & Trapnell, P. D. (1996). A dyadic-interactional perspective on the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 88–162). New York, NY: Guilford Press.

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, England: Cambridge University Press.