A Connectionist Approach to the Organization and Continuity of Working Models of Attachment

R. Chris Fraley

University of Illinois at Urbana-Champaign

ABSTRACT Most research on adult attachment dynamics has been conducted under the assumption that working models are generalized cognitive-motivational structures that are highly stable and activated in a wide array of circumstances. Recent research, however, suggests that people develop attachment representations that are relationship specific, leading them to hold distinct working models in different kinds of relationships. The objective of this article is to outline a connectionist approach to the working model construct that has the potential to explain how global and relationship-specific working models are organized within the same mental system and how different learning environments can support continuity in those representations over time.

One of the core assumptions of adult attachment theory is that individuals construct mental representations, or *working models*, of the self and significant others based on their interpersonal experiences. Individuals who have had a history of warm and responsive interactions with their caregivers, for example, are assumed to develop secure representations of themselves and significant others, whereas individuals who have had a history of inconsistent or unresponsive caregiving develop insecure representations of themselves and others. Importantly, these representations are believed to play a crucial role in the way people interpret and understand their social relationships. As such, they are valuable constructs for understanding interpersonal processes, personality dynamics, and development.

Traditionally, attachment researchers have conceptualized working models in a trait-like fashion, assuming these representations are relatively stable over time and influential across a wide array of

Correspondence concerning this article may be sent to R. Chris Fraley, University of Illinois at Urbana-Champaign, Department of Psychology, 603 East Daniel Street, Champaign, IL 61820. E-mail: rcfraley@uiuc.edu.

Journal of Personality 75:6, December 2007

DOI: 10.1111/j.1467-6494.2007.00471.x

relational contexts, including relationships with parents, friends, and romantic partners. In recent years, however, scholars have called into question the assumption that working models have trait-like properties (Baldwin, Keelan, Fehr, Enns, & Koh-Rangarajoo, 1996; La Guardia, Ryan, Couchman, & Deci, 2000; Pierce & Lydon, 2001). Baldwin and his colleagues (1996), for example, demonstrated that there is considerable within-person variability in the expectations and beliefs that people hold about significant others in their lives. A person may consider his or her spouse to be warm, affectionate, and responsive, while simultaneously viewing his or her mother as being cold, rejecting, and aloof. The fact that substantial within-person variation exists in the way people relate to others raises a number of controversial questions about how working models should be conceptualized in research on adult attachment.

The objective of this article is to offer a theoretical approach to the working model construct that has the potential to clarify a number of questions about how working models operate in different contexts and how they change over time. One of the themes of this article is that classic models of memory and cognition do not provide an ideal foundation upon which to consider these kinds of issues. As an alternative, I explore working models within the framework of connectionist theories of memory. Connectionist, or Parallel Distributed Processing (PDP), models are based on the assumption that familiar memory phenomena, such as schematic processing, are the emergent properties of the activity of networks of massively interconnected neurons (Rumelhart, McClelland, & the PDP Research Group, 1986). Connectionist models have become increasingly influential in cognitive science over the past two decades because they have proven useful for understanding learning, memory, and perception. Moreover, connectionist ideas have provided a much needed bridge between diverse subdisciplines in psychology. Researchers with backgrounds ranging from developmental psychology (Elman et al., 1996; Schultz, 2003), social psychology (Queller, 2002; Smith, 1996), personality (Read & Miller, 1998; Shoda, LeeTiernan, & Mischel, 2002), to neuroscience (Rumelhart, McClelland, et al., 1986) have drawn upon connectionism as a framework for understanding some of the most enduring questions in psychology.

I begin this article with a brief discussion of the distinctions between classic symbolic models of cognition and connectionist models. Next, I review some contemporary debates regarding the

question of whether working models are best viewed as global, generalized structures, relationship-specific structures, or some combination of these two alternatives. I argue that a connectionist perspective offers some novel ways to think about these issues and will report a series of simulations that demonstrate that a simple connectionist model offers (a) a mechanism by which the same representational network can store representations of multiple significant others, both global representations and relationship-specific ones; (b) an account of how representations can have hierarchical properties, as has been documented by some researchers (e.g., Overall, Fletcher, & Friesen, 2003), without explicitly positing hierarchical relationships among elements; and (c) a novel framework in which to understand continuity and change in attachment patterns.

Although this article has been inspired primarily by debates in the study of attachment, the connectionist framework offers the potential to address parallel issues in the study of personality more generally. As many readers are aware, the last few decades have been riddled with controversy over the legitimacy of the trait concept (see Kenrick & Funder, 1988). Some critics have offered social-cognitive alternatives to trait models, arguing that such models provide a stronger foundation for understanding the proximate mechanisms underlying individual differences in behavior (Cervone, 1997). One theme that will be emphasized in this article is that, at least for individual-difference constructs that are assumed to have important cognitive components, a connectionist perspective offers a valuable way to understand both the trait-like properties of representational systems as well as their more context-dependent and differentiated features. It is possible, in other words, for a single framework to capture what have emerged as "alternative" positions in contemporary personality psychology.

A BRIEF OVERVIEW OF CONNECTIONIST MODELS

Connectionist theories were initially developed to model the dynamics of simple neural systems (see Churchland & Sejnowski, 1992, for a review). A prototypical connectionist network is composed of a number of *units*, loosely analogous to neurons, that are connected to one another via excitory and inhibitory pathways (see Figure 1). Like biological neurons, connectionist units can vary in their activity

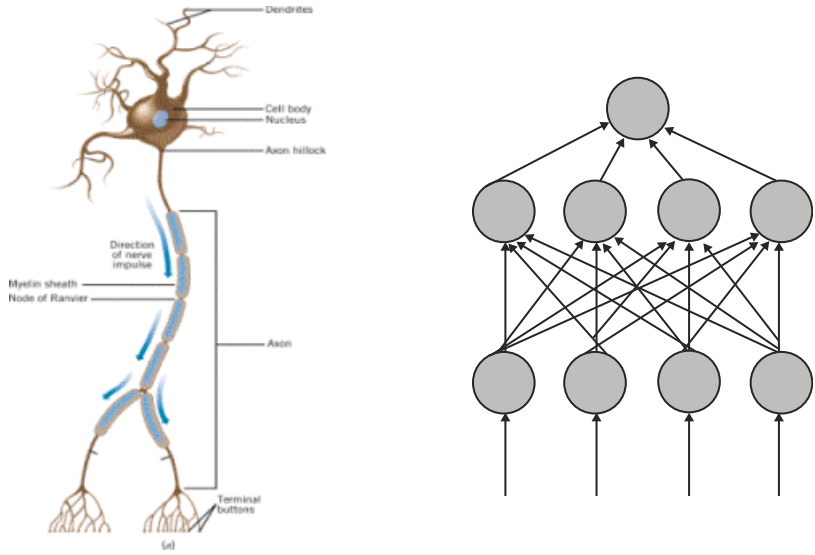

Figure 1
Illustrations of the way activation flows through a biological neuron (left) and artificial neural network (right).

levels, sometimes being relatively inactive and, at other times, exhibiting increases or decreases in activity. Importantly, an active unit can pass its activation along to other units to which it is connected. Thus, activation originating from the external world or from a specific locus in the network can spread across a network in much the same way that activation may spread in a biological neural network (see Figure 1).

Although certain features, such as connectivity, are common to all connectionist networks, there are a variety of different architectures—that is, ways of organizing the connections and functions of units—used in connectionist simulations. For example, in some models there is a layer of units corresponding to feature detectors in a perceptual system. Activation from those perceptual units may pass through an intermediate level of units (e.g., "hidden layers") before being channeled into *response units*—units that enable the behavior enacted or decision reached by the system. Other networks may be composed of units that are fully connected to one another. These *recurrent networks* are frequently used in social-psychological applications of connectionist theory (see Smith, 1996).

Connectionist perspectives on cognition differ in some important ways from traditional, symbolic theories of cognition. The most important difference concerns the conceptualization of representation. In classic symbolic models, a concept is represented as a node in an associative or hierarchical network (e.g., Collins & Quillian, 1969). These symbols are discrete and do not overlap with other representations. Concepts can be combined or joined through the addition of conjoining features (''isa'' links; see Anderson, 1993). In contrast, in a connectionist network, knowledge is distributed across the connections among multiple units, and, importantly, the same units can be involved in the representation of independent patterns. In a connectionist model, a representation can be nothing more than the pattern of activity across units and, as such, may not be a ''thing'' (e.g., a node or unit) that is operated upon in any strict sense.

Another way in which traditional and connectionist models differ is with respect to process. Classic models of memory tend to focus on the different kinds of processes that operate upon memory units. They may, for example, postulate distinct encoding and retrieval processes or assume that specific processes operate upon elements that are stored in memory. In connectionist models, the process of retrieval is not viewed as a distinct operation per se but, instead, as a reinstatement of a pattern of activation that corresponds to a concept. Whether the pattern is reactivated through internal or external means, the basic process is the same.

Finally, and most importantly for our purposes, in a connectionist network, many of the representations involved in cognition are not built into the system in advance. Rather, they are acquired over the course of the network's learning history. In a typical connectionist simulation, the network may be exposed to a variety of different training patterns (e.g., words) to determine if the network can accurately distinguish different classes of stimuli. When the network performs incorrectly (e.g., by responding as if it has been presented with the word ''coffin'' instead of ''coffee''), the connections among units are modified in a way that reduces the probability that a similar error will be made in the future. There are a variety of ''learning rules'' that can be implemented in artificial neural networks. One of the most intuitive is referred to as the *delta rule*, a rule inspired by Hebbian learning theories (see Rumelhart, Hinton, & Williams, 1986). In this form of supervised learning, the connections (also known as ''weights'') between any two units are adjusted in a manner

that is proportional to the magnitude of the error produced by the network. This kind of learning rule functions to create stronger connections between units that tend to be simultaneously active during learning. There is much debate over what kinds of learning rules are most efficient, as well as which rules are the most biologically plausible (see Rumelhart, Hinton, & Williams, 1986, for a discussion). The important point for our purposes is that, regardless of the precise way in which connection weights are updated over time, the network gradually constructs a set of representations that allow it to understand and interact with its world. In symbolic models of memory and cognition, the basic concepts often exist in the model's "lexicon" from the start, and learning and development per se are not typically addressed.

In many respects, the connectionist revolution has altered the course of the cognitive sciences (Macdonald & Macdonald, 1995). One reason for the popularity of connectionist models is that they are able to account for some basic psychological phenomena with a minimal number of assumptions. For example, connectionist networks can capture the schematic functioning of human memory systems (Rumelhart, Smolensky, McClelland, & Hinton, 1986). They can generalize their knowledge to novel stimuli in both appropriate ways (e.g., responding to similar stimuli in the same fashion) and inappropriate ways that mirror human errors (e.g., misusing irregular verb tenses). Moreover, like natural memory systems, they exhibit "graceful degradation"— gradual impairments in performance as the system becomes physically compromised. This is not to say that connectionist frameworks are perfect; they have many limitations (see Macdonald & Macdonald, 1995). Nonetheless, because they have proved useful in many lines of cognitive inquiry, their potential for understanding personality dynamics is worthy of consideration.

APPLICATIONS OF CONNECTIONISM TO ATTACHMENT

Attachment researchers have tended to conceptualize working models as generalized representations—representations that capture the broad, as opposed to specific, relational themes across a variety of interpersonal experiences. This approach, which has sometimes been referred to as a "trait" or "individual-centered" approach (see

Kobak, 1994; Lewis, 1994), has been popular for a number of reasons. For example, if it is the case that early childhood experiences with caregivers lead to the formation of cognitive structures that are relatively general and stable, then a relational mechanism exists that can be used to understand how it is that people create continuity and coherence across their important relationships. Although there would undoubtedly be variations from one relationship to the next in how the person relates to significant others, the trait perspective implies that there will be a common thread tying together the individual's thoughts, feelings, and behavior across these unique contexts.

Despite its appeal, the trait approach to the study of attachment has been criticized on at least two grounds. First, scholars have observed that people exhibit different attachment patterns across different relationships. For example, people who are relatively secure with their mothers may or may not be secure with their romantic partners (Baldwin et al., 1996). Such findings have been interpreted as suggesting that working models may be too context dependent to be meaningfully viewed as characteristics of persons rather than situations. In addition, researchers have noted that the test-retest stability of attachment patterns is low, even when attachment patterns are assessed across relatively short-term intervals (e.g., Baldwin & Fehr, 1995). If working models are not highly stable across a period as brief as 2 weeks, for example, how can they reflect general, enduring features of people's personalities?

In the following sections I address these issues, focusing primarily on the ways in which a connectionist perspective may help advance the way we consider questions regarding global versus specific models and stability and change. One of the key points I will make is that traditional models of cognition do not allow us to resolve these issues easily. However, when working models are viewed from the lens of connectionist theory, an intriguing set of solutions to these problems emerges.

Are Working Models Generalized Cognitive Structures or Representations That Are Specific to Relational Contexts?

The observation that people do not always exhibit similar attachment patterns in different contexts was initially documented in research on infant attachment (e.g., Fox, Kimmerly, & Schafer, 1991).

Most of the early research on infant attachment had focused on infants' attachments to their mothers. Once researchers began focusing on other family members as well, it became apparent that children who are secure with their mothers may or may not be secure with other caregivers, such as fathers. Fox et al. (1991) provided a review of these data, arguing that the correspondence between attachment classifications with mother and father were roughly equivalent to a correlation of .30. Similar findings have been obtained in the study of adult attachment in the social-personality tradition. Baldwin and his colleagues (1996) have shown that there is considerable within-person variability in the expectations and beliefs that people hold about significant others. For example, people may report being relatively secure with their parents but report insecurities with their romantic partners. In fact, according to research by Baldwin and his colleagues (1996) and Klohnen, Weller, Luo, and Choe (2005), the average correlation between security measured in romantic relationships and security measured in parental relationships is approximately .20. If people truly hold a generalized working model of attachment, it might seem that the security experienced across different relationships would be more consistent.

The fact that within-person variation exists in the way people relate to important others in their lives raises a number of questions about how working models should be conceptualized. One possibility, albeit an extreme one, is that there is no such thing as a global model of attachment. It may be the case that the ratings obtained in commonly used self-report measures of attachment are based on the on-line inferences that a person makes when instructed to think back across his or her important relationships (i.e., a person's rating is a weighted average of the security of his or her relationship-specific representations). Most researchers, however, have argued that although relationship-specific attachment representations exist, general or more abstract representations exist as well (see Collins & Read, 1994). In an influential paper, Collins and Read (1994) proposed that these distinct representations are hierarchically arranged, such that relationship-specific representations (e.g., those relevant to one's spouse) are nested within representations of broader relational categories (e.g., romantic partners), which, in turn, may be nested under even broader categories (e.g., others; see Figure 2). Overall and colleagues (2003) tested his hypothesis in a study in which participants were asked to rate their security with respect to a variety of different

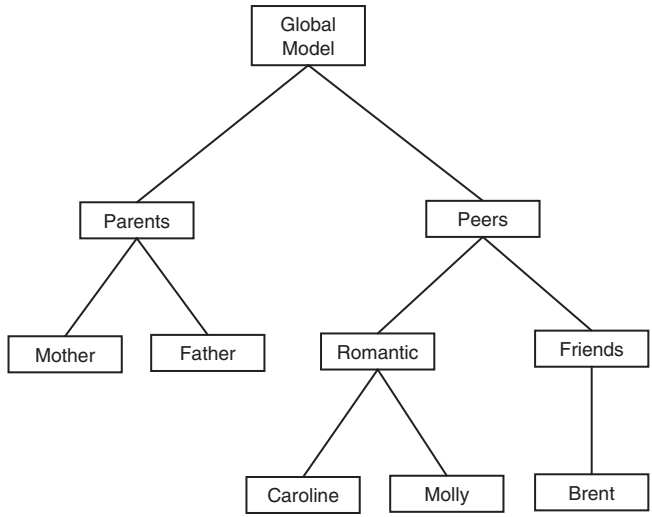

Figure 2
The hierarchical model of the organization of working models of attachment.

people in their lives, including parents and romantic partners. They fit the data to a hierarchical factor model in which security in each specific relationship was assumed to be influenced by higher-order factors (i.e., romantic, familial, and peer relationships), which were, in turn, organized by a more overarching factor (e.g., a global working model of others). Their analyses suggested that a hierarchical model was able to reproduce the data well.

Although the hierarchical framework has proven useful in helping to explain why people may vary in security from one relationship to the next, it is important to note that the hierarchical model is a difficult one to evaluate empirically. The one empirical test of the model that has been published (Overall et al., 2003) did not examine a crucial alternative explanation, one easily derived from a developmental perspective on attachment dynamics, namely, that if one kind of relationship-specific representation is forged in part on the basis of those that already exist, then we would expect a modest degree of association in security across these different life domains. For example, if one relationship-specific representation (i.e., that pertaining to one's mother) was constructed before another one (i.e., that concerning one's partner), and if the former played a role in shaping the

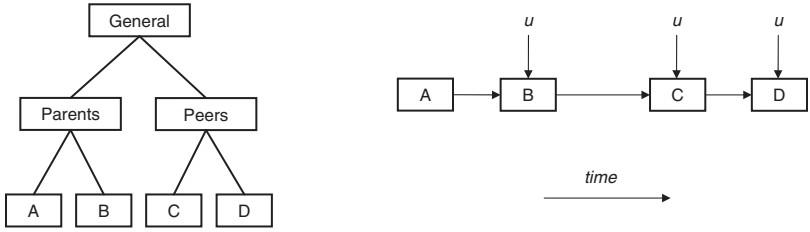

Figure 3
Alternative models for explaining the pattern of associations among security as rated in four relationships, A, B, C, and D. The left-most panel illustrates the structure of working models according to a simple hierarchial perspective. The right-most panel shows the relations among relationship-specific working models according to a simple developmental model. The influences of unique experiences that are uncorrelated with existing representations are denoted as "u."

latter, then the two sets of relational experiences would be similar (and, thus, correlated across people; see Figure 3). In this situation, a hierarchical model that postulates a global representation would be able to explain the data, but, in fact, there is no global model in this scenario. Longitudinal data would be needed to differentiate these alternative accounts for the pattern of associations among measures of security in different relationships.

Another potential limitation of the hierarchical perspective is that traditional models of cognition do not provide a clear framework for understanding how general representations might be abstracted from relationship-specific experiences. The key problem is that there is no obvious way to demonstrate by using classic models of memory that global models can be constructed without building into the model, a priori, a set of processes that enable this to take place (see Rumelhart & McClelland, 1986). As such, traditional models do not explain how both global and relationship-specific representations develop; they simply assume that both kinds of representations exist.

One of the important discoveries in early connectionist research was that simple connectionist models could extract the underlying structure of a set of patterns after repeated exposures to those patterns. Thus, if a network were trained to recognize a variety of idiosyncratic stimuli (e.g., specific cats that may have differed from one

another in some unique ways, such as size, hair color, spot patterns), it would develop a representation that captured the features that those stimuli had in common (e.g., meows, has four legs and a tail, prefers milk without coffee, and is aloof). Importantly, the network would develop a representation of a prototypical example even if it were never exposed to the prototype per se. This finding suggests that the same processes that support the learning of specific exemplars enable the development of a more abstract or global representation of those exemplars. In other words, there is not a specific cognitive process that gives rise to the development of the abstracted representation; instead, the global representation emerges naturally as knowledge concerning the specific exemplars is learned. This finding is crucial because it suggests that global representations of attachment can be born from relationship-specific experiences.

Although previous researchers have demonstrated the ways in which global or "prototypical" patterns can be learned by artificial neural networks, one of the goals of this article is to illustrate these ideas specifically in the context of attachment theory. Thus, in this section, I present a simulation—couched in the language of attachment theory—that demonstrates how a simple connectionist network can extract a global representation based on relationship-specific experiences.

I begin with a basic version of the problem. In the context of a single relationship, a relationship partner does not always behave in a consistent fashion. A parent who is warm and responsive on most occasions may be unhelpful and distant at other times. Thus, one challenge for the network is to develop a unified representation of a specific exemplar, given that behavior of the exemplar itself is a dynamic over time.

The details of the connectionist network used in these simulations are described in the Appendix. For the sake of discussion, let us assume that the various nodes in the network correspond to concepts such as "caring," "warm," "sensitive to my needs," and "cold." We assume that the caregiver has a latent profile for these traits, such that he or she is truly caring, warm, sensitive, and not cold. However, in each interaction, this latent profile will be expressed imperfectly. For example, during the first trial, the caregiver may exhibit caring and warm qualities but not express sensitivity or coldness. On another interaction, the caregiver may come across as cold. The important point is that although the caregiver has a specific profile of

qualities, the qualities expressed are partial reflections of that profile. The network never sees the "true" profile in any one interaction; it only experiences random permutations of it over time.

The key question is whether the network is capable of extracting a representation of the latent profile despite being exposed only to statistical derivations of it. To examine this issue, a latent profile of qualities was constructed and 50 derivations of it were created. These derivations were presented to the network one at a time. On each trial, the network was presented with the pattern and, after the activation stabilized, the weights were adjusted according to the delta rule. The network's ability to reproduce various patterns, including the unseen prototype, was tested by presenting the pattern to the network and correlating the network's response (i.e., the profile of activation levels of the nodes after the network had settled into a stable pattern of activation) with the test pattern. Fifty simulations were run, using a randomly selected prototype/latent profile for each one. The average of the results are reported below.

The network was able to reproduce all of the patterns it experienced remarkably well (average $r = .90$). Most importantly, however, it was able to reproduce the prototypical pattern accurately ($r = .99$), even though it was never exposed to the prototypical pattern per se. In other words, the network abstracted knowledge regarding what was common to the patterns it was experiencing.

The previous demonstration shows that a simple connectionist network can exhibit properties that are critical to our understanding of attachment and cognition. Namely, it shows that general or global models can be constructed on the basis of repeated experiences—experiences that are similar in nature but not necessarily redundant. It is noteworthy that there is not a special process that functions to construct this abstract representation. Instead, the global representation emerges from the basic processes that allow the network to acquire knowledge about any one interaction.

Based on the same principles, it is also possible for the same network to acquire representations of distinct prototypes. For example, if the network were exposed repeatedly to derivatives of two correlated patterns, it would extract two prototypes, one for each pattern. For the first few trials, the network would attempt to assimilate the features of the new pattern into its existing knowledge base, thereby creating an opportunity for a variety of inferential "errors" that are of interest to the psychodynamically inclined. However, after

repeated experiences with the second pattern, it would construct a representation uniquely suited to that pattern.

Summary. Connectionist models have several implications for how theorists conceptualize the general versus specific problem in attachment research. First, on the basis of varied experiences with a person, individuals can develop representations both of those specific experiences and whatever is common to them—a global representation or a prototype. Second, if the individual interacts with multiple relationship partners, a representation that captures the core features of each relationship is extracted as well as one that captures what is common to each of those relationships.

It is important to note that the same network is responsible for representing both the global and specific features of relationships. In other words, there is not a portion of the network that holds knowledge about mother and a distinct portion that contains knowledge about the romantic partner. The knowledge is fully distributed across the network. Although this is a subtle matter, it provides an interesting point of comparison to the hierarchical metaphor that is commonly used in attachment theory. A more important note, perhaps, is that the connectionist framework suggests that the similarity between the more "objective" features of significant others is responsible for the way representations become organized. For example, within a connectionist network, if a new relationship partner partly resembles one's mother, knowledge regarding the mother will be brought to bear on the interpretation of the new person—even though the new person exists in a different social category (i.e., potential romantic partner vs. a parent). In contrast, the hierarchical perspective suggests that the new partner will be categorized as a romantic partner and that knowledge about previous romantic partners will be more likely to guide the development of the new relationship than knowledge about one's mother.

Continuity and Change in Representational Models

The previous simulations demonstrate that a simple connectionist model is capable of developing relationship-specific and global attachment representations on the basis of repeated experiences with caregivers. The network acquires its knowledge by adjusting its connection weights in response to new information. As it learns each

pattern, the weights have to be reconfigured slightly to accommodate the new information. The fact that the network's weights are gradually revised over time raises some intriguing questions about continuity and change in representational states. Specifically, if the values of the network's weights are evolving as it learns new patterns, what is the fate of the representations developed early in the network's learning history? This question is a critical one from an attachment perspective. One of the core assumptions of attachment theory is that the representations constructed early in life are relatively stable and have consequences for interpersonal relationships across the lifespan. If it is the case, however, that these representations are being gradually modified and revised as the individual is exposed to new environments, it seems possible that representations of those early experiences may be reorganized substantially over time.

These kinds of issues are especially relevant in light of recent debates concerning the stability of attachment patterns. Some scholars have noted that attachment security is not particularly stable over time, with estimates hovering between .20 to .60 across various time intervals (see Baldwin & Fehr, 1995; Fraley, 2002). According to critics, if attachment security is not stable over brief periods of time, it seems unlikely that working models constructed in early childhood will continue to persist into adulthood. The objective of the simulations reported in this section is to explore issues concerning continuity and change using a simple connectionist network. One of the themes of this section is that patterns of continuity and change exhibited by a connectionist network depend crucially upon the network's learning history. In other words, there is nothing about connectionist networks per se that make representations stable or unstable; instead, it is the way in which the network acquires knowledge that determines its ability to retain that knowledge over extended periods of time.

The same network that was used previously was used for these simulations (see Appendix for details). For each simulation, a random 40-unit pattern (consisting of +1's and −1's) was generated and designated as the *initial pattern*. This pattern was designed to represent the presence or absence of features (e.g., warm, caring, cold, distant) that could be used to characterize the individual's early caregiving environment. Next, a series of patterns were generated that were gradual distortions of the initial pattern. These patterns

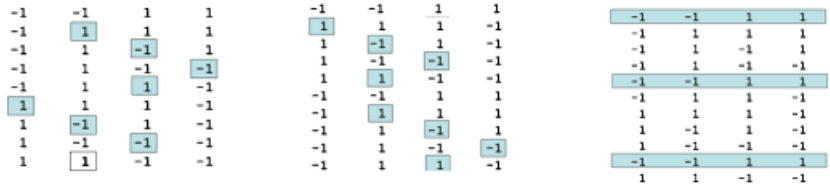

Figure 4

Examples of patterns learned by the network in the three simulations. Left panel: Each pattern was created by flipping sign of a random element from the immediately preceding pattern. Center panel: The same patterns were used as were used in the first simulation, but the order in which they were learned was random. Right panel: Patterns evolved gradually over time using the same rules as in the first simulation, but every once in a while the initial pattern was fully represented.

were created in an iterative fashion by randomly flipping the sign on one unit from the previous pattern. Thus, the second pattern was based on the first pattern, with one unit, selected at random, flipped (see the left-most panel of Figure 4). The third pattern was based on the second pattern, with the value of one unit flipped, and so on. Thus, each subsequent pattern was highly similar to the one prior to it but became increasingly dissimilar to the initial pattern over time. This series of patterns was designed to represent a gradually evolving sequence of caregiving "environments" such that the final environment was uncorrelated with the one the network initially experienced. One hundred patterns were created using this iterative technique.

At the beginning of each learning trial, the network was exposed to the target pattern. The activation was allowed to flow through the network for 40 cycles (enough time to allow the network to converge on a stable pattern of activation), after which the weights were updated according to the delta rule. The network learned the initial pattern first, followed by each of the 99 remaining graded patterns. The graded patterns were presented in the order of decreasing similarity to the initial pattern. Thus, as learning trials continued, the network was exposed to patterns that overlapped less and less with the one initially seen. The overlap between any two temporally adjacent patterns, however, was high.

Because there is an element of randomness in both the way the patterns are generated and the way in which the network behaves,

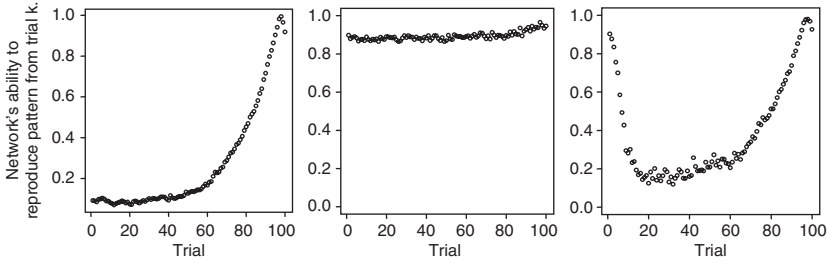

Figure 5
The network's ability, after the learning trials are complete, to reproduce each of the 100 patterns it learned in three conditions. In the left-most panel, the patterns the network learned were gradually evolving over time. In the central panel, the patterns the network learned were diverse and unstructured. In the right-most panel, the patterns the network learned were gradually evolving, with the exception of periodical appearances of the initial pattern.

the results reported here are based on averaging across the output of 50 distinct simulations. To explore the network's ability to represent each pattern, each pattern was re-presented at the end of the final set of learning trials (i.e., once the learning phase was over and the network had been exposed to each of the 100 patterns). The network's ability to represent the pattern was quantified as the correlation between the test pattern and the output of the network. The left-hand panel of Figure 5 illustrates how these correlations changed as a function of the order in which the patterns were learned. The first important feature to note about these data is that there is a strong recency effect, such that the network is best able to reproduce patterns that it experienced most recently. The second important finding is that not only is the model less adept at reproducing patterns it learned early in the developmental process, it is actually unable to produce the earliest patterns any better than chance. This result suggests that the way in which early experiences are represented in the weight matrix is gradually "overwritten" with time.

Notice also the way in which the right-most lip of the curve folds. Although there is strong evidence of a recency effect when the network's performance is evaluated against all of the patterns it experienced over the course of its development, performance is maximized for patterns it learned a few trials before the end of the simulation. This occurs because the network's weights are essentially constructing

a moving average of the patterns it is seeing. Thus, the pattern from, say, $k - 3$ trials, is much more prototypical of its recent experiences than the most recent pattern or a pattern $k - 6$ trials back in time.

In summary, these findings indicate that it is not possible for a connectionist network to retain accurate representations of early experiences if the caregiving environment gradually evolves over time. In the process of readjusting itself to learn new environmental patterns, the network loses its ability to represent the initial pattern accurately.

It is important to note that the network's performance is constrained by its learning history. In other words, this simulation assumed that, over the course of development, peoples' caregiving environments gradually evolve. The change is slow, making it unlikely that a person will notice the change as it occurs (in much the same way that people do not notice that their hair is growing until they look back at photos from the recent past). Whether peoples' caregiving environments evolve in this fashion, however, is unknown and worthy of debate. It is also possible that, while there is an element of randomness and change in our caregiving environments, there is an enduring degree of stability in those environments as well.

Under what conditions should the network be able to retain a representation of the initial pattern? In principle, the network should be able to represent all of the patterns accurately (initial and graded) if they are linearly predictable and separable (see Rumelhart & McClelland, 1986). By presenting the patterns in an ordered sequence, as was done previously, it is possible that the network was "forced" to adjust its weights in a way that did not allow it to establish stable representations of any one pattern in particular—including the initial pattern. If the same patterns were to be presented in a random order within a single set of learning trials, however, it is possible that the network would be able to represent each pattern adequately.

To test this hypothesis, 50 simulations were conducted in the same manner described previously. The only exception was that, instead of each pattern being presented sequentially over a set of learning trials, each pattern had an equal probability of being presented within a set of 100 learning trials (see the central panel of Figure 4 for an example). Conceptually, this situation is equivalent to an individual experiencing a diverse array of relational contexts within the same developmental window rather than navigating a developmental course that gradually evolves over time. It is important to note that the network was exposed to the *same* patterns that it learned in

the previous simulation; only the *order* in which the patterns were presented to the network was different.

When the network was exposed to a distributed learning history, it was able to reproduce any pattern, including the ones it experienced early in learning, with considerable accuracy (see the central panel of Figure 5). There was also evidence of a slight recency effect, such that the more recently learned patterns were recalled with the most precision, but this effect was not particularly pronounced. In summary, when the network was exposed to environmental patterns in a manner such that those patterns were not gradually evolving (but were just as variable), the network was able to preserve its knowledge of earlier patterns quite well. In other words, the network extracted a stable prototype that was retained in the mental system, despite the network's need to adapt to new contexts.

In the final simulation, I varied the learning environment such that, as in the first simulation, the patterns that the network experienced were gradual variations on one another. However, at randomly selected points in time, the network was exposed to the original, prototypical pattern (see the right-hand panel of Figure 4 for an example). Thus, although the network was experiencing a gradually evolving caregiving environment, the early pattern would periodically resurface, as may be the case when a specific experience is relatively salient to an individual, such as a particularly fond or tragic memory.

The performance of the network was quite different from that observed previously (see the right-hand panel of Figure 5). Notice that, as was the case in the first simulation, the network exhibited a recency effect such that patterns it had experienced more recently were the ones that the network was best able to reproduce. Importantly, however, the network also exhibited a primacy effect, such that its ability to reinstate patterns experienced early in development was intact.

Summary. The previous simulations have important implications for theories of continuity and change in attachment. First, they suggest that there is nothing about connectionist networks per se that enables them to exhibit stability or instability in the representations they construct. The key factor that influences the stability of a representation is the network's learning history. If the initial environment changes gradually over time, the network's ability to represent the early environment gradually fades, favoring memories for more recent experiences (see Lewis, 1997). If the environment evolves to a

similar degree, but in a nonordered fashion (i.e., if the network experiences a variety of different caregiving environments across its development), representations of early experiences continue to exist and are readily reactivated when the context is appropriate. According to the third simulation, representations of early experiences have a privileged status in the memory system when those experiences are recurrent—even if rare—across development. For example, if the primary caregiver's behavior gradually changes over time, but the prototypical pattern periodically reemerges, the network is capable of maintaining a representation of that prototype, despite the gradual evolution of the caregiver's behavior and even when the caregiver has not behaved in that fashion recently. In short, the network is capable of developing representations of the same person's behavior across multiple contexts (i.e., what is expected based on recent experiences and what is prototypical or based on early experiences).

DISCUSSION

The primary objective of this article was to consider some important issues in the study of adult attachment from the perspective of connectionist models of memory. One of the major issues concerns the organization of working models or, more specifically, how global and relationship-specific models are structured, whether they can jointly influence behavior, and, indeed, whether global models have any ontological basis. Previous research has shown that people exhibit different attachment patterns in different relational contexts, raising questions about whether it makes sense to conceptualize and measure attachment security in a global or trait-like fashion. One solution to this problem has been to acknowledge that both global and relationship-specific attachment representations exist and that both play some role in shaping thoughts, feelings, and behavior in close relationships. This hypothesis, however, has been a difficult one to test because credible alternative explanations exist for how it is that measures of security measured across different contexts can become correlated in the fashion entailed by a hierarchical model. Moreover, the hierarchical assumption is unsatisfactory on theoretical grounds because the hierarchical relations among objects are built into the model a priori, therefore making it a less powerful explanation for the associations between measures of security assessed in different relational contexts.

One of the advantages of a connectionist framework is that it offers an account for how relationship-specific models develop and, importantly, how global models can be extracted from those relationship-specific experiences. In a connectionist model, the global representation emerges naturally from the same learning process that enables relationship-specific patterns to be learned. This global model is readily activated and applied to novel contexts in which there is partial overlap between the characteristics of the new interaction partner and the global representation. Another advantage of the connectionist framework is that it offers a different metaphor than the hierarchical one for conceptualizing the organization of working models. In a connectionist network, working models for various relational contexts are not discrete things; instead, knowledge for different relational contexts is distributed across the network. This latter point has important implications for understanding how existing attachment representations come to shape future experiences. If an individual is forging a new relationship with someone who exhibits specific qualities (e.g., a reluctance to self-disclose), any existing representation that involves this feature will become active and will be brought to bear in making sense of the new person. It is less relevant within a connectionist framework whether the new person falls into a similar social category (e.g., parent, partner, peer) than whether there is overlap at the level of psychological features.

A connectionist framework also offers some useful insights into continuity and change in attachment representations. Specifically, it suggests that the way in which a person's developmental history is structured has important implications for the stability of working models. If the individual is attached to someone whose behavior gradually evolves over time, the representations of that person constructed early in development will gradually change such that they no longer resemble what they were once. In contrast, if a person's various relational experiences are less structured, the individual will develop multiple representations of those experiences, each of which will be highly stable over time. Most importantly, even when aspects of a significant relationship gradually evolve over time, if the core elements are reinstated from time to time, early representations will exhibit a high degree of stability.

These simulation findings are important because they suggest a mechanism by which early attachment experiences can be retained in a mental system, even when they have been dormant for most of the

time. One of the interesting claims made in the attachment literature is that, even when they are not being used, representations of early experiences exist in the mind and can be reactivated under appropriate situations. Sroufe, Egeland, and Kreutzer (1990), for example, argued that "earlier patterns may again become manifest in certain contexts, in the face of further environmental change, or in the face of certain critical developmental issues. While perhaps latent, and perhaps never even to become manifest again in some cases, the earlier pattern is not gone" (p. 1364). The simulations reported here suggest that this is indeed possible, so long as the relational context in which the pattern was initially acquired is not gradually evolving over time.

To the personality psychologist, many of the issues that have been discussed here should seem familiar. Historically, personality psychologists have assumed that people hold basic personality traits that influence behavior across a variety of contexts. Yet, empirically, behavior is not highly consistent from one context to the next. One of the proposed solutions to the problem is to view personality as a system of "if . . . then" contingencies for behavior (i.e., the CAPS model; Mischel & Shoda, 1995). From this perspective, a person's behavior may not be consistent across contexts because he or she has adapted a certain set of skills and norms for each relational context. Within that context, however, the person may behave in a consistent way. This approach is similar to the one endorsed here, with one critical exception. Namely, the connectionist perspective suggests that the mind can construct unique "rules" for behavior in different situations but that a more global set of rules is abstracted and used as well. In other words, because the same network can develop and use relationship-specific and global representations, behavior in any one context is likely to be a product of both. The CAPS model has the potential to capture the idiosyncrasies of a person's behavior across different contexts but doesn't provide a means to capture what is common across contexts. A connectionist approach has the potential to accommodate both.

REFERENCES

Anderson, J. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Baldwin, M. W., & Fehr, B. (1995). On the instability of attachment style ratings. *Personal Relationships*, **2**, 247–261.

Baldwin, M. W., Keelan, J. P. R., Fehr, B., Enns, V., & Koh-Rangarajoo, E. (1996). Social cognitive conceptualization of attachment working models: Availability and accessibility effects. *Journal of Personality and Social Psychology*, **71**, 94–104.

Cervone, D. (1997). Social-cognitive mechanisms and personality coherence: Self-knowledge, situational beliefs, and cross-situational coherence in perceived self-efficacy. *Psychological Science*, **8**, 43–50.

Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, MA: MIT Press.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal memory*, **8**, 240–247.

Collins, N., & Read, S. (1994). Cognitive representations of attachment: The structure and function of working models. In K. Bartholomew & D. Perlman (Eds.), *Attachment processes in adulthood: Advances in personal relationships* (Vol. 5, pp. 53–90). London: Kingsley.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness*. Cambridge, MA: MIT Press.

Fox, N., Kimmerly, N. L., & Schafer, W. D. (1991). Attachment to mother/attachment to father: A meta-analysis. *Child Development*, **62**, 210–225.

Fraley, R. C. (2002). Attachment stability from infancy to adulthood: Meta-analysis and dynamic modeling of developmental mechanisms. *Personality and Social Psychology Review*, **6**, 123–151.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, **43**, 23–34.

Klohnen, E. V., Weller, J. A., Luo, S., & Choe, M. (2005). Organization and predictive power of general and relationship-specific attachment models: One for all, and all for one? *Personality and Social Psychology Bulletin*, **31**, 1665–1682.

Kobak, R. (1994). Adult attachment: A personality or relationship construct? *Psychological Inquiry*, **5**, 42–44.

La Guardia, J. G., Ryan, R. M., Couchman, C. E., & Deci, E. L. (2000). Within-person variation in security of attachment: A self-determination theory perspective on attachment, need fulfillment, and well-being. *Journal of Personality and Social Psychology*, **79**, 367–384.

Lewis, M. (1994). Does attachment imply a relationship or multiple relationships? *Psychological Inquiry*, **5**, 47–51.

Lewis, M. (1997). *Altering fate: Why the past does not predict the future*. New York: Guilford Press.

Macdonald, C., & Macdonald, G. (Ed.) (1995). *Connectionism: Debates on psychological explanation*. Oxford, UK: Blackwell.

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.

Overall, N., Fletcher, G. J. O., & Friesen, M. (2003). Mapping the intimate relationship mind: Comparisons between three models of attachment representations. *Personality and Social Psychology Bulletin*, **29**, 1079–1094.

Pierce, T., & Lydon, J. (2001). Global and specific relational models in the experience of social interactions. *Journal of Personality and Social Psychology*, **80** (4), 613–631.

Read, S. J., & Miller, L. C. (1998). *Connectionist models of social reasoning and social behavior*. Mahwah, NJ: Erlbaum.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–364). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1986). PDP models and general issues in cognitive science. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed processing: explorations in the microstructure of cognition* (Vol. 1, pp. 110–146). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vols. 1 & 2). Cambridge, MA: MIT Press.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential though processes in PDP models. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 7–57). Cambridge, MA: MIT Press.

Queller, S. (2002). Stereotype change in a recurrent network. *Personality and Social Psychology Review*, **6**, 295–303.

Schultz, T. R. (2003). *Computational developmental psychology*. Cambridge, MA: MIT Press.

Shoda, Y., LeeTiernan, S., & Mischel, W. (2002). Personality as a dynamical system: Emergence of stability and constancy from intra- and inter-personal interactions. *Personality and Social Psychology Review*, **6**, 316–325.

Smith, E. R. (1996). What do connectionism and social psychology offer each other? *Journal of Personality and Social Psychology*, **70**, 893–912.

Sroufe, L. A., Egeland, B., & Kreutzer, T. (1990). The fate of early experience following developmental change: Longitudinal approaches to individual adaptation in childhood. *Child Development*, **61**, 1363–1373.

Stone, G. O. (1986). An analysis of the delta rule and learning of statistical associations. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing* (Vol. 1, pp. 444–459). Cambridge, MA: MIT Press.

APPENDIX

The model employed in these simulations was a synchronous, auto-associator model with 40 units, similar to that used by Rumelhart, Smolensky, McClelland, and Hinton (1986). The activation of each unit during processing cycle k was a function of its activation during cycle $k-1$ and the net input it received from other units. The net input for unit i was defined as the sum of the activation values of all

units weighted by the strength of their connection to unit *i*:

$$net_i = \Sigma a_j w_{ij} \tag{1}$$

During each cycle, activation values were updated according to a simple nonlinear function:

$$\begin{aligned} \Delta a_j &= E \times net_i(1 - a_i) - Da_i \quad [\text{if } net_i > 0] \\ \Delta a_j &= E \times net_i(a_i - (-1)) - Da_i \quad [\text{if } net_i \leq 0] \end{aligned} \tag{2}$$

where E and D are global parameters corresponding to excitation and decay rates, respectively (these values were set to .15 in the present simulations). When the net activation reaching unit j was positive, the unit's activation increased in a manner proportional to the input and the difference between the unit's activation and its activation ceiling (+1). When the net activation reaching unit j was negative, the unit's activation decreased in a manner proportional to the net input and the difference between the unit's activation and its activation floor (-1). A decay factor was also incorporated into the model that pushed the activation of unit j toward zero as the absolute magnitude of unit j's activation increased.

During the learning phases of the various simulations to follow, the network was exposed to a stimulus pattern (i.e., a sequence of 1's and -1's). Once the activation values of the units in the network had settled into a stable pattern, the weights among units were updated according to the *delta rule*. According to the delta rule (see Stone, 1986), the change in a weight between units j and i is proportional to the difference between the activation of unit j and its "desired" level of activation (given the input):

$$\Delta w_{ij} = \eta \delta_i a_j \tag{3}$$

In this equation, η is the learning rate, which was set to .01 in the present simulations. δ_i is the discrepancy between the "desired" output of unit i (the external input to the unit) and the input to unit i from unit j. When the difference is positive, the unit j is not sending unit i enough activation. Thus, the weight between these two units is increased. When the difference is negative, unit j is sending too much activation to unit i. The weight between these two units is adjusted accordingly.