Psychological Review
2010, Vol. 117, No. 1, 61–92

0033-295X/10/$12.00 DOI: 10.1037/a0018131

A Neural Network Model of the Structure and Dynamics of Human Personality

Stephen J. Read, Brian M. Monroe, Aaron L. Brownstein, Yu Yang, Gurveen Chopra, and Lynn C. Miller
University of Southern California

We present a neural network model that aims to bridge the historical gap between dynamic and structural approaches to personality. The model integrates work on the structure of the trait lexicon, the neurobiology of personality, temperament, goal-based models of personality, and an evolutionary analysis of motives. It is organized in terms of two overarching motivational systems, an approach and an avoidance system, as well as a general disinhibition and constraint system. Each overarching motivational system influences more specific motives. Traits are modeled in terms of differences in the sensitivities of the motivational systems, the baseline activation of specific motives, and inhibitory strength. The result is a motive-based neural network model of personality based on research about the structure and neurobiology of human personality. The model provides an account of personality dynamics and person–situation interactions and suggests how dynamic processing approaches and dispositional, structural approaches can be integrated in a common framework.

Keywords: personality, neural network models, motivation, goals, traits

Personality is stable—"set like plaster" (James, 1890/1981, p. 126), yet behavior is highly responsive to situations. Personality is about what is unique to the individual, but it is also about what is shared across people (Allport, 1962). Personality is structured but, at the same time, dynamic. Personality is all of these things and more, but it is not clear how they all fit together.

Here, we present a motive-based neural network model of human personality that addresses these and other issues. The model integrates much of what is known about personality.

Currently, work on the structure of personality and work on the dynamics of personality proceed largely independently (for recent discussions see Cervone & Shoda, 1999; Funder, 2001; Mischel & Shoda, 1998). We should note that these two literatures use the term structure in different ways (Cervone, 2005). Structural approaches to personality are concerned with the *interindividual*, psychometric structure of something like the Big Five (John, Nauman, & Soto, 2008); the statistical structure of personality across people. Dynamic approaches are concerned with the intraindividual or within person structure of the processing systems responsible for personality dynamics.

Stephen J. Read, Brian M. Monroe, Aaron L. Brownstein, Yu Yang, and Gurveen Chopra, Department of Psychology, University of Southern California; Lynn C. Miller, Annenberg School of Communication, University of Southern California.

Brian M. Monroe is now at the Programme in Computational Cognition for Social Systems, Institute of High Performance Computing, Singapore Agency for Science Technology and Research (A*STAR), Singapore. Aaron L. Brownstein is now at the Department of Psychology, Touro College. Yu Yang is now at the Johnson Graduate School of Management, Cornell University.

Correspondence concerning this article should be addressed to Stephen J. Read, Department of Psychology, University of Southern California, Los Angeles, CA 90089-1061. E-mail: read@usc.edu

The most widely accepted view of personality structure, the Big Five (and related work) focuses on describing interindividual personality structure but does not provide a model of the psychological dynamics that underlie this structure (Pervin, 1990). Conversely, most attempts to describe the dynamics of personality (the intraindividual structure of personality mechanisms) do not make close contact with interindividual, structural approaches. For example, many recent attempts to understand the dynamics of human personality focused on goals and related motivational constructs (e.g., Cantor & Kihlstrom, 1987; Emmons, 1991; Little, Salmela-Aro, & Phillips, 2006; L. C. Miller & Read, 1987; Mischel & Shoda, 1995; Pervin, 1989; Read & Miller, 1989). These attempts typically did not provide an account of how these motivational dynamics might be responsible for the observed interindividual structure of personality.

To date, there have been only preliminary attempts to provide an integrated model of the interindividual structure of personality and the underlying dynamics of personality. Providing the basis of such an account is a central goal of this article.

A central part of our argument is that personality arises from structured and organized motivational systems. The resulting model allows us to map from the structure and dynamics of these motivational systems to the observed interindividual structure of personality and vice versa. Modeling personality in terms of organized motivational systems also enables us to integrate goal-based approaches to personality, such as those proposed by Read and Miller (L. C. Miller & Read, 1987, 1991; Read & Miller, 1989) and Mischel and Shoda (1995, 1998), into a dynamic model that can explain central features of personality structure.

One weakness of current goal-based models of personality is that the motivational systems, as they are characterized in these models, are relatively undifferentiated and unorganized. There is no account of how these systems are structured. As a result, it is not clear how the motivational systems in these models could result in the interindividual structure of human traits. We argue that the dynamics of organized motivational systems within individuals can produce the structural (or psychometric) relations observed when looking across individuals, such as in analyses of the trait lexicon and personality inventories.

Because personality is the result of a complicated interaction among a large set of elements, we believe that simulating personality provides advantages that traditional approaches do not. Simulations provide a microscope on the underlying complex personality dynamics in response to changing situations. They allow us to see how these elements might interact with each other in ways that are not possible with more traditional methods.

Shoda and Mischel (1998; Mischel & Shoda, 1995; 2008) have also presented a recurrent neural network model of personality, the cognitive-affective process system model (CAPS), which is focused on understanding personality in terms of underlying goal-based units and has shown the usefulness of neural network modeling in understanding the underlying dynamics of personality. However, these researchers have not tried to capture the interindividual structure of human personality. In contrast, a fundamental goal of our model is to capture major aspects of this structure. One central difference is that in our model, mediating units are organized into specific systems, such as approach and avoidance systems, whereas in the CAPS model, the mediating units form a relatively free form network.

Plan of the Article

First, we outline a motive-based model of personality, review the areas of research and theory in personality and neuroscience that served as the foundation of our model, and describe the implications of that work for a model of the structure and dynamics of personality. Second, we describe the neural network implementation of that theoretical model. Third, we present a number of simulations that test aspects of the model. Finally, we discuss the implications of the model for such issues as person–situation interactions and personality dynamics, as well as for understanding how the structural, dispositional approach to personality and the dynamic approach can be integrated.

A Neural Network Model of Personality

Theoretical Background

In our model, we aim to capture the intraindividual structure and resulting dynamics of human personality in terms of hierarchically structured motivational systems. We model personality in terms of two general levels of motivations, with an additional general control system that operates on the motivational systems. At the broadest level are two general motivational systems, which have been variously termed approach and avoidance systems, behavioral approach systems (BAS) and behavioral inhibition systems (BIS; Gray, 1991; Gray & McNaughton, 2000), or behavioral facilitation and BIS (Depue & Collins, 1999). The approach system governs response to rewarding stimuli and strongly parallels the broad trait of extraversion, whereas the avoidance system governs response to punishment and aversive stimuli and strongly parallels the broad trait of neuroticism, particularly its anxiety and fearfulness facets. Each of these two broad motivational systems encompasses and moderates a set of more specific motives, such as affiliation, sex, dominance, avoiding social rejection, and avoiding physical harm. The behavior of specific motives is a joint function of characteristics of the broad motivational system of which it is a part and of its own specific parameters.

Moderating the activity of these motivational systems is a general "control system," characterized as a disinhibition and constraint (e.g., Clark & Watson, 1999; Tellegen & Waller, 2008) system. The inhibitory processes of this control system moderate the activity of the motivational systems and related behavior.

In developing the broad motivational structures in our model, we drew heavily on two sources. First, we drew from work on the lexical analysis of trait language (e.g., Digman, 1997; Goldberg, 1981; John & Srivastava, 1999; Peabody & DeRaad, 2002; Saucier & Ostendorf, 1999) and work on the structure of different trait inventories (e.g., Lee & Ashton, 2004; McCrae & Costa, 1999; Tellegen, & Waller, 2008; Wiggins & Trapnell, 1996). This research provides information on the nature and structure of individual differences, especially on the Big Five: Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness to Experience.

Second, we relied on numerous sources of information about the potential neurobiological bases of human motivation and personality. Recent models of temperament (e.g., Clark & Watson, 1999; Pickering & Gray, 1999; Rothbart & Bates, 1998; Tellegen & Waller, 2008) have identified major dimensions of temperament (e.g., neuroticism, extraversion, BIS, BAS) and provided evidence for their possible biological bases (also see Cloninger, 1987, 1998; Zuckerman, 2005). Further, research by Davidson (e.g., Davidson, Jackson, & Kalin, 2000) suggests that differences in the activation of the right and left prefrontal cortex (PFC) correspond to chronic individual differences in positive and negative affect and are related to differences in Extraversion and Neuroticism, respectively. The dimensions identified in this work have close parallels with four of the major dimensions identified in the Big Five tradition: Extraversion, Neuroticism, Agreeableness, and Conscientiousness.

We were also interested in more specific traits and how they depended on more specific motives. Again, we drew on two different literatures. First, we relied on two goal-based models of traits, one by Read and Miller (e.g., L. C. Miller & Read, 1987, 1991; Read & Miller, 1989) and the other by Mischel and Shoda (e.g., Mischel & Shoda, 1995; Shoda & Mischel, 1998), which provided guidance on how specific trait terms could be related to specific motives.

Second, evolutionary analyses (e.g., Bugental, 2000; Fiske, 1992; Kenrick & Trost, 1997) of the tasks that all humans must solve, recent work in affective neuroscience, and recent work on goal taxonomies (e.g., Chulef, Read, & Walsh, 2001) provided further information about specific motivational systems that underlie individual differences. In the evolutionary analyses, it was argued that a set of brain or motivational systems has evolved to handle the tasks that all humans face. Among these systems are mating, nurturance of young, affiliation and bonding with peers, establishing dominance hierarchies, insuring attachment to caregivers, avoiding physical harm, and avoiding social rejection. In addition, recent work in affective neuroscience (for a review, see Panksepp, 1998) provides evidence for a similar set of motivational systems. A hierarchical taxonomy of human goals (Chulef, Read, & Walsh, 2001) provided further corroboration.

Broad motivational and control structures.

Analysis of personality measures and the lexical analysis of trait language. Personality measures (e.g., Eysenck, 1983, 1994; Lee & Ashton, 2004; McCrae & Costa, 1999; Tellegen & Waller, 2008; Wiggins, & Trapnell, 1996; Zuckerman, 2002) and the lexical analysis of trait terms (e.g., Digman, 1997; Goldberg, 1981; John & Srivastava, 1999; Peabody & De Raad, 2002; Saucier & Ostendorf, 1999) provide considerable evidence for what is termed the Big Five: Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness to Experience, with the strongest evidence for the first four. Researchers (e.g., McCrae & Costa, 1999; Hofstee, de Raad, & Goldberg, 1992; Lee & Ashton, 2004; Monroe, Read, & Miller, 2006) have proposed that each of these broad factors has a number of subcomponents. For example, Extraversion has separate facets for energy level, gregariousness, and dominance. Digman (1997) and Wiggins and Trapnell (1996) argued that an important distinction cutting across the Big Five is that between agentic (individually focused) and communal (social) traits; thus we might expect agentic and communal aspects of Extraversion (Depue & Morrone-Strupinsky, 2005), Neuroticism, and Conscientiousness.

Temperament and neurobiological models of personality. Work on temperament and the neurobiological bases of personality (e.g., Clark & Watson, 1999; Gray, 1987a, 1987b; Gray & McNaughton, 2000; Pickering & Gray, 1999; Rothbart & Bates, 1998; Zuckerman, 2005) provided considerable evidence for at least three broad, biologically based individual differences: approach (extraversion), avoidance (neuroticism), and disinhibition and constraint. There is also evidence that the third dimension of disinhibition and constraint divides into two factors of Agreeableness and Conscientiousness (Clark & Watson, 1999; Zuckerman, 2005). These biologically based differences largely map onto four dimensions of the Big Five: Extraversion, Neuroticism, Agreeableness, and Conscientiousness.

Approach. One broad system produces behavioral approach and the experience of positive emotion. It corresponds to Tellegen, Watson, and Clark's (1999) Positive Activation (PA) factor (Markon, Krueger, & Watson, 2005; Tellegen, 1985; Watson, Wiese, Vaida, & Tellegen, 1999), which represents the extent to which an individual experiences actively positive emotions.

The approach system corresponds to Gray's (1987b, 1991) behavioral approach system (BAS), which is activated by reward or nonpunishment and produces positive emotions (such as hope and happiness). When activated, the BAS results in active approach, accompanied by feelings of energization and positive affect. Panksepp (1998) argued for a similar system, which he calls a seeking system. A number of researchers (e.g., Depue & Collins, 1999; Panksepp, 1998) have argued that dopamine plays a central role in these systems. Increasing dopamine levels lead to greater activation of the system and to increased exploration and vigor of approach.

Depue and colleagues (Depue & Collins, 1999; Depue & Morrone-Strupinsky, 2005) described a higher order construct of Extraversion, which involves positive incentive motivation and corresponds to the approach system. Extraversion includes two subfactors of agency and affiliation. The agency subfactor involves the incentive to approach, explore, and engage a potentially rewarding stimulus. The neurological core of agency is found in the ventral tegmentum, ventral pallidum, and nucleus accumbens, which are connected to each other by dopamine pathways. The affiliation subfactor involves feelings of warmth, affection, and calmness that are experienced when in contact with an affiliative stimulus. Neurologically, affiliation is based on opiate systems throughout the brain.

Davidson, Jackson, and Kalin (2000) reviewed considerable evidence demonstrating that the left and right prefrontal cortices (PFCs) are differentially involved with approach related and withdrawal related emotions and motivations (also see Davidson, 2003; Harmon-Jones & Sigelman, 2001; W. Heller, 1993; W. Heller, Schmidtke, Nitschke, Koven, & Miller, 2002; Schmidtke & Heller, 2004; Tomarken et al., 1992). Davidson (1992, 1995, 1998, 2001, 2003; Tomarken et al., 1992) suggested that the approach system corresponds to an affective style characterized by increased positive affect, sensitivity to positive stimuli, recovery from negative stimuli, approach behavior, and reduced depression and inhibition. This affective style involves higher activation in the left hemisphere PFC than in the right hemisphere PFC; moreover, an active left PFC may inhibit the amygdala, thereby further reducing negative affect. Structurally, this affective style corresponds to Depue and Collins's (1999) dopamine system and Tellegen's PA (Tomarken et al., 1992).

Avoidance. Another major system, an avoidance system, produces the experience of negative emotion and behavioral inhibition or withdrawal. It corresponds to Tellegen, Watson, and Clark's (1999) Negative Activation factor, which represents the extent to which individuals experience actively negative emotions such as distress or fear (Markon, Krueger, & Watson, 2005; Tellegen, 1985; Watson, Wiese, Vaida, & Tellegen, 1999). The earliest, systematic empirical account of the avoidance system may have been provided by Hans Eysenck's (1967, 1983, 1994) Neuroticism factor (N), which is characterized by anxiety, depression, guilt, low self-esteem, tension, irrationality, shyness, moodiness, and emotionality.

Gray (1987a, 1987b, 1988, 1991; Gray & McNaughton, 2000) has provided the most detailed analysis of the avoidance system, which he termed the behavioral inhibition system (BIS), and which seems similar to the trait of anxiety. The BIS is sensitive to signals of punishment, threat, nonreward, and novelty and responds with behavioral inhibition, attention, and arousal, accompanied by anxiety. Gray (1987, 1988, 1991; Gray & McNaughton, 2000) recognized that BIS/anxiety is closely related to Eysenck's Neuroticism factor but maintained that the BIS represents a more fundamental neurological system. (Gray later reformulated his conceptualization of the BIS [Gray & McNaughton, 2000; Smillie, Pickering, & Jackson, 2006], which is covered in more detail in the discussion.)

Unlike many others, Panksepp (1998) does not identify a general system for managing avoidance or sensitivity to threat. Instead, he identifies two systems that manage specific types of threat. The fear system is activated by physical threat. The panic or separation-distress system manages social attachment, such as that between parent and child, or lovers, and is activated by separation or the loss of a close relationship. Fear and panic could form two major components of the more general BIS.

Davidson and colleagues (Davidson, 1992, 1995, 1998, 2001, 2003; Tomarken et al., 1992) demonstrated that chronically high activation in right hemisphere PFC and the amygdala produces an affective style characterized by negative affect, heightened sensitivity to negative stimuli and stress, behavioral inhibition, and withdrawal. This affective style represents a strong avoidance system and corresponds to Gray's BIS and Tellegen and colleagues' (1999) Negative Activation.

Constraint. In addition to two broad approach and avoidance systems, several researchers (e.g., Clark & Watson, 1999; Rothbart & Bates, 1998; Tellegen & Waller, 2008) have argued for a general disinhibition and constraint system that underlies major aspects of personality, particularly self-control and emotion regulation. Clark and Watson (Clark & Watson, 1999, p. 403) proposed that "disinhibition versus constraint" (DvC), a "tendency to behave in an undercontrolled versus overcontrolled manner," is related to serotonin regulation and is an important dimension, along with positive and negative emotionality, underlying personality differences.

> Disinhibited individuals are impulsive and somewhat reckless and are oriented toward the feelings and sensations of the immediate moment; conversely, constrained individuals plan carefully, avoid risk or danger, and are controlled more strongly by the longer term implications of their behavior. (Clark & Watson, 1999, p. 403)

Low serotonin functioning is generally associated with negative outcomes, but the nature of the negative outcome depends on the activation of other systems. Depue and Collins (1999) argued that serotonin functions by exerting "a tonic inhibitory influence on information flow, regardless of the type of information being conveyed" (Clark & Watson, 1999, p. 414). This agrees with Spoont's (1992) argument that higher levels of serotonin increase the signal to noise ratio, thereby focusing activity on a single system and inhibiting competing systems: Low levels of serotonin result in rapid switching from one activity to another, moderate levels stabilize activity around a single system, and high levels produce behavioral rigidity. Thus, individuals with low serotonin functioning should be hypersensitive to a signal. Such individuals might be over reactive to cues to both reward and punishment. For example, both affective aggression, which seems to be related to incentive activation of the BAS, and responses to fear or social rejection cues should be higher in low serotonin individuals. A second consequence of this hypersensitivity might be greater emotional and motivational lability, resulting in somewhat disorganized behavior.

Child temperament researchers (e.g., Derryberry & Rothbart, 1997; Eisenberg, 2002; Eisenberg, Smith, Sadovsky, & Spinrad et al., 2004; Rothbart & Bates, 1998; Kochanska & Knaack, 2003) have argued for a more effortful control system, in which constraint is the result of planful, executive processes. There is debate as to whether this system operates at an automatic level (as seems consistent with Clark and Watson's, 1999, and Depue and Collins, 1999, argument) or is the result of planful, controlled thought (e.g., Rothbart & Bates, 1998; see Carver, 2005, for a discussion of these issues). It has been suggested that both exist (Carver, 2005; Eisenberg et al., 2004).

Rothbart and Bates (1998) argued that these three dimensions have subcomponents that correspond with more specific motivational systems. Neuroticism may have two subcomponents: fearful distress and separation distress (corresponding to Panksepp's fear and panic systems). There is some disagreement about the location of irritable distress or hostility (which may come from blocking the BAS). Rothbart and Bates (1998) and McCrae and Costa (1999) have placed this under Neuroticism, whereas other researchers (e.g., Ashton & Lee, 2007; Lee & Ashton, 2004; Peabody & DeRaad, 2002) have argued that irritable distress or hostility (which may come from blocking the BAS or approach behaviors) belongs to (negative) Agreeableness. Extraversion may have both an energy–activity level component and a sociability component.

Specific motives and traits. Human personality is much more differentiated than just the broad motivational and control systems described above. To capture this differentiation, we model more specific traits in terms of more specific human motives that are organized under the broad motivational systems discussed above. In doing so, we were guided by the following literature.

Traits as goal-based, motive-based structures. L. C. Miller and Read (1987, 1991; see also Read & Miller, 1989) have argued that traits are goal-based structures, represented in terms of the goals, plans, resources, beliefs, and behavioral styles of the type of individual that can be characterized by the trait. Thus, personality can be captured by configurations of goals, plans, resources, and beliefs. For example, the trait *helpful* can be represented in terms of a goal of helping others, plans for achieving that goal, resources needed to achieve the goal, and beliefs related to the goal (e.g., whether one's actions would actually assist the other and whether the other desired assistance).

Mischel and Shoda's (1995) CAPS model also suggests that personality can be understood in terms of similar components, called cognitive-affective units (e.g., goals, plans, expectancies, etc.). One important difference between Read and Miller's (1989, 1998; L. C. Miller & Read, 1991) approach and Mischel and Shoda's (1995) approach is that Read and Miller (1989, 1998; L. C. Miller & Read, 1991) argued that traits can be analyzed in terms of these goal-based components, whereas Mischel and Shoda (1995) do not make a similar claim. Read and Miller (1989, 1998; L. C. Miller & Read, 1991) have explicitly mapped out the relations between their goal-based components and the trait lexicon and personality structure. Mischel and Shoda (1995) have not examined how individual differences in their cognitive-affective units can map onto specific traits or personality structure, such as the Big Five.

We should note that we are not using the term goal as necessarily referring to a consciously represented mental structure. Goals refer to things that people (and animals want), but of which they may not be conscious. For example, both people and animals are interested in mates, food, status, and dominance, and social animals care both about being rejected from the group and about behaving cooperatively. Obviously, some goals are specific to humans, and some can be consciously represented, but that is not necessary. We should also note that most aspects of the current model are compatible with recent work (e.g., Gosling & John, 1999; Gosling, Kwan, & John, 2003) demonstrating considerable cross-species continuity in broad personality dimensions. Our broad approach and avoidance motivational structures are found in all animals, and many of the more specific motives we discuss can be found across a wide range of species.

Evolutionary analyses of social tasks and taxonomies of human motives. The model we propose is heavily motive-based: Individual differences in personality and behavior can be understood largely in terms of the behavior of underlying motivational systems. But what are the basic sets of human motives that underlie personality differences? For our model we drew on two sources of information. First, we drew upon several recent evolutionary analyses of the motivational underpinnings of social tasks. Second, we drew upon a recent extensive taxonomy of human motives (Chulef, Read, & Walsh, 2001).

Several researchers (e.g., Bugental, 2000; Fiske, 1992; Kenrick & Trost, 1997) have argued that human beings have evolved specific brain systems specialized for handling our most important social tasks. They argued, based on both evolutionary and empirical considerations, that a variety of tasks need to be solved by human beings to survive and reproduce, including (a) status, (b) coalition formation, (c) affectional relationships, (d) self-protection, (e) mate choice, (f) parenting, (g) attachment, and (h) play or exploration. Most, if not all, of these motives are shared with other social mammals. Some of these motives are approach related, whereas others are avoidance related. They are among the more specific motives that underlie the two broad motivational systems. A similar set of distinctions can be found in Panksepp's (1998) review of the neurobiology of human (and other animals') motivational systems.

A further source of information about a basic set of human motives is Chulef, Read, and Walsh's (2001) recent cluster analysis of a detailed list of 135 human motives. Although their results provide a more fine-grained structure than the broad evolutionary analyses, their broad clusters of motives generally correspond to the evolutionary tasks.

Agentic and communal/affiliative systems. Another major distinction in our analysis of motives is between agentic and communal motives. Depue and Morrone-Strupinsky (2005; see also Church, 1994; Digman, 1990; Wiggins & Trapnell, 1996) have argued that extraversion is composed of two lower order factors: agency and communion or affiliation. This suggests that the approach system may encompass two general classes of motives: "Affiliation reflects enjoying and valuing close interpersonal bonds, and being warm and affectionate . . . *agency*, reflects social dominance and the enjoyment of leadership roles, assertiveness" (Depue & Morrone-Strupinsky, 2005, pp. 314–315). Depue and Morrone-Strupinsky have outlined a detailed argument for a set of brain mechanisms that underlie a general communal trait. Being responsive and caring towards others is tightly tied to neuromodulators such as oxytocin, prolactin, and brain opiates (Depue & Morrone-Strupinsky, 2005; Panksepp, 1998).

This distinction has been identified in a number of domains. Wiggins and Trapnell (1996) and Digman (1997) have argued that agency and communion are two major distinctions in trait structure. Cross-cultural researchers, such as Triandis (1995) and Hofstede (1980) have argued for a closely related difference between cultures, individualism versus collectivism. This distinction is also apparent in Chulef, Read, and Walsh's (2001) goal taxonomy.

General Description of the Model

Using this literature, we developed a hierarchically organized neural network model of personality. At the lowest level of the model are a number of relatively specific motivational systems that manage different motivational domains and their related behavior (e.g., authority relations, mating, avoiding physical harm, avoiding social rejection). At the next level, two overarching motivational systems, approach and avoidance influence the lower level systems and integrate over them (Cacioppo, Gardner, & Berntson, 1997). Finally, there is a general disinhibition and constraint system that moderates all aspects of the model.

The approach and avoidance systems are modeled by organizing the specific motives into two general motive systems, whose parameters can be independently manipulated to capture broad individual differences in these systems. Disinhibition and constraint is the broadest level of the model; it is a general inhibitory system that moderates activity in both the approach and avoidance systems and the lower level systems that they moderate. The impact of this inhibitory system is to sharpen the differences between more highly and more weakly activated nodes by increasing the signal to noise ratio. Various researchers (e.g., Depue & Collins, 1999; Zald & Depue, 2001) have suggested that its operation is based on the neurotransmitter serotonin. This disinhibition and constraint system is a general control system that moderates the behavior of the fundamental approach and avoidance motivational systems.

The two broad motivational systems in the model should capture two major, broad dimensions of personality: extraversion and neuroticism. Note that in our approach, neuroticism is essentially limited to fearfulness and anxiety.

Each of the two broad motivational systems moderates a number of more specific motives. Among the approach related motives are (a) social bonding, (b) dominance and the development of authority relations in groups, (c) exploration and play, (d) caring and parenting, and (e) mating. Among the avoidance related motives are (a) avoiding physical harm and (b) avoiding social separation. These motives ultimately play a role in a variety of more specific traits and subfactors of personality.

Another structural component cuts across the two broad motivational systems and the more specific motives just discussed. These specific motivations may also be organized into two broader systems: agentic motives and communal motives.

In addition to the central role played by the motivational systems and by the general inhibitory process (disinhibition and constraint), the model has two other critical components. One component provides a representation of the situations to which the individual is responding. Situations are modeled by a feature layer, which represents salient or motive-relevant attributes of the situation. A second component represents the resources that an individual possesses (e.g., sense of humor, money) that are important in the pursuit and attainment of the individual's goals, and which Read and Miller (1989; L. C. Miller & Read, 1987) have argued are key components of traits. Resources are modeled by a resource layer, which represents the presence or absence of various motive relevant resources that the individual directly possesses as part of their person, such as wit or a store of jokes. There are other kinds of resources (e.g., alcohol or a computer), which an individual may possess, but which are considered part of the situation (because they are not part of the person); these are modeled in the feature layer.

A diagram of the neural network model (which will be discussed shortly) can be seen in Figure 1. In our model, individual motives are activated as a result of interactions with situations. The motive activations are a function of (a) the situation, (b) experience (i.e., knowledge and memory), which influences weight strength in the model, and (c) innate individ-

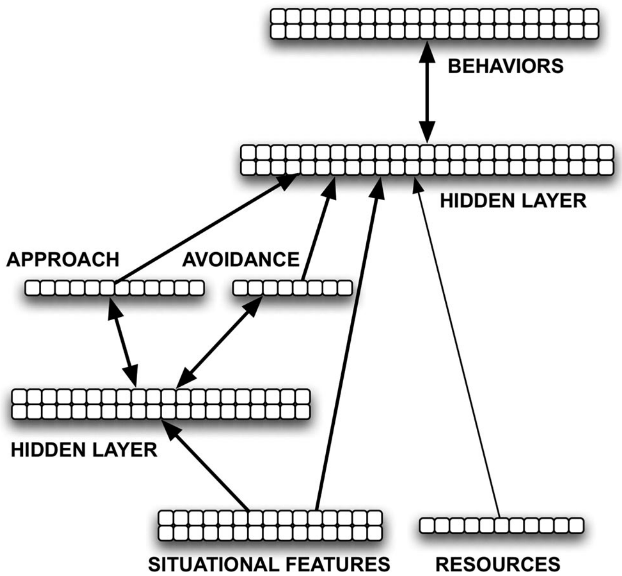

Figure 1. Structure of the new virtual personalities neural network model. Double-headed arrows represent bidirectional connections among layers, which represent feedback relations among nodes and enable the network to function as a recurrent or parallel constraint satisfaction network. Single-headed arrows represent unidirectional connections in the direction of the arrow.

ual differences, represented by (a) baseline motive activations for each specific motive in the model, which can differ from person to person, (b) individual differences in the overall sensitivities of the approach and avoidance systems, which affect the activations of individual motives and have broad impact, and (c) individual differences in the disinhibition and constraint system, which moderate the activity level of the entire system and further focuses (or defocuses) motivations.

Implications. A central claim of this model is that human personality can be understood in terms of a hierarchically organized motivational system. We argue that the structure of human personality is the result of the operation and interaction of the components of this system. Even if some of the specific components of the systems are ultimately revised, we believe that the basic approach of modeling personality as a hierarchy of interacting motivational systems will prove valuable.

A considerable amount of work has been done on the structure of human personality, and there have been various theoretical attempts to describe some of the processes that may be responsible for that structure. However, the accounts of that underlying structure and the potential interactions among different structures have largely been verbal, with the exception of a computational simulation by Pickering (2008) of Gray's revised reinforcement sensitivity theory (Gray & McNaughton, 2000; also described in Smillie, Pickering, and Jackson, 2006). With verbal models it is difficult, if not impossible, to examine the potential interactions among systems. A major advantage of computational models is that they make it possible to explicitly model the interrelations among different systems and to model the dynamics of how these systems may interact with each other over time. In doing so, computational models should help us determine which kinds of systems and relations are plausible.

Relation to our previous virtual personalities model. Our previous virtual personalities model (Read & Miller, 2002) tested some of the basic principles outlined above in a very abstract way, but it was somewhat limited. The previous model (a) had very simplified situational and personality representations, (b) could not learn, (c) had one-to-one feature to goal mappings (each feature linked to only one goal and each goal linked to only one feature), and (d) had one-to-one goal to behavior mappings.

Richer representations of situations and personality. Previously, each situation node corresponded to one abstract type of situation: a social situation, a social rejection situation, a physical threat situation, and so on. Here, we used more realistic features that corresponded to important aspects of two classes of situations: social and work. For example, we used number of people present; presence of music or alcohol; location in a cubicle, a conference room, or a bar; presence of one's boss or co-workers, and so on. This allows the network to represent a variety of specific situational features, as well as configurations of those features.

We also developed more realistic and specific behaviors. For example, behavior nodes correspond to such things as tell jokes, dance, drink alcohol, and work hard. Previously, the behaviors were quite abstract: social behavior, helpful behavior, fearful behavior, and so on.

Learning matters. In our earlier model, the network did not learn; weights were hand coded. Unfortunately, the complexity of hand coding prevented us from modeling realistic representations of situations and personality. However, with learning we could model much more complex and realistic representations. Further, learning allows us to model how early temperament and different socialization experiences affect learning to interpret and respond to different situations.

From one-to-one mappings to learned configural linkages from situations and goals. We also tried to implement more realistic assumptions about the structure and organization of the underlying motivational systems and their relation to the world. In the real world, the same situational feature might activate a range of motives, and a given motive might be activated by different configurations of situational features.

In our earlier model, we reduced that complexity: Each goal was directly activated by one and only one situation. This had several undesirable consequences. First, because of this direct mapping, we could not model how different individuals might learn different representations and interpretations of a situation: Something either was or was not an "avoid social rejection" situation. Second, there was no way to learn representations from configurations of situational features or to learn that such configurations of multiple features might activate a goal. Thus, the previous model could not capture the realistic web of connections among situations and goals and the resulting interactions among them in the enactment of behavior.

Here, each situation is represented by a number of much more concrete situational features. For example, one might have a party situation that would be defined by features for music, alcohol, other people, large room, low lighting, and so on. This had several implications for the current model. First, the activation of a goal should be the result of input from a configuration of situational features. Second, the same feature could be related to different goals.

From single source to multisource activation from goals: Implications for behavior. An additional important difference between the models is in how behaviors are activated. Previously, each behavior received input only from a single corresponding goal (which in turn only received activation from a single situation). However, here, a behavior is activated by inputs from a configuration of multiple goals, situational features, and resources, all of which go through a hidden layer to the behavior layer. Thus, the triggering of a behavior is the result of a much more complex set of inputs. This more complex model allows us to capture more aspects of L. C. Miller and Read's (1987; Read & Miller, 1989) model of traits, in which they argued that traits consist of configurations of goals, plans, resources, and beliefs (also see Mischel and Shoda's (1995) cognitive affective units). The structure of the previous model was equivalent to arguing that only goals played a role in the representation of traits or were responsible for behavior.

Further, the previous version of the model did not allow for the possibility of either equifinality (a goal can be achieved by different behaviors) or multifinality (a single behavior can achieve multiple goals). The current model allows for both possibilities.

Introducing hidden layers: Implications for learning conjunctive representations. We introduced hidden layers between the different layers, with a hidden layer between the situational features and the goal layer and between all other layers and the behavior layer. Hidden layers add much greater computational power to a network, allowing it to learn conjunctions of inputs that were important in predicting behavior. One result is that the model can now learn to represent situations and traits in terms of configurations of features, and behavior will be a result of those configurations.

Technical Description of the Current Model

Neural network models have three important components: the network architecture (how the nodes in the network are organized and connected, and how activation flows), the activation updating function for the nodes, and the learning rule for weights between nodes (Bechtel & Abrahamsen, 1991; McClelland & Rumelhart, 1986).

Basic network architecture. The basic architecture can be seen in Figure 1, and a description of the different nodes in the network is in Tables 1A–1D. There are two input layers: a situational feature layer and a resource layer. The situational feature layer defines the different situations to which the model responds. It uses a localist representation (each node corresponds to a single feature) consisting of 29 features (See Table 1 for a list of the features) that are used to specify various specific situations in two general contexts: work and parties.

The resource layer represents personal resources (e.g., wit, money, etc.) that are directly possessed by or a part of an individual (see Table 2 for a list of the 11 resources used). Things that may be present in the environment but that are not part of an individual are represented in the situational feature layer. We view personal resources as a component of personality that should influence the kinds of behaviors that are most likely to be enacted (Read & Miller, 1989).

The situational feature layer connects, through a hidden layer, to two goal layers: an approach layer (12 goals) and an avoidance layer (8 goals; see Table 3 for a list of the goals). Organizing the goals into two separate layers represents the idea that the approach and avoidance systems are functionally separate systems with different characteristics. Two layers also make it easier to set parameters differently for the two systems and to model inhibition separately within the two goal systems. The goals in the goal layers are guided by the evolutionary analyses mentioned previously (e.g., Bugental, 2000; Fiske, 1992; Kenrick & Trost, 1997) and by our recent taxonomy of human goals (Chulef, Read, & Walsh, 2001). We note, however, that the field does not have a consensually agreed on taxonomy of human motives.

Table 1
Situational Feature List

Situational features	Situational features
At home	At party
In conference room	At work
Urgent work	Work to do
With friends	With strangers
With subordinates	With disliked acquaintance
With boss	In break room
In office/at desk	Conflict situation
TV	With potential date
With relatives	With date
With kids	At restaurant
With 0 others	At bar
With 1 other	Alcohol
With 2 or more others	Dancing
With romantic partner	Wedding/formal party
Age difference >7 years	

The situational features, the goal layer, and the resource layer are all directly connected to a hidden layer, which then connects to the behavior layer, which has a variety (43) of behaviors that can be enacted in work and party settings (e.g., give orders, work extra hard, dance, drink alcohol, tell jokes, etc.; Table 4). The situational features, resources, and behaviors in the model are not intended to be representative but rather to demonstrate the plausibility of our approach with relatively realistic representations. True representativeness is not possible as there are no consensually accepted descriptions of representative situations and situational features.

The hidden layers mediate the transformation from input to output layers. Networks consisting of one or more hidden layers can perform more complex transformations than can those without hidden layers. The hidden layers should allow the network to learn to respond to conjunctions of inputs, such as conjunctions of situational features, goals, and personal resources.

In Figure 1, double-headed arrows represent bidirectional connections among layers, which represent feedback relations among nodes and enable the network to function as a recurrent or parallel constraint satisfaction network. Single-headed arrows represent unidirectional connections in the direction of the arrow.

Each individual node has a bias (not shown), representing the degree of baseline or chronic activation. Biases are implemented by a weight from a node set to a constant activation of 1, for which the weight represents the extent to which the node's activation is influenced by the bias. Bias weights can change during learning.

The Leabra implementation. We use the Leabra (local, error-driven and associative, biologically realistic algorithm) implementation framework (O'Reilly & Munakata, 2000) in the PDP++ program. Leabra is designed as a biologically realistic architecture. We chose the Leabra architecture because it allowed us to model several important processes in our theoretical model of the dynamics of motivational systems: chronic activation of mo-

Table 2
Resource List

Resources	Resources
Social skills	Things to talk about
Job skills	Money
Face saving skills	Attention span
Quick thinking	Status
Wit	Time
Intelligence	

tives, differences in sensitivities to inputs of individual motives and motivational systems (approach and avoidance), and inhibitory processes. However, we are not claiming that these features of our theoretical model are necessarily realized by the specific neurobiological processes modeled by Leabra. Leabra has several interesting claims about the possible biological implementation of these features and raises interesting hypotheses for future research. But our major interest in Leabra was that it allowed us to relatively easily implement many of the key theoretical characteristics of our model. Our key theoretical assumptions do not depend on whether certain aspects of the neurobiological bases of Leabra are correct. We consider each component in more detail below and describe how features of the Leabra architecture map onto features of our personality model.

The activation function. The activation function in Leabra results in an S- or sigmoid-shaped output activation. Sigmoidal functions are common in computational modeling because they have a minimum and maximum activation, consistent with this property of real neurons, and because such functions are more computationally powerful than are linear activation functions (Hertz, Krogh, & Palmer, 1991; Read, Vanman, & Miller, 1997; Rumelhart & McClelland, 1986).

Possible output activation ranges from 0 to 1. As the output activation can be thought of as representing the summed firing frequency of a neuron, a node cannot have a negative activation. The activation function in Leabra is intended as a useful approximation of key electrophysiological properties of a real neuron, but one that is computationally tractable.

Based on the electrophysiological properties of real neurons, the Leabra activation function includes three types of currents or channels in the neuron, also referred to as ionic conductances. There is an excitatory conductance (e), controlled by excitatory neurotransmitters, such as dopamine or norepinephrine, an inhibitory conductance (i), controlled by inhibitory neurotransmitters, such as gamma-aminobutyric acid (GABA), and a leak current (l), which is a constant process due to the constant flow of ions in and out of the neuron. This distinction between different channels allows us to capture individual differences in the role of different neurotransmitter systems.

Again, we emphasize that we are not making strong claims about the role of actual neural conductances as the basis of individual differences in personality. Rather, we are using the conductances as a convenient way of implementing individual differences in sensitivity to input and individual differences in general inhibitory processes. Although Leabra raises interesting hypotheses about how these individual differences might be neurobiologically implemented, our theoretical model does not depend on this implementation.

Output activation in Leabra is calculated in two steps. First, the activation level or membrane potential of the neuron (V_m) is calculated as a function of its current activation plus its input activation and then the output activation of the neuron is calculated as a function of the difference between the V_m and its firing threshold.

The change in V_m is calculated from the three conductances and their corresponding reversal potentials E as

$$\Delta V_m(t) = \tau \sum_c g_c(t)\overline{g_c}(E_c - V_m(t)). \tag{1}$$

There are three channels or conductances (c): e, the excitatory input, i, the inhibitory input, and l, the leak. Each input is separated into two components, the current input level (g_c) and g_{bar_c}, which indicates the relative number of channels for each of the types of current. g_{bar_c} can be thought of as the proportion of influence for each current. So the change in V_m is a function of the sum of each conductance times the difference between the current V_m and the reversal potential for that channel.

The excitatory net input conductance $g_e(t)$ can be treated as the proportion of excitatory channels that are open and computed as a function of sending activations times the weights:

$$g_e(t) = \langle x_i w_{ij} \rangle = \frac{1}{n}\sum_i x_i w_{ij}. \tag{2}$$

This term largely corresponds to the standard way of computing input activations. However, in Leabra, unlike most neural networks, the input activation is standardized by the number of input connections (1/n). The input from any bias node is also added to this term.

The inhibitory conductance is computed by the k-winners-take-all (kWTA) function described in the following, and leak is a parameter set in the model. As we discuss in more detail, kWTA inhibition is operationalized by this inhibitory current in the activation function for the neurons.

Table 3
Goal List

Approach goals		Avoidance goals	
Friendship	Mastery	Avoid rejection and embarrassment	Avoid loss of control
Sex/romance	Exploring	Avoid guilt	Avoid interpersonal conflict
Be liked	Fun	Avoid failure	Avoid effort
Help others	Fairness–equality–justice	Avoid physical harm	Avoid risk and uncertainty
Dominance	Uniqueness		
Achievement	Material gain		

Table 4
Behavior List

Behaviors	Behaviors	Behaviors	Behaviors
Eat and drink	Stay at periphery	Help others with work	Ensure work distributed fairly
Drink alcohol	Self-disclose	Order others what to do	Wear something distinctive
Relax	Ask others about self	Dance	Steal
Play practical joke	Talk politics	Ask other to dance	Kiss up
Tease and make fun of	Gossip and talk about others	Ask for date	Be cheap
Try new dance steps	Talk about work (job-related)	Kiss	Mediate
Intro self to others	Tell jokes	Do job	Give in
Surf web	Compliment others	Extra effort job	Procrastinate
Explore environment	Ignore others	Find new way to do job	Pretend to work
Leave	Insult others	Improve skills	Stay with comfortable others
Be silent	Clean up	Confront other about slacking	

Once V_m is calculated, the output activation of the node is calculated as

$$y_j = \frac{\gamma[V_m - \Theta]_+}{\gamma[V_m - \Theta]_+ + 1}. \tag{3}$$

V_m represents the membrane potential, Θ is the threshold for the firing of the node, and V_m must exceed the threshold for the node to fire. γ is the gain parameter and represents the sensitivity of the node to its inputs, once the firing threshold is passed. Finally, the + subscript indicates that the terms in the brackets are set to 0 if the term is less than 0.

Several aspects of these two equations are important in our model. First, the conductances (g_{bar_c}; the relative number of channels) and the gain parameter (γ) capture differences in the sensitivity of a node. Higher conductances and gains lead to more steeply accelerated activation functions; they result in higher output for the same input. One important difference between these parameters is that the gain parameter (γ) only affects activations that exceed threshold [Θ]), whereas conductances also affect below threshold activation.

Second, each node has a firing threshold, such that V_m must exceed the threshold Θ before the node can fire. This controls for the effect of random noise as well as allowing for control of the sensitivity of the node. The likelihood of firing can be influenced by both the threshold of the node [Θ]) and by its baseline activation (which contributes to V_m).

Although these parameters provide a convenient way to model key theoretical aspects of our model, such as baseline activation of motives, differences in sensitivity to inputs, and degree of inhibition, our theoretical model does not depend on this neurobiological implementation. Our central concern is capturing the theoretical and functional aspects of our model of personality.

A final aspect of the behavior of a node that is important (although it is not represented in the activation function) is accommodation. Accommodation can be thought of as a node "fatiguing" after constant firing. One commonly cited example of this is the observation that if you continually pronounce a word, it starts to lose its meaning. We wanted to capture the idea that after enacting the same behavior for a while, this behavior and associated goals will start to fatigue, and the activation of other behaviors and goals will increase and thus be enacted.

Inhibition. A fundamental aspect of Leabra is a general mechanism for inhibition. We were interested in this for several reasons. First, we wanted to capture the idea that goals compete for control of behavior; when the organism is actively pursuing one goal, competing goals should be inhibited. Second, having a global inhibitory field seemed a useful way to implement aspects of personality, such as conscientiousness and disinhibition and constraint. The result of a strong inhibitory field is that only the most strongly activated nodes can fire. Such inhibitory processes allow for selectivity and focus in a variety of important cognitive processes (Martindale, 1991; Nigg, 2000).

The inhibition function we use is a version of the kWTA algorithm (Majani, Erlarson, & Abu-Mostafa, 1989). Leabra has two variations of this algorithm. The standard kWTA algorithm allows no more than *k* nodes out of a total of *n* (in a layer) to become active at a given time, whereas average kWTA allows, on average, *k* nodes to be active. The average kWTA algorithm allows more than *k* nodes to be active if the input activations are sufficiently strong, whereas strict kWTA will not. In Leabra, inhibition is set for each individual layer. Thus, different layers can have different maximum numbers of nodes active. Inhibition is implemented in the kWTA algorithm by manipulating the inhibitory input in the equation presented above for calculating the V_m.

Inhibition could be captured by using a large number of inhibitory neurons. However, this incurs a large computational cost in large models, so O'Reilly and Munakata (2000) developed this version of kWTA, which achieves the same outcome at much less cost.

Learning. Learning in Leabra combines two different forms of learning: associative, Hebbian learning and error-correcting. The associative, Hebbian form enables the network to be sensitive to the correlational or statistical structure of the inputs, whereas error-correcting learning enables the network to capture a specific task structure (whether a certain output is right or wrong). Error-correcting learning also allows us to model reward and punishment learning (O'Reilly & Munakata, 2000). Differential sensitivity to actual and potential reward and punishment is an important aspect of individual differences in personality.

Error-correcting learning in Leabra is similar, although not identical to, the better known delta rule (Widrow & Hoff, 1960). It uses a contrastive Hebbian learning (CHL) algorithm developed for the Boltzmann machine and then generalized by O'Reilly (1996). This algorithm compares the activation of the network in a plus phase (when both inputs and desired outputs are presented to the network) with its activation in a minus phase (when only the inputs are presented). CHL then adjusts weights to reduce the difference in activation between the two phases. One way to think about what CHL is doing is that in the minus phase, the network is using its inputs to predict its outputs, and the plus phase represents what the output activations actually should be.

As O'Reilly and Munakata (2000) note, in some situations, correlational structure and task structure are different kinds of information, and combining the two kinds of information provides a more powerful learning mechanism. They are combined as a weighted average of the weight change calculated by each learning rule. It is important to note that the error-correcting component in Leabra enables learning in multilayer networks with hidden units. The interested reader can find a detailed description of these learning mechanisms in O'Reilly and Munakata (2000).

Relation of Model Parameters to Individual Differences

We used Leabra because its features allow us to directly implement key aspects of our theoretical model, such as the role of individual differences in inhibition, individual differences in the sensitivity and chronic activation of different layers (especially the approach and avoidance layers), and individual differences in the baseline activation of motives. It also provides a flexible learning rule (CHL) that is more biologically plausible then back-propagation. However, our theoretical model does not depend on the specific neurobiological implementation.

In our model, the following individual differences can be modeled in terms of specific features of Leabra. First, the gains and conductances of nodes determine their sensitivity to inputs. Thus, manipulating the gains and conductances of nodes in the approach and avoidance layers should allow us to model individual differences in reward sensitivity (BAS) and punishment sensitivity (BIS). Second, the CHL rule allows us to model individual differences in learning in response to rewards and punishments; for example how differences in temperamental sensitivity might influence how individuals learn to respond differently to the same stimuli. These two features should allow us to model important aspects of extraversion and neuroticism. Third, the bias weights of motives in the approach and avoidance layers can be used to model the chronic or baseline activation of those motives and, thus, the extent to which they guide behavior. This allows us to model individual differences in the importance of specific motives in the representation of specific traits. These bias weights and the baseline activation parameter V_m_init could also be used to model differences between sets of motives within the approach and avoidance layers, such as the differences between agentic and communal goals. Fourth, manipulating the degree of inhibition within different layers, such as the approach, avoidance, and behavior layers (as operationalized by kWTA) will influence the selectivity and focus of the network. This should allow the network to capture certain aspects of the general factor of disinhibition and constraint (also related to aspects of conscientiousness and impulsivity). Consistent with suggestions by researchers such as Depue (1996) and Zuckerman (2005), this may allow the network to capture aspects of the role of serotonin in personality.

Training Procedure for the Personality Model

Our model first needed to be trained. The network had five different meaningful layers (situational features, resources, approach and avoidance goals, and behaviors). We did not just train the pairing between the initial inputs (situational features and resources) and the final outputs (behaviors) because that would not insure that the network learned all the appropriate relations, especially for such things as situational feature to goal pairings. To insure learning of relations among the layers, we trained five different pairs of relations among different layers (e.g., features to goals, goals to behavior). An individual training event consisted of active units in a layer that served as the input and an active unit in a layer that served as the output. The individual training events paired one input layer with one output layer, with exceptions noted below. The model was trained on input–output patterns for each of the following layer pairings: situations–goals, situations–behaviors, goals–behaviors, and resources–behaviors. In addition, the model was trained on a set of events that involved pairing the simultaneous activation of three input layers (situations + goals + resources) with the behavior layer.

We should note that Leabra allows one to train relations among any arbitrarily chosen pair of layers or even among multiple layers. Although the goal layers mediate between the situational features and the behavior layer, we believe that individuals do have access to aspects of the current state of the goal systems. People are often aware of how much they want different things. Thus, the activation of goals can serve both as target activations and input activations.

One can view learning in this context as "wiring up" the network in a reasonable way. We used this technique to have the network associate situational features with the goal affordances of situations and to associate the goal affordances of situations with relevant behaviors. We assume that weight strength is a function of experience and hard-wired biases.

Definitions of the training events were determined by a consensus of the authors. For the situational feature layer, this meant creating a set of situations, from conjunctions of input features, within which the agent would act. These training situations and their features can be seen in Tables 1–5. Outputs were then paired with these situations, to create the training events involving situational inputs. For the situation–goal events, for each particular situation there was a single goal output. So for example, the training events that paired the situation "taking a break (at work) by yourself" with its associated goals consisted of four active units on the feature (input) layer, "in the break room," "at the workplace," "with work to do," and "with zero others/alone," and one active unit per event in the goal layers, for example, either "have fun" or "avoid effort." For the situation–behavior events, an input situation was paired with one active output (behavior) unit at a time. For goal–behavior events, one input (goal) was paired with one output (behavior), and for resource-behavior events, one input (resource) was paired with one output (behavior). For the situation + goals + resources events, multiple units could be active in each of the three input layers, and these input combinations were paired with one output behavior.

We also wanted the network to be roughly sensitive to the relative frequency with which things co-occur in the environment, which should be represented by weight strength. To capture this, we varied the relative frequency of different features and their frequency of co-occurrence. Estimates of the relative frequency of the inputs and their frequency of co-occurrence were based on the judgments of Stephen J. Read, Brian M. Monroe, Aaron L. Brownstein, Yu Yang, and Gurveen Chopra. We first independently rated the relative frequency of occurrence and co-occurrence of the different inputs and outputs and then discussed the ratings until we reached consensus. For example, we judged the situation "party in a restaurant/bar" to be approximately twice as common as the situation "family birthday party." Thus, there were twice as many of the former events, compared with the latter.

Again, the important point here was to come up with a plausible rather than a clearly representative frequency distribution. A comprehensive description of the frequency of occurrence and co-occurrence of social situations and behaviors does not exist.

Because each of the inputs varied with respect to the number of outputs associated with it (for example, there were more behaviors associated with the goal "have fun" than with behaviors associated with "avoid guilt"), inputs with more outputs would have fewer instances per output than those with relatively few outputs, in order to keep the total frequency of the input in line with our judgments. Thus, for example, it was judged that the goal "have fun" would appear approximately 2.5 times as often as "avoid guilt." But because "have fun" has 12 behaviors associated with it, and "avoid guilt" only two, each behavior with "have fun" has to share the total time with many more behaviors, so in the end each "have fun" behavior pairing was seen less often than each "avoid guilt" behavior pairing. Specifically, the pairing of "have fun" with "tell jokes" would be seen seven times, whereas the pairing of "avoid guilt" with "leave" would be seen 17 times. This is appropriate because if more things are associated with an input, each will be less strongly associated with the input. Because there were 11 other behaviors paired with "have fun," each would be seen seven times, and the goal "have fun" would be seen a total of 84 times (12 behaviors $\times$ 7 replications), versus 34 total events for "avoid guilt" (2 behaviors $\times$ 17 replications), thus preserving the overall relative frequency of those goals.

There were 356 unique training events, each one replicated for frequency considerations so that there were 4,642 events seen per training epoch. During each epoch, the order of events was randomly permuted. Training proceeded for 25 epochs; testing over a range from five to 500 epochs showed no significant increases in performance after about 25 epochs.

Testing the Model

We first tested whether the model learned the relationships in the training events and generated reasonable behaviors in response to relatively novel situational inputs. These test whether the model learns and behaves appropriately but not whether it can model personality. We then present a series of simulations that test how well the model explains various aspects of the structure and dynamics of personality.

Basic validation of the network.

Validation 1: How well did the model learn the relations between situations and behaviors? After training the model as described above, we tested it by using the 15 different situations from the training phase as inputs (See Table 5). During testing, we turned on accommodation, which can be thought of as the fatiguing of a node, allowing other nodes to become active. For each input we recorded both the behavior with the highest initial activation and the subsequent two behaviors activated. These behaviors were then coded for correctness in terms of whether they had been paired with that situation during training. This was done for five random initializations of the starting weights.

Results. The model performed well (See Table 6). Looking only at the first behavior, over 93% of the time it generated a behavior that was paired with the situation during training. A random response would result in an appropriate response somewhat less than 20% of the time. And of the first 3 behaviors

Table 5
Fifteen Original Situation Descriptions

Situation	Features
Individual assignment	At work; in office; with 0 others; work to do; urgent
Working with one other	At work; in office; at desk; work to do; urgent work; with 1other
Working together on urgent project	At work; conference room; in an office; conflict situation; work to do; urgent; with 2 or more others; with subordinates; with disliked acquaintance
At a group meeting	At work; in an office; conference room; conflict situation; work to do; with 2 or more others; with friends; with boss; with disliked acquaintance
Review with boss	At work; with boss; office; conflict situation; urgent
Taking a break with coworkers	In break room; at work; TV; work to do; with 2 or more others; with friends
Taking a break by yourself	In break room; at work; in office; at desk; work to do; with 0 others
Party at work	Party; conference room; work; alcohol; work to do; with 2 or more others; with friends; with boss; with subordinates; difference >7 years
Social engagement At boss' house	With boss; with strangers; with disliked acquaintance; with friends; with 2 or more others; conflict situation; with subordinates; wit hromantic partner
Dance	Dancing; with friends; with potential date; with strangers; with 2 or more others; alcohol
Trying to get a date	Party; restaurant; alcohol; with 1 other; with potential date
On a date	Restaurant; alcohol; with 1 other; with date
Family birthday party	Home; party; with 2 or more others; with romantic partner; with relatives; with kids; age differences >7
Wedding party at a fancy restaurant	Party; wedding or formal party; restaurant; dancing; alcohol; with 2 or more others; with friends; with romantic partner; with kids; age differences >7
Party in a restaurant that has a bar	Party; bar; restaurant; dancing; alcohol; with 2 or more others; with friends; with strangers; with potential date

generated, 72% of the time the model generated a behavior that was initially paired with the situation during training, again a rate much greater than chance.

Validation 2: Sensitivity of the model to changes in situational features. We examined the sensitivity of the model to minor changes in situational features by creating three variations of each of the 15 training situations (see Table 5). A typical variation would add one or two features and remove one or two. After training on the original 15 situations, the model was then tested on a random sequence of the 45 modified situations, using accommodation. We recorded the first three behaviors that were activated in response to each situation and coded whether the modified situation activated the behavior that corresponded to the original unmodified situation. Five random initializations were trained and tested as above.

Results. The model was fairly robust, with little decline in performance (See Table 6). For the first behavior generated in response to the modified situations, 92% of the time the model generated a behavior that was paired with the matching original situation during learning and for the first three behaviors generated, 74.4% of the time an appropriate behavior was generated.

Personality simulations. The model successfully learned. We now examine its ability to capture major aspects of human personality. First, in Simulations 1–4, we tested the ability of the model to capture several aspects of the broad factors of extraversion and neuroticism, by manipulating different parameters of the approach and avoidance layers. In Simulation 1, varying the sensitivity of the approach and avoidance layers to situational cues allows us to capture broad aspects of the impact of extraversion and neuroticism on behavior. In Simulation 2, we showed how recent work by Cacioppo, Gardner, and Berntson (1997) on positive and negative evaluative systems, could provide a more detailed model of extraversion and neuroticism, by helping to better specify the parameters of the two motive systems: a positivity offset (approach layer) and negativity bias or gain (avoidance layer). In Simulation 3, we show that other aspects of extraversion and neuroticism can be modeled in terms of differences in strength of learning in response to the reward and punishment value of situations. In Simulation 4, we examined whether extraversion (approach) and neuroticism (avoidance) necessarily have orthogonal influences on behavior or whether they could have an interactive, nonlinear impact. Following this, in Simulation 5 we show that the general factor of disinhibition and constraint can be modeled by differences in the strength of inhibition among motives in the approach and avoidance layers. In Simulation 6, we show that the impact of individual traits on behavior can be modeled in terms of patterns of individual differences in the chronic activation of specific motives and personal resources. Further, in Simulations 6 and 7, we demonstrate how the model can capture person–situation contingencies. Simulation 7 also addresses Fleeson's (2001, 2007) work showing that the magnitude of intraindividual variability in personality related behaviors is equivalent to the magnitude of interindividual variability in personality traits. Finally, in Simulation 8 we show that the model can capture aspects of other specific traits, specifically Downey's concept of rejection sensitivity (Downey & Feldman, 1996; Romero-Canyas & Downey, 2005).

Table 6
Results of the Validations

Validation	Correct behaviors	Errors	Percentage correct
1			
First behaviors	75	5	93.3%
Total behaviors	311	87	72.0%
2			
First behaviors	225	18	92.0%
Total behaviors	891	228	74.4%

Simulation 1: Extraversion and Neuroticism: Impact of differences in sensitivity of the approach and avoidance layers on approach and avoidance behaviors. We tested whether manipulating the sensitivity of the motive layers enables us to model individual differences in approach and avoidance motivation. We manipulated sensitivity with the excitatory conductance parameter (g_bar_e) on the motives in the approach and avoidance layers after training but before testing. Higher excitatory conductances have a multiplicative effect on the strength of inputs and should lead to increased impact of those motives on behavior.

Gains were set at 100 on both layers for all simulations. The comparison condition was one of the initializations of Test 1 and had default excitatory conductances of 1 for both layers. Four other conditions were constructed with excitatory conductances as follows (approach listed first): 1.2 and 0.8, 1.1 and 0.9, 0.9 and 1.1, and 0.8 and 1.2.

We coded a behavior as approach versus avoidance on the basis of the relative number of approach versus avoidance goals with which it was paired during training. Twenty-eight of the behaviors were related only to approach goals during training, five were related to more approach goals than avoidance goals, seven were related only to avoidance goals, and one was related to more avoidance than approach goals. Finally, one behavior was paired with one approach and one avoidance goal, and one behavior was not paired with any goals.

With accommodation on the behavior layer, we input each of the 15 training situations, recorded the first 3 behaviors generated in response to each situation, and then coded the number of approach and avoidance behaviors activated in different conditions. We also recorded the average activation level for each layer, averaged across the 15 situations.

Results. The average activation of the two layers was strongly influenced by their sensitivities (see Table 7). For the approach layer, the average activation ranged from 0.000 when the conductance was 0.8 to 0.200 when the conductance was 1.2. For the avoidance layer, the average activation ranged from 0.123 when the conductance was 0.8 to 0.234 when conductance was 1.2.

Further, the relative sensitivities had a strong impact on the likelihood of generating approach and avoidance related behaviors (See Table 7). When the approach excitatory conductance was 1.2 and avoidance was .8, 87% of the behaviors were approach related, whereas when the excitatory conductance was 0.8 for approach and 1.2 for avoidance, 44% of the behaviors were approach related. Since so many more of the behaviors were approach related, it is not surprising that even when the avoidant layer had a higher conductance, many of the behaviors were approach related.

The impact of the relative sensitivities of the two motivational layers on the relative frequency of approach and avoidance related behaviors is consistent with a large literature on the impact of

Table 7
Results of Simulation 1: Effect of Differences in Sensitivity of Approach and Avoidance Systems

BAS *g* _bar_e	BIS *g* _bar_e	Number of approach behaviors	Proportion of approach behaviors	Number of avoidance behaviors	Average BAS activation	Average BIS activation
0.8	1.2	20, 1 N	44%	24	.000	.234
0.9	1.1	25, 4 N	56%	16	.005	.220
1.0	1.0	22, 4 N	49%	19	.061	.200
1.1	0.9	35, 1 N	78%	9	.142	.172
1.2	0.8	39	87%	6	.200	.123

Note. N stands for behaviors that are equally associated with approach and avoidance goals. BIS = behavioral inhibition systems; BAS = behavioral approach systems.

extraversion and neuroticism on approach and avoidance related behaviors. Individuals high on Neuroticism are more likely to be socially withdrawn (Clark & Watson, 2008; John, Nauman, & Soto, 2008; Widiger & Smith, 2008), whereas individuals high on Extraversion are more likely to be socially involved and spend more time with people (Mehl, Gosling, & Pennebaker 2006; Watson & Clark, 1997). Further, Avila and Torrubia (2008) reviewed literature indicating that individuals high in avoidance are more vigilant to threat related stimuli, whereas individuals high in Approach are more vigilant to cues indicating reward.

Gable and her colleagues (Gable, Reis, & Elliot, 2000; Elliot, Gable, & Mapes, 2006) have extensively examined the impact of approach and avoidance social goals on social interaction. Gable, Reis and Elliot (2000), using Carver and White's (1994) BIS–BAS scale, found that BAS predicted the frequency of positive social events, although BIS did not predict the frequency of negative social events. Elliot, Gable, and Mapes (2006) measured specific social approach and avoidance goals (not general BIS and BAS). They found that friendship approach goals positively predicted frequency of positive events, whereas friendship avoidance goals positively predicted frequency of negative events. Further, friendship approach goals negatively predicted frequency of negative events, whereas friendship avoidance goals did not predict frequency of positive events.

Simulation 2: Positivity offset and negativity bias in the approach and avoidance systems. Recent work by Cacioppo, Gardner, and Berntson (1997) suggested that the approach and avoidance systems might differ both in their sensitivities and in their baseline activations. They argued that positive and negative evaluation do not form a single bipolar dimension, but instead should be conceptualized in terms of separate dimensions and separate systems for positive and negative activation (also see Cacioppo, Gardner, & Berntson, 1999; Ito & Cacioppo, 2000, 2001). This is consistent with our model's separate approach and avoidance layers.

They further argue that the two systems have different functional forms, with a positivity offset and a negativity bias. Positivity offset means that the positive evaluation system has a somewhat higher baseline activation than does the negative evaluation system. Thus, in the presence of weak or no inputs, the positive evaluation system will be more highly activated than the negative evaluation system. Negativity bias means that the negative evaluation system is much more sensitive to input. Each unit of input to the negative evaluation system results in a higher level of output than does the same unit of input to the positive evaluation system.

Thus, in the absence of strong situational cues, there will be a mild positive evaluation and a tendency to approach or explore interesting things. However, as the strength of situational cues increases, the negativity bias means that the negative evaluation system will respond more strongly to negative cues than the positive evaluation system will to equally strong positive cues. One can visualize this in terms of what happens as an individual moves closer to a goal object with both rewarding and punishing features (for example, in classic work on approach–avoidance conflicts by N. E. Miller, 1959, a rat approaches a goal object with both positive and negative features.) When the individual is far away, the inputs from both positive and negative cues will be fairly weak, and the positivity offset should lead to stronger approach motivation than avoidance motivation. But as the individual gets closer to the goal and the inputs get stronger, at some point the negativity bias results in the avoidance motivation being stronger.

To simulate positivity offset, we set the baseline activation of the approach layer higher than that of the avoidance layer, by changing the resting V_{m} of nodes in a layer, using the V_{m}_init parameter, which controls the resting V_{m} (the default activation of the node in the absence of specific inputs). In Leabra, nodes send activation only when the V_{m} exceeds the threshold value θ (see Equation 1). Thus, if the threshold is constant, higher baseline activations reduce the distance to threshold, reducing the input needed to fire. In the current simulations, the V_{m} _init for the approach layer was .27, and for the avoidance, layer it was .20, with the threshold (θ) for the goal layers set at .3.

To capture negativity bias for avoidance, we manipulated the maximum excitatory conductance $\bar{g}_e$ for the avoidance nodes. This parameter has a multiplicative effect on the impact of the incoming excitation. Conductance $\bar{g}_e$ applies to the entire range of excitatory inputs, affects the V_{m} of the node, and thus, can influence the likelihood that a node will exceed threshold and fire. In this simulation, $\bar{g}_e$ for the approach layer was 0.9 and for the avoidance layer, 1.2. We used a gain of 100 for both motivational systems.

We manipulated the strength of the input cues to the two goal layers by scaling the strength of the weights from the situational layer to the goal layers, with a parameter in Leabra for scaling the strength of the links between two layers. We took the hidden layer between the situational inputs and the goal layers and varied the weight scaling from this hidden layer to the goal layer across three levels: 0.5, 1, and 1.5. The scaling factor is multiplied by the existing weights to give the acting weights. (This is simply a way to capture the effect of getting closer to a stimulus on the strength of the stimulus cues. This does not have any theoretical relevance.)

We presented the network with each of the 15 test situations and recorded the number of avoid behaviors that were activated in response to each weight scaling. Here, as in the previous simulation, an avoid behavior is defined as a behavior that was paired with more avoid goals than approach goals during initial learning.

Results. With a weight scaling of .5, one out of 15 behaviors were avoid behaviors, with a weight scaling of 1, seven out of 15 behaviors were avoid behaviors, and with a weight scaling of 1.5, seven out of 15 behaviors were avoid behaviors. Thus, as the strength of the inputs increases, the relative impact of the avoid layer also increases. This is consistent with Cacioppo, Gardner, and Berntson's (1997) characterization of the positive and negative evaluation systems.

Ito and Cacioppo (2005) have recently provided evidence for individual differences in the strength of both positivity offset and negativity bias and for their impact on the evaluation of social stimuli. Individuals with higher positivity offset formed more positive impressions of a neutrally described target, whereas individuals with stronger negativity bias formed more negative impressions of negatively described targets.

Simulation 3: Extraversion and Neuroticism: Impact of individual differences in reward and punishment values on learning approach and avoidance behaviors. Individuals differ in the extent to which they experience the same event as rewarding or punishing. This should influence learning and lead to individual differences in the likelihood of approach and avoidance oriented behaviors. In a large literature on reinforcement learning (e.g., Sutton & Barto, 1998), researchers have examined both the neurobiological and computational foundations of such learning (e.g., O'Doherty, Buchanan, Seymour, Raymond, & Dolan, 2006; O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003; Schönberg, Daw, Joel, & O'Doherty, 2007; Schultz, Dayan, & Montague, 1997). The idea is that the organism learns to predict whether outcomes will be rewarding or punishing by using prediction error, the difference between the predicted reward or punishment and the actual outcome, to update its prediction. Moreover, a growing body of work indicates that firing levels of dopaminergic neurons indicate the level of prediction error for reward. How punishment prediction error is signaled is very much an open question.

To examine the impact of differential reward and punishment sensitivity, we manipulated the magnitude of reward and punishment value during initial training by varying the training activations for target goals in the approach and avoidance layers in three simulations: in Simulation 1, training values were 1.0 for approach goals and .4 for avoidance goals, in Simulation 2, training values were .4 for approach goals and 1 for avoidance goals, and in a baseline simulation, approach and avoidance goal values were both set at .7.

To enable the network to be sufficiently sensitive to the changes in reward and punishment values, we set the gain for both layers to 20 rather than 100 as in the preceding simulations. Experimentation indicated that with higher gains the model was less sensitive to differences in reward and punishment value. This evidently occurs because higher gains tend to make the nodes behave more like binary neurons rather than have a continuous value.

To test the network, we used the 15 training situations, with accommodation on, and recorded the first three behaviors activated in response to each testing situation, resulting in 45 behaviors. We then coded how many behaviors were approach oriented and how many were avoidance oriented, as described previously. In addition, we summed the total activation across the approach and avoidance goal layers to look for differences in summed activation.

Results. The manipulations of punishment and reward values had the expected effect (see Table 8). As the relative reward/punishment values shifted from approach 1, avoidance .4 to approach .4, avoidance 1, there was an increase in the number of avoidance behaviors from 6/45 (13%) to 15/45 (33%). Further, the summed activations of the approach and avoidance nodes showed a corresponding shift from approach 1.64, avoidance .95 to approach .69, avoidance: 1.63. Thus, the manipulation of the reward and punishment values during learning did have a consistent effect on level of activation of the corresponding goal layers and on the corresponding type of behavior.

The results for avoidance are consistent with considerable work showing that those high in BIS or neuroticism learn responses to punishing cues faster than individuals low in BIS or anxiety (e.g., Zinbarg & Mohlman, 1998; for a review see Avila & Torrubia, 2008) and develop stronger negative expectations. Although the evidence for BAS impact is not as consistent, Pickering and Smillie (2008) and Smillie, Pickering, and Jackson (2006) have recently reviewed the evidence indicating that higher BAS sensitivity leads to stronger learning of rewarded cues.

Some of the clearest evidence of the impact of individual differences in BIS and BAS on reward and punishment learning is provided by Smillie, Dagleish, and Jackson (2007). They showed that higher BAS leads to greater response-sensitivity to rewarded stimuli, measured by a signal detection analysis, and higher BIS leads to greater sensitivity to punished stimuli.

Table 8
Results of Simulation 2: Individual Differences in Sensitivity to Reward and Punishment Value

	Reward and punishment values during training					
	Approach: 1.0 & Avoid: 0.4		Approach: 0.7 & Avoid: 0.7		Approach: 0.4 & Avoid: 1.0	
Variable	*n*	%	*n*	%	*n*	%
Number of avoidant behaviors	6/45	13	13/45	29	15/45	33
Summed activation of approach layer	1.640		.96		0.69	
Summed activation of avoidance layer	0.95		1.07		1.63	

Schönberg, Daw, Joel, & O'Doherty (2007) showed that there are strong individual differences in the extent to which individuals generate reinforcement learning signals when learning reward-based decision tasks and that the strength of this signal is positively related to reward learning. Individuals who exhibit strong prediction error signals in the striatum better learn to make the optimal rewarded choice in a decision making task. Although this work only addresses reward, it suggests that the approach and avoidance systems might influence behavior through their effect on the strength of reward and punishment cues and their impact on learning.

Discussion of Simulations 1, 2, and 3. In Simulations 1, 2, and 3, we examined how individual differences in approach and avoidance systems may influence behavior. In Simulations 1 and 2, we studied the impact of differences in sensitivity of the two systems to inputs that signal potentially rewarding versus punishing outcomes. In Simulation 2, we also tested the impact of differences in baseline activation. In Simulation 3, we examined whether the development of individual differences in behavior could be due to differences in the extent to which outcomes are seen as punishing or rewarding during learning.

However, in these simulations, all responses by the network receive feedback. But what if the feedback were contingent on the type of behavior, as is frequently the case? Eiser, Fazio, Stafford, and Prescott (2003) noted that in the real world, approach and avoidance behaviors typically provide asymmetric information: Approach behaviors, which involve interacting with the situation, typically result in feedback as to whether the situation is rewarding or punishing. In contrast, successful avoidance behaviors mean that the organism does not interact with the situation and fails to receive feedback.

To examine this potential asymmetry, Eiser et al. (2003) constructed a connectionist network that chose whether or not to "eat" a "bean" (approach or avoid the bean) that was potentially either good or bad and that then learned to categorize the beans as good or bad, based on feedback when it "ate" a bean (no feedback was received when it did not "eat"). Receiving feedback only when it "ate" a bean led to asymmetries in learning whether beans were good or bad. The network learned to evaluate bad beans perfectly, whereas they often categorized good beans as bad, thereby giving up the opportunity to attain a rewarding outcome. For the bad beans, if the individuals choose the bean, they get punished, and if they avoid the bean, they keep their initial negative expectancy. However, for good beans, if they choose them, they get rewarded, whereas if they avoid them because they think they are bad, they do not learn otherwise. Fazio, Eiser, and Shook (2004) showed this same learning asymmetry with people.

In a further simulation, Eiser et al. (2003) showed that increasing the likelihood of the network choosing a bean (approaching) increased accuracy in categorizing good beans. (Fazio, Eiser, & Shook, 2004, showed the same pattern with people.) However, the down side of making the network more approach oriented was that it was more likely to "die" from eating too many bad beans.

Although they did not examine individual differences in approach and avoidance orientation, this work suggests that such differences could influence learning asymmetries. Compared with approach oriented individuals, avoidance oriented individuals, because they avoid potentially negative situations, will be less likely to correct misconceptions and will miss some rewarding situations. However, Eiser et al.'s (2003) simulation, in which more approach oriented networks were more likely to "die" from encountering too many bad beans, suggests that in dangerous environments, being more approach oriented may not always be the best option.

Simulation 4: Interactions between approach and avoidance in the control of behavior. A further implication of the current model is that the influence of the approach system and the avoidance system on behavior should not necessarily be additive, as the systems compete with each other for the control of behavior (e.g., Pickering & Gray, 1999). Thus, one might often expect nonlinear relationships between the two systems. For example, suppose the approach system is initially more active, but then the activity of the avoidance system increases until it is more active. Because the approach and avoidance systems compete for the control of behavior, as the relative strength of the avoidance system increases there will be a sudden shift to avoidance behavior. This is relevant to understanding how extraversion and neuroticism could interact.

One example of such a nonlinear relationship between extraversion and neuroticism might be found at a party where, as long as the approach motivation is higher than the avoidance motivation, an individual will be fairly sociable, but then at some point as the avoidance motivation increases (social anxiety increases), there will be a sudden shift to avoidant behaviors, such as staying at the periphery of the action, being silent, or even leaving. The point of the current simulation was to demonstrate that our model could capture such nonlinearity.

We held approach activation constant but varied the avoidance activation from being less than approach to being considerably more. We did this by setting $\bar{g}_e$ (excitatory conductance) at 1 for the approach layer and then manipulating $\bar{g}_e$ for the avoidance layer, starting at 0.8 and increasing it in steps of 0.1 to 1.6. We used the default gain of 100 for both goal layers, with accommodation on.

Results. We used two situations out of the set of 15 ("party at work" and "party at restaurant") that activated both approach and avoid related goals. We recorded the behaviors and the average layer activation for each situation and for each level of $\bar{g}_e$ for the avoidance system. As can be seen in Table 9, for both situations, as $\bar{g}_e$ increased for the avoidance system, there were sudden increases in avoidance activation and a sudden shift from sociable behaviors (e.g., drink alcohol, gossip, dance) to avoidant, nonsocial behaviors (e.g., leave, be silent, stay at periphery). We consider the implications of this nonlinear relationship in the Discussion section.

Simulation 5: Disinhibition and constraint and behavioral switching or impulsivity. We tested whether our model could capture aspects of disinhibition and constraint (related to Conscientiousness) by varying the level of inhibition on both motives and behavior. First, we wanted to confirm that higher inhibition led to lower overall activation of the motive layers. Second, and more importantly, we tested whether higher inhibition would reduce the sensitivity of the network to changing inputs and reduce the likelihood that the network would switch to different behaviors. That is, with higher inhibition, would the network be more likely to stay focused on a current goal as the situational cues changed than to switch to new goals and behavior?

Table 9
Simulation 4: Results of Simulation of Interaction Between BIS and BAS in the Control of Behavior

BIS *g* _bar_e	Behaviors	Average BAS activation	Average BIS activation
	Party at work		
0.8	Tease and make fun, drink alcohol, gossip	0.121223	1.73E-14
0.9	Tease and make fun, drink alcohol, gossip	0.121223	8.36E-06
1.0	Tease and make fun, drink alcohol, order others	0.119589	0.0700957
1.1	Tease and make fun, introduce self, be silent	0.123967	0.108918
1.2	Gossip, tell jokes, give in	0.142863	0.164092
1.3	Drink alcohol, leave, introduce self	0.101414	0.389281
1.4	Leave, introduce self, improve skills	0.0881733	0.408919
1.5	Leave, introduce self, improve skills	0.0829425	0.42897
1.6	Leave, introduce self, improve skills	0.0914679	0.446203
	Party at restaurant		
0.8	Tease and make fun, dance, ask other to dance	0.161413	6.08E-16
0.9	Tease and make fun, insult others, ask other to dance	0.161414	1.20E-06
1.0	Tease and make fun, drink alcohol, explore environment	0.121996	0.103078
1.1	Insult others, explore environ, gossip	0.124009	0.110283
1.2	Tell jokes, steal, drink alcohol	0.135796	0.200727
1.3	Stay at periphery, compliment others, explore environ	0.155491	0.215646
1.4	Stay at periphery, be silent, introduce self	0.153316	0.220754
1.5	Stay at periphery, be silent, introduce self	0.151624	0.22454
1.6	Stay at periphery, be silent, introduce self	0.128478	0.335672

Note. BIS = behavioral inhibition systems; BAS = behavioral approach systems.

To examine the impact of inhibition on switching behavior, we did the following. First, for both motive layers we manipulated the k parameter for the kWTA average algorithm to be either $k = 1$ nodes active on average (high inhibition) or $k = 4$ (low inhibition). Second, we allowed the previous state of the network in response to earlier situations to influence subsequent states of the network when it was exposed to new situations. In everyday life, when one is exposed to a new situation, there is residual activation from the previous situation. For example, if one is sitting in one's office alone, working, and someone then walks by the open door, one's response will be partially influenced by one's previous state. To capture this, we set to .5 the Leabra++ parameter for the proportion of activation that carries over from a previous exposure to an input. Every time the network was exposed to a new situation, half of the activation of each node in response to the previous situation was carried over to the activation of the same node. When this parameter is set to 0, the default, no activation carries over.

What this carryover of activation means for the impact of inhibition is the following: when $k = 1$ then on average only the activation from one motive in a layer will be carried over to the new set of activations, whereas when $k = 4$ then on average the activation from four motives will be carried over. (kWTA sets the activation of losers to close to 0.) As a result, as the network is exposed to the sequence of situations, the network in the high inhibition condition ($k = 1$) receives influence from only one motive in each layer, on average, whereas the network in the low inhibition condition ($k = 4$) received carryover activation from four motives in each layer, on average. Thus, the network in the high inhibition condition is responding to a more restricted set of motives. As a result, the network in the high inhibition condition should show less switching.

We then exposed the network to the sequence of 15 testing situations and recorded what behavior was activated. We then determined whether the activated behavior changed when a new testing situation was presented or whether the activated behavior stayed the same. If we count the first behavior activated in response to the first situation as one then there are 15 possible behavior changes across the 15 situations. We also summed the activations over each layer and then over each of the 15 events. This was done for five different initializations of the network.

Results. Manipulating the level of inhibition had a consistent impact on both the overall activation of the motive layers and the tendency of the network to switch among behaviors (see Table 10). Overall activation in both layers was considerably lower in the high ($k = 1$) inhibition condition compared with the low inhibition condition ($k = 4$). Further, there was less switching among behaviors in the high inhibition condition. The only exception was in Initialization 2, for which in the low inhibition condition, the network seems to have gotten "stuck" in the same behavior for all the situations.

Thus, inhibition (as operationalized by the k parameter) does affect the extent to which the network remains focused on a goal and behavior (constrained) versus switching to a new goal and behavior (disinhibited). This suggests that the model can capture at least some aspects of impulsivity: When inhibition is low, the network is more prone to switch to a new goal. Consistent with our argument that this inhibitory process captures part of conscientiousness, Nigg et al. (2002; also see Clark and Watson, 2008) have shown that low conscientiousness predicts attention-deficit/hyperactivity disorder, specifically attentional and organizational problems.

Simulation 6: Role of individual motives in defining specific personality traits. In the preceding, we tested general characteristics of the model, such as the behavior of the approach and avoidance layers, how they could capture aspects of the two broad factors of extraversion and neuroticism, and whether aspects of the general factor of disinhibition and constraint could be captured by

Table 10
Simulation 5: Disinhibition and Constraint: Results of Impact of Variations in Inhibition

	Low inhibition			High inhibition		
Initialization	Behavior change	Approach activation	Avoidance activation	Behavior change	Approach activation	Avoidance activation
1	10	27.6	29.0	7	4.9	5.4
2	2	32.8	32.4	8	4.2	5.3
3	8	33.5	33.5	6	10	1.7
4	8	30.7	27.9	5	4.2	3.7
5	11	25.4	26.4	6	11.1	0.4

For low inhibition, $k = 4$. For high inhibition, $k = 1$.

inhibitory processes. Here, we focus on capturing more specific traits by manipulating specific goals and resources, consistent with L. C. Miller and Read's (L. C. Miller & Read, 1991; Read & Miller, 1989) goal-based model of personality.

We created six distinct personalities by manipulating the biases on goals and resources that theory suggested should be central to a corresponding trait. The six traits form three pairs (sociable–shy, confident–anxious, industrious–lazy), corresponding to one of three major dimensions of the Big Five: Extraversion (sociable–shy), Neuroticism (confident–anxious), and Conscientiousness (industrious–lazy). These dimensions also correspond to the three major dimensions in Eysenck's (1983, 1994) Extraversion, Psychoticism, and Neuroticism (EPN) and in Tellegen's (1985) model. Manipulating the biases on goals and resources changes the baseline activation of the corresponding nodes.

The biases for the six traits are in Table 11. For example, consider the representation of the trait sociable. The biases for the approach goals of friendship, being liked, sex and romance, exploring, and fun are set well above 0, and the goal of dominance is set below 0. Further, the biases for the avoidance goals of avoid rejection and/or embarrassment and avoid interpersonal conflict are set above 0. Finally, the biases for the resources of social skills, things to talk about, wit, quick thinking, face-saving skills, and intelligence are set above 0. The specific biases for the different traits were the result of discussion among Stephen J. Read, Brian M. Monroe, Aaron L. Brownstein, Yu Yang, and Gurveen Chopra as to which goals and resources characterized an individual with each of the six traits.

Our goal is to provide an initial demonstration that differences in the chronic activation of relevant goals and resources can plausibly capture individual traits. For this purpose it is most important to arrive at a plausible representation of traits, rather than a precisely accurate one. We aim to show that if one represents traits in this way, one gets behavioral responses to situations that are consistent with the relevant trait. Assuming that we can demonstrate this, further work should allow for improvement in the precision of the representations.

To test the result of defining traits in this way, we ran six simulations, each with a different patterns of biases that corresponded to one of the six traits. For each simulation, we tested with the 15 training situations and recorded which behaviors were activated across the 15 situations.

Results. The behaviors generated in response to each of the fifteen situations for each of the six traits can be seen in Table 12. The extent to which each of the generated behaviors exemplifies the relevant trait dimension is plotted in Figures 2A, B, and C. There is one figure for each trait pair and its corresponding dimension: Extraversion, Conscientiousness, and Neuroticism. Situations are ordered along the x-axis and sorted into the two groups of work situations (1–9) and social situations (10–15). Situations 8 and 9 are somewhat mixed, as 8 is a party at work and 9 is a social engagement at the boss' house. Within the work situations, they are ordered both in terms of how work oriented they are (1 = break by yourself, 2 = break with others, versus active working) and how many others are present (3 = individual assignment, 4 = working with one other, 5 = working together, urgent project, 6 = group meeting, and 7 = review with boss). Within the social situations, they are ordered in terms of whether they are romantic or not (10 = trying to get a date, 11 = on a date, 12 = at a dance, 13 = party at a restaurant, 14 = wedding reception, 15 = family birthday). The y-axis represents how strongly the behavior exemplifies the relevant trait. For the sociable–shy pair, the y-axis represents how Extraverted the behaviors are on a scale from 1 (*highly introverted*) to 10 (*highly extraverted*). For the confident–anxious pair, the axis represents how anxious–avoidant (Neuroticism) the behaviors are on a scale ranging from 1 (*not at all anxious–avoidant*) to 10 (*highly anxious–avoidant*) and for the industrious–lazy pair, the y-axis ranges from 1 (*not at all conscientious*) to 10 (*highly conscientious*). Ratings were obtained by having four of the authors rate all 43 behaviors on each of the three dimensions and then averaging the rating for each behavior.

As expected, networks with different personalities exhibited quite different patterns of behavior in response to the 15 situations. For example, sociable personalities (see Figure 2A) tended to behave in a more extraverted fashion in work situations when others were present and in social situations, particularly romantic situations, than did shy personalities. It is interesting to note that the sociable and shy personalities did not differ in the individual assignment situation, when there is no one else present, and in the wedding and family birthday situations, when there are familiar others present. This pattern is consistent with the idea that behavior is a function of both personality and the situation and that there is often an interaction between personality and situation, such that individual differences in some situations are nonexistent in others.

Confident personalities (see Figure 2C) exhibited much less anxious behavior in many of the social situations than did anxious personalities. However, there were no real differences in the anxiousness of behaviors in the work related setting, although in general, in work situations, the confident personalities put extra

Table 11
Simulation 6: Settings of Goals and Resources for Simulation of Individual Personality Traits

Goal	Sociable	Shy	Confident	Anxious	Industrious	Lazy
Approach goals						
Friendship	1.5	0	0	0	0	0
Sex and romance	0.6	0	0	0	0	0
Be liked	1.5	0	0.6	0	0	0
Help others	0	0	0	0	0	−0.9
Dominance	−0.6	0	0.6	0	0	0
Achievement	0	0	0.9	0	2.1	−0.9
Exploring	0.6	0	0.9	−0.9	0	−0.3
Fun	0.9	0	0	0	0	0.9
Mastery	0	0	0.9	0	2.1	−0.9
Fairness, equality, justice	0	0	0	0	0	0
Uniqueness	0	0	0	0	0.9	0
Material gain	0	0	0	0	0.9	0
Avoid goals						
Rejection or embarrassment	0.9	1.8	−0.6	1.2	0	0
Guilt	0	0	−0.6	1.2	0	−0.6
Failure	0	0	−0.9	1.2	1.8	−0.6
Harm physical	0	0	−0.6	1.2	0	0
Loss of control	0	0	−0.6	1.2	0	0
Interpersonal conflict	0.6	1.8	−0.6	0.9	0	0
Effort	0	0	−0.6	0	−0.9	2.1
Risk uncertainty	0	0	−0.6	1.2	0	0
Resources						
Social skills	2.1	0	1.8	0	0	0
Things to talk about	2.7	0	1.2	0	0	0
Money	0	0	1.2	0	0	0
Attention span	0	0	1.8	0	1.2	0
Job skills	0	0	2.4	0	1.8	0
Face saving skills	0.9	0	0.9	0	0	0
Time	0	0	0	0	1.8	0.9
Intelligence	0	0	2.4	0	1.8	0
Quick thinking	1.2	0	1.8	0	0.9	0
Wit	0.6	0	1.8	0	0	0
Status	0	0	2.4	0	0	0
Happy	0	0	0	0	0	0
Mellow	0	0	0	0	0	1.8
Depressed	0	0	0	0	0	0

effort into their job, whereas the anxious personalities simply did their job. This pattern of interaction between type of situation (work versus social) and personality trait again indicates that this model might be able to capture important aspects of person–situation interactions in human personality.

The lazy and industrious personalities also exhibited somewhat different patterns of behavior (see Figure 2B), with the industrious individual consistently either trying to find a new way to do their job or putting extra effort into their job, with the single exception of asking someone to dance at a dance. In contrast, the lazy person exhibited much less effort, doing such things as procrastinating, kissing up, drinking alcohol, or doing their job in a situation where the industrious individual was exhibiting extra effort.

Despite these consistent differences between lazy and industrious individuals, it is interesting that both the industrious and lazy individual did ask someone to dance when they were at a dance. This is an example in which the situational press is sufficient to over ride strong individual differences. These results again provide an example of person–situation interactions.

Simulation 7: Interindividual versus intraindividual variability in personality. Fleeson (2001, 2007) has examined the relative magnitude of interindividual and intraindividual variability in personality and found that for major factors of the Big Five, the variability within an individual in personality related states is typically of the same magnitude as the variability in personality traits between different individuals. Further, he has shown that this within person variability can be explained by situational contingencies. For example, variability in extraverted states is contingent upon characteristics of specific situations, such as the number of people present or their relationship to the subject (Fleeson, 2001, 2007).

This contingency between situational features and trait related behavior is conceptually similar to the results in Simulation 6. In that simulation, traits are defined in terms of goals and resources and behavior is partially a function of the extent to which situational features activate different goals. In Simulation 6, we found that there were typically high levels of variability from situation to situation in expression of trait related behavior.

In Simulation 6, we simulated midlevel traits (i.e., sociable–shy, confident–anxious, industrious–lazy) taken from three of the Big Five. Here, we simulate a broader trait dimension, specifically the communal component of Extraversion, by manipulating both the conductances of the approach and avoidance layers and the bias weights of the communal goals in the approach layer: friendship, sex–romance, be liked, help others, and fairness–equality–justice.

Table 12
Simulation 6: Behaviors Generated for Each Trait for Each of 15 Situations

Situation	Confident	Anxious	Lazy	Industrious	Shy	Sociable
Break by yourself	Ask for date	Surf web	Procrastinate	Find new way	Procrastinate	Tease and make fun
Break with coworkers	Talk politics	Ask other about self	Procrastinate	Find new way	Tease and make fun	Ask others about self
Individual assignment	Extra effort	Do job	Do job	Extra effort	Extra effort	Extra effort
Working with 1 other	Extra effort	Do job	Do job	Extra effort	Ensure fairness	Ask other to dance
Working together on urgent project	Extra effort	Do job	Extra effort	Extra effort	Ensure fairness	Gossip
Group meeting	Extra effort	Do job	Ensure fairness	Extra effort	Extra effort	Gossip
Review with boss	Extra effort	Do job	Kiss up	Extra effort	Ensure fairness	Kiss up
Party at work	Extra effort	Leave	Tease and make fun	Extra effort	Talk politics	Tease and make fun
Social engagement boss house	Extra effort	Leave	Extra effort	Find new way	Talk politics	Ask others about self
Trying to get a date	Self disclose	Be silent	Drink alcohol	Find new way	Pretend to work	Kiss
On a date	Self disclose	Stay at periphery	Drink alcohol	Find new way	Give in	Kiss
Dance	Ask other to dance	Introduce self	Ask other to dance	Ask other to dance	Introduce self	Ask other to dance
Party, restaurant or bar	Extra effort	Drink alcohol	Tease and make fun	Extra effort	Stay at periphery	Kiss
Wedding, fancy restaurant	Drink alcohol	Leave	Introduce self	Find new way	Introduce self	Ask others about self
Family birthday	Drink alcohol	Clean up	Confront other	Find new way	Tell jokes	Gossip

Our goal with this simulation is to demonstrate the same kind of intraindividual variability and the same kind of situational contingency that Fleeson (2001, 2007) has demonstrated. For our introverted individual, we left all goal bias weights at baseline, and we set the approach excitatory conductance at 0.9 and the avoidance excitatory conductances at 1.1. For our extraverted individual, we increased the bias weights to the communal goals to 2.1, and we set both the approach and avoidance excitatory conductances to 1.0. As in Simulation 6, we presented the 15 testing situations and recorded which behavior is most strongly activated for each situation.

Results. The behaviors generated in response to each situation can be seen in Table 13, and the extent to which each behavior exemplifies Extraversion can be seen in Figure 3. The ratings of the behaviors on Extraversion are the same as the ratings used in the sociable–shy simulation in Simulation 6. As can be seen, there are both substantial differences between the extraverted and introverted individual, as well as substantial variability across situations for the extravert.

Fleeson (2007) has argued that the role of situational contingencies in within person variability can be understood in terms of the if–then, person–situation contingencies in Mischel and Shoda's (1995, 1998) CAPS model. However, he did not provide a computational example. The current simulation demonstrates that our model can provide a computational implementation of Fleeson's point.

Simulation 8: Rejection sensitivity. Earlier simulations focused on relatively broad personality characteristics, ranging from broad Big Five factors, to more midlevel subcomponents of those factors, such as sociability and industriousness. Here we focus on a more specific personality characteristic, Downey's concept of rejection sensitivity (Downey & Feldman, 1996; Romero-Canyas & Downey, 2005). Rejection sensitivity is the disposition to readily perceive and strongly react to possible rejection cues in the behavior of other people. Much of Downey's (e.g., Downey & Feldman, 1996) work has focused on the impact of rejection sensitivity on attentional and affective responses to possible rejection cues, although some work has shown that rejection sensitive individuals are likely to respond behaviorally to cues to rejection with either hostility or social withdrawal.

Our current model has several avoid goals that are relevant to rejection sensitivity, avoid social rejection and avoid interpersonal conflict, and a number of behaviors that are relevant to social withdrawal. However, we do not have any hostility related behaviors. Thus, in the current simulation, we focused on manipulating bias weights on these two avoid goals and examining their impact on withdrawal related behaviors.

For the rejection sensitive individual, we set the bias weights for avoid social rejection and for avoid interpersonal conflict to an elevated level of 2.1, whereas for the normal individual these were left at baseline. For the rejection sensitive individual, the excitatory conductance for the approach layer was set at 0.9, and for the avoidance layer, it was set at 1.2. For the normal individual, the excitatory conductances were 0.9 and 1.1 respectively. Each network then received the 15 standard situations, and the first behavior activated was recorded.

Results. The behaviors generated in response to the 15 different situations for both the rejection sensitive and the normal individual can be seen in Table 14. As predicted, the rejection sensitive individual responds to social situations by consistently being socially withdrawn: they leave, they stay quiet, or they stay at the periphery. In contrast, the normal individual performs a much wider range of normal social behaviors. Thus, this simulation indicates that the current model can capture more specific and narrow range dispositions.

Admittedly, the situational features and behaviors in the current model are only a rough approximation of what one would do if one were designing a model that was specifically of rejection sensitivity. Instead of having cues such as the availability and number of individuals or the nature of the social situation, ideally, one would want a much more detailed representation that would allow one to represent specific behavioral cues to possible rejection. Further, one would like to be able to represent other behavioral responses, such as hostility, as well as cognitive and affective responses. Nevertheless, the current simulation demonstrates that one can

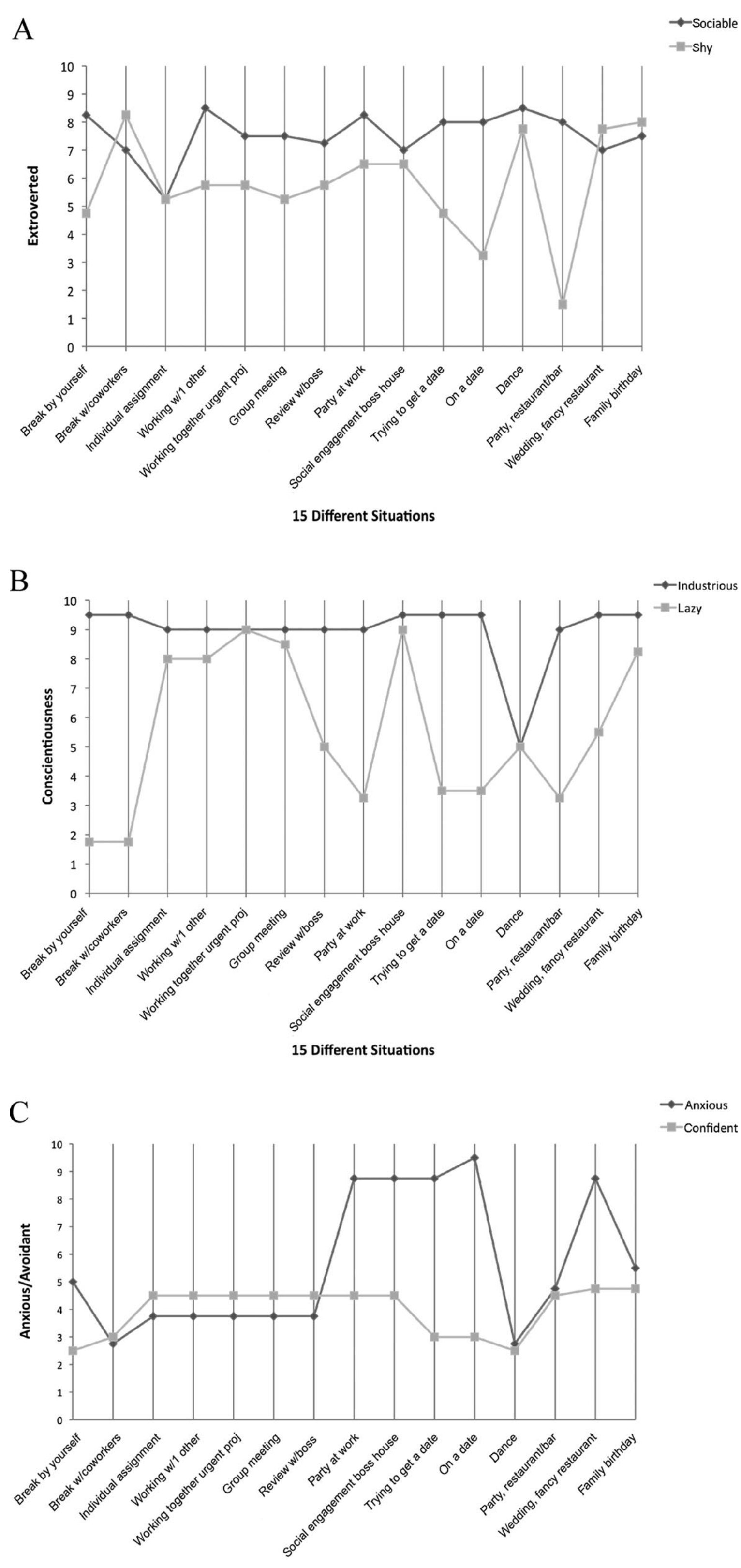

Figure 2. Simulation 6: Ratings for each of the generated behaviors for each of the three trait pairs: (A) extraversion (sociable–shy), (B) conscientiousness (industrious–lazy), and (C) anxious/avoidant (neuroticism; confident–anxious). w/ = with; proj = project.

Table 13
Simulation 7: Behaviors for Extrovert and Introvert

Input	Extrovert behavior	Introvert behavior
Break by yourself	Clean-up	Introduce self
Break with coworkers	Introduce self	Ask other about self
Dance	Compliment other	Ask other to dance
Family birthday	Eat and drink	Ask other about self
Group meeting	Ensure fairness	Ignore other
Individual assignment	Clean-up	Leave
On a date	Compliment other	Eat and drink
Party at work	Introduce self	Wear distinctive clothes
Party, restaurant or bar	Compliment other	Leave
Review with boss	Mediate	Kiss up
Social engagement boss house	Introduce self	Introduce self
Trying to get a date	Compliment other	Stay at periphery
Wedding, fancy restaurant	Gossip	Eat and drink
Working together on urgent project	Stay with and comfort other	Stay with and comfort other
Working with 1 other	Clean-up	Ensure fairness

Note. For extrovert, changing biases to 2.1 for all communal goals; *g* _e (approach) = 1.0; and *g* _e (avoid) = 1.0. For introvert, all baseline biases, *g*_e (approach) = 0.9 and *g*_e (avoid) = 1.1.

capture the basic features of rejection sensitivity by manipulating the baseline activation of relevant motives.

Discussion

Summary of Simulations

We first demonstrated, in the validation simulations, that the model successfully learned relationships between different situations and the behaviors that are appropriate in those situations. The model is not learning to associate just a single behavior with a single situation, but is learning many-to-many relationships: Any situation is associated with several different behaviors, and conversely, any behavior is associated with several different situations.

Our second set of simulations showed the model's ability to capture different aspects of personality (See Table 15 for the key

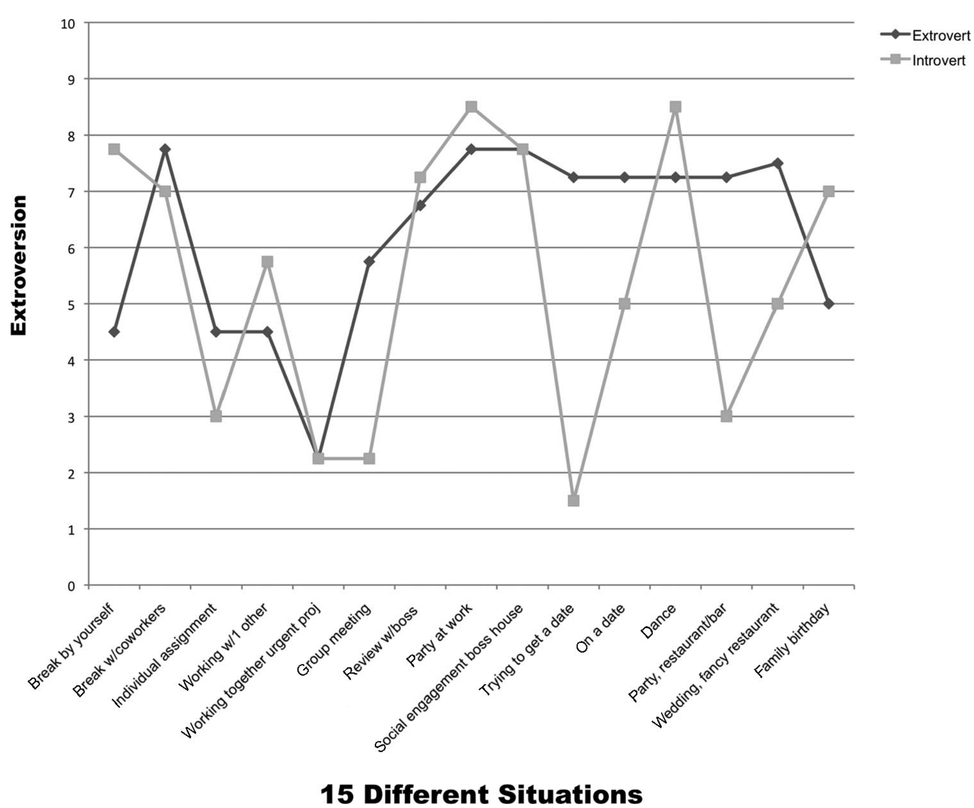

Figure 3. Simulation 7: Extraversion ratings for generated behaviors for an extrovert versus an introvert. w/ = with; proj = project.

Table 14
Simulation 8: Behaviors for Rejection Sensitive and Normal Individual

Situational input	Rejection sensitive individual behaviors	Normal individual behaviors
Break by yourself	Introduce self	Introduce self
Break with coworkers	Be silent	Ask others about self
Dance	Be silent	Ask other to dance
Family birthday	Stay at periphery	Ask others about self
Group meeting	Leave	Ignore others
Individual assignment	Be silent	Leave
On a date	Stay at periphery	Eat and drink
Party at work	Be silent	Wear distinctive clothes
Party, restaurant/bar	Tell jokes	Leave
Review with boss	Be silent	Kiss up
Social engagement boss house	Tell jokes	Introduce self
Trying to get a date	Be silent	Stay at periphery
Wedding, fancy restaurant	Leave	Eat and drink
Working together on urgent project	Be silent	Stay with comfort other
Working with 1 other	Leave	Ensure fairness

Note. For the rejection sensitive individual, biases for avoid rejection = 2.1, and biases for avoid interpersonal conflict = 2.1; g _e (approach) = 0.9; g _e (avoid) = 1.2. For the normal individual, bias for avoid rejection = baseline, and bias for avoid interpersonal conflict = baseline; g _e (approach) = 0.9; g _e (Avoid) = 1.1.

parameters in these simulations). We first focused on the behavior of the approach and avoidance goal layers, as they are central to our model and to its ability to represent two major dimensions of personality: extraversion and neuroticism. The relative sensitivity of the two layers should play an important role in the model's ability to simulate individual differences. In Simulation 1, we manipulated the relative sensitivities of the two layers by varying the relative excitatory conductances for the approach and avoidance layers, from 1.2 and 0.8, respectively, to 0.8 and 1.2, respectively. As expected, as the relative sensitivity of the avoidance layer increased, the average activation of the avoidance goals and the frequency of activation of avoidance behaviors increased, and conversely, as the relative sensitivity of the approach layer increased, the average activation of the approach goals and the frequency of approach behaviors increased. These results are consistent with the idea that individuals with a more sensitive approach system should exhibit a broad pattern of approach related or extraverted behavior, whereas individuals with a more sensitive avoidance system should exhibit a broad pattern of avoidance related or neurotic behaviors.

In Simulation 2, we examined potential differences in the activation dynamics of the approach and avoidance systems. Cacioppo, Gardner, and Berntson (1997) have argued that positive and negative evaluation systems (corresponding to approach and avoidance systems) have somewhat different activation dynamics. We gave the approach system a positivity offset; giving it a higher baseline activation by increasing its baseline V_m, and we gave the avoidance system a negativity bias by increasing its excitatory conductance. As predicted, with weak input to the two layers, the approach layer strongly influenced behavior, but as the strength of the input increased, the avoidance layer had an increasing influence on behavior.

In Simulation 3, we examined the impact of individual differences in sensitivity to reward and punishment cues during learning on approach and avoidance behaviors. This can be thought of as examining the impact of initial temperament differences on learned sensitivity to reward and punishment. As expected, higher reward values for approach goals led to relatively more frequent approach related behaviors, whereas higher punishment values for avoidance goals led to more frequent avoidance related behavior. Thus, the model successfully captures the potential impact of initial differences in sensitivity to reward and punishment on learning and the frequency of approach and avoidance related behaviors.

In Simulation 4, we tested whether there were nonlinear relations between the approach and avoidance systems in generating behavior. By holding the baseline sensitivity of the approach system constant and varying the sensitivity of the avoidance system from lower than the approach system to higher, we showed that as the sensitivity of the avoidance system increased, the model suddenly switched from producing approach related behaviors (e.g., dance, tease/make fun) to avoidance related behaviors (e.g., stay at periphery, leave). Thus, this model clearly suggests that the relationship between the approach and avoidance systems is interactive. Further, this simulation suggests that when the sensitivity of the avoidance system is high enough, then it will largely prevent the system from engaging in approach related behaviors. One example of this would be in a social situation, where sufficiently high levels of social anxiety would largely inhibit social approach oriented behaviors.

In the first 4 simulations, we examined the ability of the model to capture broad individual differences between extraversion (or approach) and neuroticism–emotional stability (avoidance). In Simulation 5, we showed that varying the level of inhibition in the network resulted in behavior that was consistent with the concept of disinhibition and constraint and with some aspects of impulsivity. Specifically, higher inhibition on the two motive layers led to less sensitivity to changes in situational cues and a greater tendency to remain focused on previously activated motives when generating behavior.

In Simulation 6, we showed that we could model individual differences in midlevel traits, by manipulating the underlying

Table 15
Parameter Settings for the Different Simulations

Simulation	Gain motive layers	Learning rate proportion	Approach excitatory conductance	Avoidance excitatory conductance	Approach baseline activation ($V_m$init)	Avoidance baseline activation ($V_m$init)	Threshold È	Prop. of activation passthrough	kWTA (*k*) for motives	Weight Scaling to motive layer
1	100	Hebbian: .01; CHL: .99	1.2 1.1 1.0 0.9 0.8	0.8 0.9 1.0 1.1 1.2	.20	.20	.3	0	2	1
2	100	Hebbian: .01; CHL: .99	0.9	1.2	.27	.20	.3	0	2	0.5 1.0 1.5
3	20	Hebbian: .01; CHL: .99	1.0	1.0	.20	.20	.3	0	2	1
4	100	Hebbian: .01 CHL: .99	1.0	0.8 to 1.6 in 0.1 increments	.20	.20	.3	0	2	1
5	100	Hebbian: .01; CHL: .99	1.0	1.0	.20	.20	.3	.5	1 versus 4	1
6	100	Hebbian: .01; CHL: .99	1.0	1.0	.20	.20	.3	0	2	1
7	20	Hebbian: .01; CHL: .99	Extrovert: 1.0; Introvert: 0.9	1.0	.20	.20	.3	0	2	1
8	20	Hebbian: .01; CHL: .99	Rejection sensitive: 0.9; Normal: 0.9	1.1 1.2	.20	.20	.3	0	2	1

Note. Learning rate: Features to hidden layer (goals); lrate = .005; hidden layer to goals; lrate = .005; goals to hidden layer (behavior); lrate = .005; features to hidden layer (behavior); lrate = .0025; resources to hidden layer (behavior), lrate = .01. kWTA = k-winners-take-all; CHL = contrastive Hebbian learning; $V_m$init = baseline activation parameter.

patterns of goals and resources. We first identified three pairs of traits (sociable–shy, confident–anxious, and industrious–lazy) and manipulated the bias weights for the goals and resources relevant to a specific trait. Increasing the bias for a goal or resource increases its baseline activation and makes it more likely to fire. Changing the baseline activations of the relevant goals and resources had the expected effect on behavior. For example, increasing biases on sociable related goals led to increased levels of behavior involving people, whereas increasing the biases on shy related goals led to increased levels of solitary behaviors.

This simulation also showed consistent evidence of person–situation interactions. For example, although the industrious and the lazy individual differed considerably in most situations, when they were at a dance, they both asked someone to dance. Or for the sociable–shy pairings, the two individuals look quite similar in the individual assignment work setting, where there is no one else around, and in social settings with familiar others. However, they differed considerably in work settings with other people and in social settings with unfamiliar others. Finally, the confident–anxious individuals are similar in work settings but differ in social settings. Thus, this model can capture various aspects of person–situation interactions.

Simulation 7 provided further evidence that our model can capture person–situation interactions, in simulating aspects of Fleeson's (2001, 2007) work demonstrating that intraindividual variability in personality states is frequently as great as interindividual variability in personality traits. We simulated the communal aspects of extraversion and showed considerable variability both within and between individuals in their response to a range of situations.

Finally, Simulation 8 showed that we could simulate a fairly specific personality trait by manipulating a small set of motives. We simulated aspects of Downey's (Downey & Feldman, 1996) concept of rejection sensitivity by increasing the bias weights on two motives, avoid social rejection and avoid interpersonal conflict and showed that this led to a substantial increase in social withdrawal.

Relations to Other Computational Models of Personality

Read and Miller (2002) and Read et al. (2006). The current model differs in important ways from our previous model (Read & Miller, 2002). The broadest difference is that our original model was quite abstract, with very abstract representations of personality and situations, whereas this model focuses on simulating much more realistic and specific aspects of situations, personality, and behavior.

A number of changes enabled the creation of much richer and more realistic simulations. First, the current model includes learning, enabling us to capture realistic ecological frequencies. But more important, learning allowed us to examine how initial individual differences in sensitivity to reward and punishment (e.g., early temperament) could lead to personality differences due to differential learning. For example, individuals who find certain situations more punishing should learn to respond more quickly and strongly with avoidant behaviors.

Second, we introduced more realistic and richer situational representations. In the current model, each node in the situation layer corresponds to a different feature, and each situation is represented by a configuration of situational features. For example, a party situation could be defined by features for music, alcohol, other people, a large room, and low lighting. A working with others situation could be defined by features for a conference room, fluorescent lights, having work to do, and being with other people. Thus, there was no longer a one to one situational feature–goal correspondence, as in the earlier model. A goal's activation is the result of input from a configuration of situational features. Thus, the model should learn to associate configurations of situational features to different goals.

Third, our potential trait representations are much richer. Although in both models, gains on the two motive layers and inhibition are important in representing broad trait dimensions, in the current model, patterns of activated goals and resources are a key part of the definition of specific traits. This allows us to capture more specific traits than in the previous model.

We also used more concrete behaviors. Behavior nodes correspond to such things as tell jokes, dance, drink alcohol, and work hard. Previously, the behavior nodes corresponded to highly abstract behaviors, such as helpful or fearful behavior. Also, resources were present previously, but were not used. Here, they are part of the joint input into the hidden layer that activates behaviors, and they are part of trait definitions.

A final important difference is in how behaviors are activated. In the previous version, each behavior received input from a single corresponding goal. However, in the current model, a behavior is activated by a configuration of situational features, resources, and goals, all mediated through two hidden layers. Thus, the triggering of behavior is the result of a much more complex set of interactions among a number of different components. One reason for this more complex model is that we were modeling more aspects of Read and Miller's (1989; L. C. Miller & Read, 1987) model of traits, in which they argued that traits consist of configurations of goals, plans, resources, and beliefs (also see Mischel and Shoda's, 1995, concept of Cognitive Affective Units [CAUs]).

In a recent article (Read et al. 2006), we described how some aspects of the present theoretical model provided the foundation for a program (Personality enabled architecture for cognition [PAC]) to create intelligent agents that display realistic variations in personality. PAC was explicitly designed as an architecture for creating intelligent agents and not as a theoretical model of human personality. In that article, we did not systematically test the assumptions of our theoretical model but rather focused on whether it could produce plausible variations in the behavior of intelligent agents, with a symbolic, cognitive architecture. In contrast, in the present article, we systematically evaluate major assumptions of the theoretical model.

There are other major differences between the current model and PAC. The current model is implemented in a neural network architecture that was designed to be neurobiologically plausible and to capture major aspects of the neurobiology of cognition and motivation. Moreover, the neural network model has a mechanism for inhibition that allows for the dynamic adjustment of degree of inhibition as a function of the level of activation of a layer. In contrast, PAC only allows for a static, very partial manipulation of the degree of disinhibition and constraint in terms of a constant level of inhibition and a fixed number of motives that can be active.

Further, the current model has learning, which PAC does not. Thus, the current model can (a) capture individual differences in reward and punishment sensitivity in learning to respond to situations, (b) capture how chronic differences in baseline motive activations might develop, and (c) model the development of configural representations, such as the representation of different kinds of situations. Finally, the network model allows us to examine the role of constraint satisfaction processing in personality related processes.

Mischel and Shoda's CAPS model. Mischel and Shoda (1995, 2008; Shoda & Mischel, 1998) have presented a constraint satisfaction model of human personality, based on Mischel's (1973) cognitive social learning theory of personality. In their model, a set of input units, representing situational features, connect to a recurrently connected set of mediating CAUs, which are then connected to a behavior node. But, there has been no real attempt to specify the structure of underlying motivational or neural systems; different personalities are represented by different randomly connected sets of mediating CAUs. Moreover, they have not explored whether or how these motivational and cognitive units might be used to capture the interindividual structure of human personality as represented in structural models of personality. In contrast, a central goal of the current model is to begin to specify the intraindividual structure of personality in terms of structured motivational systems and to demonstrate how the interindividual structure of human personality can result from the interactions among underlying motivational and cognitive structures and processes. Further, although Mischel and Shoda (1995; Shoda & Mischel, 1998) acknowledged that work on genetics and the neurobiological bases of behavior are ultimately important in understanding the structure and dynamics of personality, Mischel and Shoda (1995; Shoda & Mischel, 1998) have not explored the relationship of genetics and the neurobiological bases of behavior to their model.

Other models of personality dynamics. Cervone (2004) and Kuhl (2000) recently proposed other accounts of the intrapersonal structure of personality mechanisms and dynamics. In the knowledge and appraisal personality architecture (KAPPA) model, Cervone (2004, 2005) argued that individual differences and cross-situational consistency in behavior can be at least partially understood in terms of idiographic differences in knowledge structures and people's appraisals of their situations. Kuhl (2000), in his personality systems interaction theory, argued that motivated behavior and personality differences can be understood in terms of the flow of energy among four basic, general systems: object recognition, intuitive behavior control, extension memory/feeling (parallel, holistic, and fast), and intention memory/thinking (serial, analytic and slow). Both models focus on the within person or intraindividual structure of the mechanisms that underlie personality but focus little attention on the interindividual structure of personality.

Pickering's computational model of Gray's reinforcement sensitivity theory. Pickering (2004, 2008) has recently presented a computational model of the revised reinforcement sensitivity theory (RST; Gray & McNaughton, 2000). In the revised RST, the BAS remains largely the same, except that it now governs sensitivity to both conditioned and unconditioned cues to reward (appetitive stimuli; which is how it had been typically treated), whereas previously, it only governed cues to conditioned reward. However, the conception of the BIS has changed considerably. Originally, the BIS governed sensitivity to conditioned cues to punishment. The fight–flight–freeze system (FFFS) now takes that role, governing sensitivity to all cues to aversive situations (both conditioned and unconditioned) and the primary emotion associated with it is fear. Because the new FFFS plays the same functional role as the old BIS, results from questionnaire measures of the BIS (punishment sensitivity) are still relevant.

In the revised version of RST, the BIS is sensitive to approach–avoidance conflict between the BAS and the FFFS, and the primary associated emotion is anxiety. A clear distinction is made between anxiety and fear, with Gray and McNaughton (2000) having argued for distinct neurobiological substrates for the two systems. However, this claim is the subject of active argument.

Although the revised RST changes the roles of the BIS and the FFFS, the structure of the revised system is quite similar to the old one. One motivational system still responds to cues of reward (BAS), and a second responds to aversive cues (FFFS).

The major addition is that the BIS is now an approach–avoidance goal conflict monitoring system. The BIS receives excitatory relationships from both the BAS and the FFFS and is highly activated only when there is high goal conflict, represented by high activation of both the BAS and the FFFS. The BIS has an inhibitory relation with the BAS and an excitatory relation with the FFFS. So, high goal conflict leads to strengthening avoidance motives and weakening approach motives. Further, the BAS and the FFFS have inhibitory relationships to each other, further helping to increase the difference between approach and avoidance motives.

Pickering (2004, 2008) presented a computational model of the new RST. One of his primary questions was the implication of the RST system structure for the extent to which major dimensions of personality should be orthogonal or correlated. He proposed that the relations among the various systems imply that the major dimensions of personality that correspond to the systems in RST should be correlated, rather than orthogonal. In a key simulation, he generated 100 individuals with different combinations of BIS, BAS, and FFFS sensitivities drawn randomly from independent normal distributions and then exposed each individual to 200 different situations composed of different strengths of reward and punishment cues. As predicted, there were strong negative correlations between the BAS activation and the FFFS activation, consistent with empirical findings of negative relationships between neuroticism and extraversion, as well as positive correlations between the BIS and the FFFS activations, paralleling empirical findings indicating an overlap between fear and anxiety. There was also a somewhat unexpected positive correlation between the BAS activation and the BIS activation, although in hindsight, this is not surprising because high BIS activation, which results from approach–avoidance conflict, depends on the joint activation of the BAS and the FFFS.

Our model is strongly influenced by Gray's (1987a, 1987b, 1988, 1991) original conception of the BIS and BAS and is also broadly consistent with the revised RST. However, there are important differences. First, in the revised RST (Smillie, Pickering, & Jackson, 2006), the approach (BAS) and avoidance (FFFS) systems have direct inhibitory relationships with each other, whereas in our model, the two systems do not. Instead, there is strong competition among potential behaviors, so that the motive systems compete indirectly through their attempt to control behavior. This indirect competition could influence the activation of the two motive systems, through the feedback relationships from the behavior layer to the goal layers.

One potential advantage of our approach is that it allows behavior to be jointly driven by both approach and avoidance motivations, if a behavior can simultaneously satisfy both types of motive. For example, an individual's choice of a behavior in a social situation, such as telling jokes, might be jointly motivated by a motive to be with others and a motive to avoid social rejection. In contrast, in Pickering's (2004, 2008) model, high Approach (BAS) and avoidance (FFFS) motives necessarily compete with one another.

Second, the revised RST has a goal conflict monitoring system, the BIS, which functions to reduce the conflict between the approach and the avoidance system by inhibiting the BAS and facilitating the FFFS when it is highly activated. In contrast, our model does not have an explicit conflict monitoring system, although the high degree of inhibition in our behavior output layer will serve to function, at least partially, as a conflict resolution system. Inhibition in our behavior layer is set so that only one behavior will be strongly active. It remains to be seen whether an explicit conflict monitoring system for motives is necessary.

Third, although our model and the revised RST are both based on two broad motivational systems, we are also trying to map out specific motives that are governed by the two broad systems. Although RST acknowledges that more specific motives exist, they are not a focus.

Poznanski and Thagard (2005). Poznanski and Thagard (2005) have recently presented a neural network model of personality for virtual characters. It is a feed-forward network in which personality is represented by single nodes corresponding to each pole of the Big Five factors. Specific personalities are coded by assigning baseline activations to each node. For example, someone who is extraverted and disagreeable would have a high activation on both the Extraverted and Disagreeable nodes. Personality nodes are treated as input nodes that send activation to behavior nodes. Behavior nodes also receive inputs from emotion nodes, situation nodes, mood nodes, and relationship nodes. Thus, activation of a given behavior is a joint function of these inputs. Further, emotion nodes receive inputs from situation, relationship, and mood nodes. Poznanski and Thagard described situational features and behaviors fairly abstractly. For example, situations are described as helpful, stressful, hostile, and so on. Behaviors are described as help, avoid, explore, persist, and so on.

This is a useful way to program virtual characters to display personality. Consistent with much current theory, it is argued that personality related behavior is based on differences in levels of the Big Five factors, represented by corresponding nodes. However, this does not provide additional insight into the structure or dynamics of human personality.

Quek and Moskowitz (2007). Quek and Moskowitz (2007) used a three-layer feed-forward neural network to simulate data that had been collected with an event-contingent recording methodology. In one simulation, they had three input nodes corresponding to possible workplace roles (boss, coworker, or supervisee), a hidden layer, and two output nodes for dominant and submissive behavior. The network was trained on a subset of the data and tested on a separate subset. It learned the relationship between work role and dominant versus submissive behavior. In a second simulation, a network learned the relationship between gender and communal behavior in different kinds of relationships, with a different data set. Quek and Moskowitz found that the network identified relationships that had been found in the literature. Thus, they demonstrated that their neural network could learn the empirical relationships between gender- and role-related characteristics of individuals and their behavior in different settings.

Implications

Bridging the gap between personality dynamics and dispositional approaches to personality. A major divide in the field of personality is between those who focus on models of personality dynamics and those who take a dispositional or trait based approach to personality (see Mischel and Shoda, 1998, and Funder, 2001, for a discussion). Dynamic, process oriented models tend to be based on constructs such as goals and motives, beliefs, and delay of gratification or self-control (e.g., Cervone, 2004; Kuhl, 2000; Little, Salmela-Aro, & Phillips, 2006; Mischel & Shoda, 1995) and have focused on how behavior changes over time and situation as a result of personality dynamics. Such approaches tend to be idiographic, focusing on understanding the personality system and personality dynamics of individuals. In contrast, dispositional or trait based approaches have tended to focus on individual differences captured by various trait constructs, such as the Big Five. Here, the focus tends to be on personality constructs that are stable across time and situations. Moreover, this approach is nomethetic and considers trait constructs and trait structure that are revealed by looking across individuals.

This distinction between dispositional and dynamic approaches is almost identical to Cervone's (2005) distinction between interindividual and intraindividual accounts of structure in personality. One conception is the interindividual, psychometric structure of something like the Big Five, the statistical structure of personality revealed when looking across people. Another sense of personality structure is the intraindividual or within person structure of the processing systems that are responsible for the behavior of an individual.

As Cervone (2005) argued, interindividual structure, such as the Big Five, does not have to be explicitly replicated at the intraindividual level. For example, just because interindividual personality structure can be described by something like five broad, largely orthogonal factors, this does not mean that the same structure will also be found in the organization of personality within the individual. Nevertheless, there must be some way in which the intraindividual structure of personality mechanisms results in the interindividual structure of personality.

Unifying these different approaches is essential for an integrated field of personality, but currently, little work has been done to attempt such unification. Mischel and Shoda (1998) argued that dynamic processing approaches, such as their CAPS model, can be integrated with dispositional approaches, and they have taken some initial steps in this direction.

Our neural network model furthers the attempt to bridge the gap between personality dynamics and a dispositional approach to personality. It is a dynamic processing model, based upon the behavior of structured motivational and processing systems. It captures aspects of personality dynamics, so that as the situation and the internal state of the "individual" changes, it produces behavior that varies across time and situations. At the same time, the structure of the motivational and processing systems, modeled with individual differences in parameters, can capture stable individual differences in behavioral tendencies—that is, dispositions. Individual differences in the baseline parameters of this system can be conceptualized as due both to initial genetic and other biological differences, as well as to experiential differences that can "tune" the parameters of the systems. As we argue below when we discuss person–situation interactions and personality dynamics, this model provides both an account of personality dynamics across time and situations and an account of broad, stable dispositions.

Relation of the current model to the structure of individual personality and the Big Five. This model is intended as a potential model of the structure of human personality. But by that we do not mean that any specific instantiation of the model will provide a replication of personality structure, such as the Big Five. Instead, we view any specific set of parameters and learning experiences as representing a particular individual or type of individual.

The Big Five is a representation of the structure of human personality across a group of people. This structure is not seen for a single person but is rather the result of the covariation among characteristics within a large sample of people. Therefore, if we created a large number of virtual individuals, each with a different random set of parameters, we would expect the resulting patterns of behavior across individuals to give us something like the Big Five.

Person–situation interactions. The current model has several important implications for thinking about person–situation interactions. These are clearer if we first review how person and situation are represented. The personality of an actor is primarily represented in terms of aspects of the motive systems, such as the conductances for the different motive systems, their thresholds, the baseline activation of individual goals, and learned weights between situational features and goals and between goals and behavior (personal resources also play a role). These representations are consistent with our argument that personality traits are goal-based structures.

Situations are represented in terms of their relevance to the actor's goals and motives: Configurations of features are linked both to the actor's goals and to the actor's behaviors. That is, a central aspect of how situations are represented is in terms of how their features influence the activation of goals. Another way to frame this is that situations are represented in terms of affordances for goal pursuit. In a recent article, Yang, Read, and Miller (in press) have argued extensively that situations can be conceptualized in terms of the goals whose satisfaction they afford (goal contents) and what happens to those goals (goal processes).

Thus, the person is represented in terms of motivational system and situations that are represented in terms of their affordances for the person's motives. Thus, we can conceptualize person–situation interactions in terms of the interaction between the motive systems (person) and the influence of situational features (situation) on the motive systems. The activation of motives is a function of stable characteristics, such as the sensitivity of the relevant motivational system and the baseline activations of individual motives, as well as the activation that motives receive from situational features. Thus, state motive activation is a joint function of stable individual differences in the motive systems and the impact of present situational features.

In this way, our work provides a specific model of precisely how person and situation might interact and result in patterns of behavior. Rather than providing only a verbal or statistical account of person–situation interactions, it provides a mechanistic, computational model for thinking about how person characteristics interact with features of the situation.

For example, one could keep a particular personality constant and examine how it responded differently to different situations. One could keep a particular situation constant and see how different personalities responded differently to the same situation. Or, one could examine what happens when both personality and situation vary. Simulations 6, 7, and 8, in which we represented individual traits by chronic baseline activations of relevant motives and resources, directly address these kinds of questions. For each of the traits, we found evidence for interactions between the characteristics of the "individual" and the features of the situations.

Personality dynamics. This model also provides a concrete way to think about the dynamics of personality over time and across situations, in terms of the interactions of situational features and the individual's underlying motivational system. Several researchers (e.g., Fleeson, 2001, 2007; D. Heller, Komar, & Lee, 2007) have recently shown that within-individual variability in personality states across time is at least as high as between-person variability in personality traits. For example, for any individual there is at least as much variability in extraversion-related behavior across the day as there is variability between individuals in extraversion. This should probably not be surprising, as the extent of extraversion-related behavior over the course of a day depends on such things as the presence or absence of other people, which vary considerably across the day.

Such variability in personality related behavior is consistent with the current model. In our model, the activation of motives and the choice of behaviors is highly dependent on inputs from the situation. As different situations are encountered, different motives will be more or less highly activated, and the activated motives will "compete" for the control of behavior. One factor influencing whether a particular motive is highly activated and drives personality is whether other motives that compete with the target motive are simultaneously active. The activation of a motive is driven not only by related situations but also by the activation of other motives. This process of person–situation interaction and concurrent motive competition will result in varying personality related behavior across time and situations, as the activation of motives changes.

As we showed in Simulation 6, the model can be set up so that the current state of the network is partially a function of its preceding state. Depending on what other situations have been previously encountered and how they activated the motives, the activation of the motive systems at the current time will differ. This suggests another way in which situational variability—specifically, recently encountered events—contributes to within person variability.

Role of interacting systems and relations to traits. Smillie, Pickering, and Jackson (2006; Corr, 2002, 2004) note that one implication of their interacting systems account of RST is that there is not a direct correspondence between the activation of motivational systems and trait related behavior. For example, since the BAS and the FFFS have an inhibitory relationship, there is not necessarily a direct relationship between the sensitivity of the BAS and the extent of extraverted behavior. Whether a sensitive BAS will lead to extraverted behavior will depend on the strength of both reward and punishment cues.

Although the interaction of motivational systems in our model is somewhat different than in RST, our model also implies that personality will be the result of interactions among systems. There will not necessarily be a direct relationship between the characteristics of a single motivational system and the hypothesized corresponding trait. For example, in our model, the approach and avoidance systems compete for the control of behavior.

Smillie, Pickering, and Jackson (2006; Corr, 2002) suggested that there will be direct relationships between the motivational systems and trait related behavior only in some situations. For example, in a situation in which there are only reward cues, behavior should be a direct result of the sensitivity of the BAS. If individuals are quite extreme on the relevant systems (e.g., very high BAS sensitivity, very low FFFS sensitivity) then the BAS should directly drive behavior. However, most situations have both reward and punishment cues, and most individuals are in the normal range on BAS and FFFS; here motivational systems interact.

The idea that most people have only moderately strong motivational systems also implies that their behavior will be highly sensitive to shifts in the distribution of cues to both reward and punishment. Thus, small shifts in the relative distribution of situational cues to reward and punishment could potentially lead to large shifts in behavior. This has strong implications for thinking about person–situation interactions; there will clearly be nonlinear relationships between personality and situations in the current model (as in Pickering's [2008] computational model of the RST).

Conclusion

This model addresses a number of issues in personality and integrates a number of aspects of personality psychology: the lexical approach to personality and the Big Five; goal-based, dynamic approaches to personality; work on the neurobiology of personality and temperament; and work on the evolutionary tasks that people address in everyday life.

A major goal with the current model is to understand human personality in terms of structured motivational systems. As part of this account, it integrates several different approaches to understanding human personality in terms of motivational systems. Specifically, it brings together work on goal-based models of personality, such as Mischel and Shoda (1995) and L. C. Miller and Read (1987, 1991; Read & Miller, 1989).

The current model also shows how stable traits or dispositions can arise from a dynamic model of personality. In doing so, it points the way toward a possible integration of the dynamic and dispositional approaches to personality. It provides an explicit account of how an individual could display broad, stable, dispositional characteristics, while showing considerable intraindividual variability in behavior across time and situations. These same mechanisms also provide the basis for more fully understanding person–situation interaction.

We believe that explicit modeling of the psychological and neurobiological mechanisms underlying personality dynamics and personality structure has the promise to provide a coherent account of a wide range of phenomena in personality. Such virtual personalities provide a powerful set of tools for hypothesis generation and testing and for theory building and data integration, thereby iteratively advancing the science of human behavior.

References

Allport, G. W. (1962). The general and the unique in psychological science. *Journal of Personality, 30,* 405–422.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review, 11,* 150–166.

Avila, C., & Torrubia, R. (2008). Performance and conditioning studies. In P. J. Corr (Ed.). *The reinforcement sensitivity theory of personality* (pp. 228–260). Cambridge, England: Cambridge University Press.

Bechtel, W., & Abrahamsen, A. (1991). *Connectionism and the mind: An introduction to parallel processing in networks.* Cambridge, MA: Blackwell.

Bugental, D. B. (2000). Acquisition of the algorithms of social life: A domain-based approach. *Psychological Bulletin, 126,* 187–219.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review, 1,* 3–25.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology, 76,* 839–855.

Cantor, N., & Kihlstrom, J. F. (1987). *Personality and social intelligence.* Englewood Cliffs, NJ: Prentice Hall.

Carver, C. S. (2005). Impulse and constraint: Perspective from personality psychology, convergence with theory in other areas, and potential for integration. *Personality and Social Psychology Review, 9,* 312–333.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67,* 319–333.

Cervone, D. (2004). The architecture of personality. *Psychological Review, 111,* 183–204.

Cervone, D. (2005). Personality architecture: Within-person structures and processes. *Annual Review of Psychology, 56,* 423–452.

Cervone, D., & Shoda, Y. (Eds.). (1999). *The coherence of personality: Social-cognitive bases of consistency, variability, and organization.* New York, NY: Guilford.

Chulef, A., Read, S. J., & Walsh, D. A. (2001). A Hierarchical taxonomy of human goals. *Motivation and Emotion, 25,* 191–232.

Church, A. T. (1994). Relating the Tellegen and the five-factor models of personality structure. *Journal of Personality and Social Psychology, 67,* 898–909.

Clark, L. A., & Watson, D. (1999). Temperament: A new paradigm for trait psychology. In L. A. Pervin & O. P. John (Eds.). *Handbook of personality: Theory and research* (2nd edition, 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 edition, pp. 265–286). New York, NY: Guilford Press.

Cloninger, C. R. (1987). A systematic method for clinical description and classification of personality variants. *Archives of General Psychiatry, 44,* 573–588.

Cloninger, C. R. (1998). The genetics and psychobiology of the seven-factor model of personality. In K. R. Silk (Ed.), *Biology of personality disorders. Review of psychiatry disorders* (pp. 63–92). Washington, DC: American Psychiatric Association.

Corr, P. J. (2002). J. A. Gray's reinforcement sensitivity theory: Tests of the joint subsystems hypothesis of anxiety and impulsivity. *Personality and Individual Differences, 33,* 511–532.

Corr, P. J. (2004). Reinforcement sensitivity theory and personality. *Neuroscience and Biobehavioral Reviews, 28,* 317–332.

Davidson, R. J. (1992). Emotion and affective style: Hemispheric substrates. *Psychological Science, 3,* 39–43.

Davidson, R. J. (1995). Cerebral asymmetry, emotion, and affective style. In R. J. Davidson & K. Hugdahl (Eds.), *Brain asymmetry* (pp. 361–387). Cambridge, MA: MIT Press.

Davidson, R. J. (1998). Affective style and affective disorders: Perspectives from affective neuroscience. *Cognition and Emotion, 12,* 307–330.

Davidson, R. J. (2001). Toward a biology of personality and emotion. In A. R. Damasio, A. Harrington, J. Kagan, B. S. McEwen, H. Moss, & R. Shaikh (Eds.), *Unity of knowledge: The convergence of natural and human science. Annals of the New York Academy of Sciences* (Vol. 935, pp. 191–207). New York, NY: New York Academy of Sciences.

Davidson, R. J. (2003). Affective neuroscience: A case for multidisciplinary research. In F. Kessel, P. L. Rosenfield, & N. B. Anderson (Eds.), *Expanding the boundaries of health and social science: Case studies in interdisciplinary innovation* (pp. 99–121). New York, NY: Oxford University Press.

Davidson, R. J., Jackson, D. C., & Kalin, N. H. (2000). Emotion, plasticity, context, and regulation: Perspectives from affective neuroscience. *Psychological Bulletin, 126,* 890–909.

Depue, R. A. (1996). A neurobiological framework for the structure of personality and emotion: Implications for personality disorders. In J. Clarkin & M. Lenzenweger (Eds.) *Major theories of personality disorders* (pp. 347–390). New York, NY: Guilford Press.

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,* 491–569.

Depue, R. A., & Morrone-Strupinsky, J. V. (2005). A neurobehavioral model of affiliative bonding: Implications for conceptualizing a human trait of affiliation. *Behavioral and Brain Sciences, 28,* 313–395.

Derryberry, D., & Rothbart, M. K. (1997). Reactive and effortful processes in the organization of temperament. *Development and Psychopathology, 9,* 633–652.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41,* 417–440.

Digman, J. M. (1997). Higher order factors of the Big Five. *Journal of Personality and Social Psychology, 73,* 1246–1256.

Downey, G., & Feldman, S. I. (1996). Implications of rejection sensitivity for intimate relationships. *Journal of Personality and Social Psychology, 70,* 1327–1343.

Eisenberg, N. (2002). Emotion-related regulation and its relation to quality of social functioning. In W. Hartup & R. A. Weinberg (Eds.), *Child psychology in retrospect and prospect: In celebration of the 75th anniversary of the Institute of Child Development. The Minnesota Symposia on Child Psychology* (Vol. 32, pp. 133–171). Mahwah, NJ: Erlbaum.

Eisenberg, N., Smith, C. L., Sadovsky, A., & Spinrad, T. L. (2004). Effortful control: Relations with emotion regulation, adjustment, and socialization in childhood. In R. F. Baumeister & K. D. Vohs (Eds.). *Handbook of self-regulation: Research, theory, and applications* (pp. 259–282). New York, NY: Guilford Press.

Eiser, J. R., Fazio, R. H., Stafford, T., & Prescott, T. J. (2003). Connectionist simulation of attitude learning: Asymmetries in the acquisition of positive and negative evaluations. *Personality and Social Psychology Bulletin, 29,* 1221–1235.

Elliot, A. J., Gable, S. L., & Mapes, R. R. (2006). Approach and avoidance motivation in the social domain. *Personality and Social Psychology, 32,* 378–391.

Emmons, R. A. (1991). Personal strivings, daily life events, and psychological and physical well-being. *Journal of Personality, 59,* 453–472.

Eysenck, H. J. (1967). *The biological basis of personality.* Springfield, IL: Charles C. Thomas.

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. Vernson (Ed.), *The neuropsychology of individual differences* (pp. 151–207). San Diego, CA: Academic Press.

Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology, 87,* 293–311.

Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review, 99,* 689–723.

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80,* 1011–1027.

Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of Research in Personality, 75,* 825–861.

Funder, D. C. (2001). Personality. *Annual Review of Psychology, 52,* 197–221.

Gable, S. L., Reis, H. T., & Elliot, A. J. (2000). Behavioral activation and inhibition in everyday life. *Journal of Personality and Social Psychology, 78,* 1135–1149.

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.

Gosling, S. D., & John, O. P. (1999). Personality dimensions in nonhuman animals: A cross-species review. *Current Directions in Psychological Science, 8,* 69–75.

Gosling, S. D., Kwan, V. S. Y., & John, O. P. (2003). A dog's got personality: A cross-species comparative approach to evaluating personality judgments. *Journal of Personality and Social Psychology, 85,* 1161–1169.

Gray, J. A. (1987a). The neuropsychology of emotion and personality. In J. 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* (2nd ed.). New York, NY: Cambridge.

Gray, J. A. (1988). The neuropsychological basis of anxiety. In C. G. Last & M. Hersen (Eds.), *Handbook of anxiety disorders. Pergamon general psychology series* (Vol. 151, pp. 10–37). Elmsford, NY: Pergamon 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.

Harmon-Jones, E., & Sigelman, J. (2001). State anger and prefrontal brain activity: Evidence that insult-related relative left-prefrontal activation is associated with experienced anger and aggression. *Journal of Personality and Social Psychology, 80,* 797–803.

Heller, D., Komar, J., & Lee, W. B. (2007). The dynamics of personality states, goals, and well-being. *Personality and Social Psychology Bulletin, 33,* 898–910.

Heller, W. (1993). Neuropsychological mechanisms of individual differences in emotion, personality, and arousal. *Neuropsychology, 7,* 476–489.

Heller, W., Schmidtke, J. I., Nitschke, J. B., Koven, N. S., & Miller, G. A. (2002). States, traits, and symptoms: Investigating the neural correlates of emotion, personality, and psychopathology. In D. Cervone, & W. Mischel (Eds.), *Advances in personality science* (pp. 106–126). New York, NY: Guilford Press.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computation.* Redwood City, CA: Addison-Wesley.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values.* Beverly Hills, CA: Sage.

Hofstee, W. K. B., de Raad, B., & Goldberg, L. R. (1992). Integration of the Big Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology, 63,* 146–163.

Ito, T. A., & Cacioppo, J. T. (2000). Electrophysiological evidence of implicit and explicit categorization processes. *Journal of Experimental Social Psychology, 36,* 660–676.

Ito, T. A., & Cacioppo, J. T. (2001). Affect and attitudes: A social neuroscience approach. In J. P. Forgas (Ed.), *The handbook of affect and social cognition* (pp. 50–74). Hillsdale, NJ: Erlbaum.

Ito, T. A., & Cacioppo, J. T. (2005). Variations on a human universal: Individual differences in positivity offset and negativity bias. *Cognition and Emotion, 19,* 1–26.

James, W. (1981). *The principles of psychology* (Vol. 1). Cambridge, MA: Harvard University Press. (Original work published 1890)

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.) *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). New York, NY: Guilford Press.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.). *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York, NY: Guilford 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.

Kochanska, G., & Knaack, A. (2003). Effortful control as a personality characteristic of young children: Antecedents, correlates, and consequences. *Journal of Personality, 71,* 1087–1112.

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 & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 111–169). San Diego, CA: Academic Press.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research, 39,* 329–358.

Little, B. R., Salmela-Aro, K., & Phillips, S. D. (2006). *Personal project pursuit: Goals, action, and human flourishing.* Mahwah, NJ: Erlbaum.

Majani, E., Erlarson, R., & Abu-Mostafa, Y. (1989). The induction of multiscale temporal structure. In D. S. Touretsky (Ed.), *Advances in neural information processing systems* (pp. 634–642). San Mateo, CA: Morgan Kaufmann.

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology, 88,* 139–157.

Martindale, C. (1991). *Cognitive psychology: A neural-network approach.* Pacific Grove, CA: Brooks/Cole.

McClelland, J. L., & Rumelhart, D. E. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models.* Cambridge, MA: MIT Press.

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.

Mehl, M. R., Gosling, S. D., & Penebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology, 90,* 862–877.

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

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.

Miller, N. E. (1959). Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation and social learning. In S. Koch (Eds.), *Psychology: A study of a science. Study* 1 (pp. 198–292). New York, NY: McGraw-Hill.

Mischel, W. (1973). Toward a cognitive social learning reconceptualizaton of personality. *Psychological Review, 80,* 252–283.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102,* 246–268.

Mischel, W., & Shoda, Y. (1998). Reconciling processing dynamics and personality dispositions. *Annual Review of Psychology, 49,* 229–258.

Mischel, W., & Shoda, Y. (2008). Toward a unified theory of personality: Integrating dispositions and processing dynamics within the cognitive-affective processing system. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 208–241). New York, NY: Guilford Press.

Monroe, B., Read, S. J., & Miller, L. C. (2006). *Mapping personality space: Uncovering the structure of personality dynamics.* Los Angeles, CA: University of Southern California.

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin, 126,* 220–246.

Nigg, J. T., John, O. P., Blaskey, L. G., Huang-Pollock, C. L., Willcutt, E. G., Hinshaw, S. P., & Pennington, B. (2002). Big Five dimensions and ADHD symptoms: Link between personality traits and clinical syndromes. *Journal of Personality and Social Psychology, 83,* 451–469.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron, 28,* 329–337.

O'Doherty, J. P., Buchanan, T. W., Seymour, B., Raymond, J., & Dolan, R. J. (2006). Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum. *Neuron, 49,* 157–166.

O'Reilly, R. C. (1996). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation, 8,* 895–938.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain.* Cambridge, MA: MIT Press.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions.* New York, NY: Oxford University Press.

Peabody, D., & De Raad, B. (2002). The substantive nature of psycholexical personality factors: A comparison across languages. *Journal of Personality and Social Psychology, 83,* 983–997.

Pervin, L. A. (1989). *Goal concepts in personality and social psychology.* Hillsdale, NJ: Erlbaum.

Pervin, L. A. (1990). A brief history of modern personality theory. In L. A.

Pervin (Ed.), *Handbook of personality theory and research* (pp. 3–18). New York, NY: Guilford Press.

Pickering, A. D. (2004). *A formal model of "new" reinforcement sensitivity theory.* Paper presented at the 12th European Conference on Personality, Groningen, the Netherlands. Retrieved from http://homepages.gold.ac.uk/aphome/ecpl2.ppt

Pickering, A. D. (2008). Formal and computational models of reinforcement sensitivity theory. In P. J. Corr (Ed.), *The reinforcement sensitivity theory of personality* (pp. 453–481). Cambridge, England: Cambridge University Press.

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* (2nd ed., pp. 277–299). New York, NY: Guilford Press.

Pickering, A. D., & Smillie, L. D. (2008). The behavioural activation system: Challenges and opportunities. In P. J. Corr (Ed.), *The reinforcement sensitivity theory of personality* (pp. 120–154). Cambridge, England: Cambridge University Press.

Poznanski, M., & Thagard, P. (2005). Changing personalities: Towards realistic virtual characters. *Journal of Experimental and Theoretical Artificial Intelligence, 17,* 221–241.

Quek, M., & Moskowitz, D. S. (2007). Testing neural network models of personality. *Journal of Research in Personality, 41,* 700–706.

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

Read, S. J., & Miller, L. C. (1998). On the dynamic construction of meaning: An interactive activation and competition model of social perception. In S. J. Read & L. C. Miller (Eds.), *Connectionist models of social reasoning and behavior* (pp. 27–68). Mahwah, 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.

Read, S. J., Miller, L. C., Monroe, B., Brownstein, A., Zachary, W., LeMentec, J. C., & Iordanov, V. (2006). A neurobiologically inspired model of personality. In J. Gratch, M. Young, & R. Aylett (Eds.), *Intelligent virtual agents: 6th Internation Conference, IVA 2006* (pp. 316–328). Berlin, Germany: Springer-Verlag.

Read, S. J., Vanman, E. J., & Miller, L. C. (1997). Connectionism, parallel constraint satisfaction processes and gestalt principles: (Re) introducing cognitive dynamics to social psychology. *Personality and Social Psychology Review, 1,* 26–53.

Romero-Canyas, R., & Downey, G. (2005). Rejection sensitivity as a predictor of affective and behavioral responses to interpersonal stress: A defensive motivational system. In K. Williams, J. Forgas, & W. Von Hippel (Eds.), *The social outcast: Ostracism, social exclusion, rejection and bullying* (pp. 131–151). New York, NY: Psychological Press.

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

Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations.* Cambridge, MA: MIT Press.

Saucier, G., & Ostendorf, F. (1999). Hierarchical subcomponents of the Big Five personality factors: A cross-language replication. *Journal of Personality and Social Psychology, 76,* 613–627.

Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *The Journal of Neuroscience, 27,* 12860–12867.

Schmidtke, J. I., & Heller, W. (2004). Personality, affect, and EEG: Predicting patterns of regional brain activity related to extraversion and neuroticism. *Personality and Individual Differences, 36,* 717–732.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275,* 1593–1599.

Shoda, Y., & Mischel, W. (1998). Personality as a stable cognitive–affective activation network: Characteristic patterns of behavior variation emerge from a stable personality structure. In S. J. Read & L. C. Miller (Eds.), *Connectionist models of social reasoning and social behavior* (pp. 175–208). Mahwah, NJ: Erlbaum.

Smillie, L. D., Dalgleish, L. I., & Jackson, C. J. (2007). Distinguishing between learning and motivation in behavioral tests of the reinforcement sensitivity theory of personality. *Personality and Social Psychology Bulletin, 33,* 476–489.

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.

Spoont, M. R. (1992). Modulatory role of serotonin in neural information processing: Implications for human psychopathology. *Psychological Bulletin, 112,* 330–350.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.

Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 681–706). Hillsdale, NJ: Erlbaum.

Tellegen, A., & Waller, N. G. (2008). Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire. In G. J. Boyle, G. Matthews, and D. H. Saklofske (Eds.), *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.

Tomarken, A. J., Davidson, R. J., Wheeler, R. E., & Doss, R. C. (1992). Individual differences in anterior brain asymmetry and fundamental dimensions of emotion. *Journal of Personality and Social Psychology, 62,* 676–687.

Triandis, H. C. (1995). *Individualism and collectivism.* Boulder, CO: Westview Press.

Watson, D., & Clark, L. A. (1997). Extraversion and its positive emotional core. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology.* (pp. 767–793). San Diego, CA: Academic Press.

Watson, D., Wiese, D., Vaida, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76,* 820–838.

Widiger, T. A., & Smith, G. T. (2008). Personality and psychopathology. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 743–769). New York, NY: Guilford Press.

Widrow, G., & Hoff, M. E. (1960). Adaptive switching circuits. Institute of Radio Engineers. Western Electronic Show and Convention. *Convention Record, Part 4,* 96–104.

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.

Yang, Y., Read, S. J., & Miller, L. C. (in press). The concept of situations. *Social and Personality Psychology Compass.*

Zald, D. H., & Depue, R. A. (2001). Serotonergic functioning correlates with positive and negative affect in psychiatrically healthy males. *Personality and Individual Differences, 30,* 71–86.

Zinbarg, R. E., & Mohlman, J. (1998). Individual differences in the acquisition of affectively valenced associations. *Journal of Personality and Social Psychology, 74,* 1024–1040.

Zuckerman, M. (2002). Zuckerman-Kuhlman personality questionnaire (ZKPQ): An alternative five-factorial model. In B. DeRaad & M. Perugini (Eds.), *Big Five assessment* (pp. 376–392). Ashland, OH: Hogrefe & Huber.

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

Received September 24, 2007
Revision received October 23, 2009
Accepted October 24, 2009 ■