Personality and Social Psychology Review
1997, Vol. 1, No. 1, 26–53

Connectionism, Parallel Constraint Satisfaction Processes, and Gestalt Principles: (Re)Introducing Cognitive Dynamics to Social Psychology

Stephen J. Read, Eric J. Vanman, and Lynn C. Miller
Department of Psychology
University of Southern California, Los Angeles

We argue that recent work in connectionist modeling, in particular the parallel constraint satisfaction processes that are central to many of these models, has great importance for understanding issues of both historical and current concern for social psychologists. We first provide a brief description of connectionist modeling, with particular emphasis on parallel constraint satisfaction processes. Second, we examine the tremendous similarities between parallel constraint satisfaction processes and the Gestalt principles that were the foundation for much of modern social psychology. We propose that parallel constraint satisfaction processes provide a computational implementation of the principles of Gestalt psychology that were central to the work of such seminal social psychologists as Asch, Festinger, Heider, and Lewin. Third, we then describe how parallel constraint satisfaction processes have been applied to three areas that were key to the beginnings of modern social psychology and remain central today: impression formation and causal reasoning, cognitive consistency (balance and cognitive dissonance), and goal-directed behavior. We conclude by discussing implications of parallel constraint satisfaction principles for a number of broader issues in social psychology, such as the dynamics of social thought and the integration of social information within the narrow time frame of social interaction.

Connectionism, neural networks, and parallel distributed processing models are among the fastest growing research areas in the study of the mind. But many social psychologists seem unsure of their relevance. The purpose of this article is to demonstrate that this burgeoning new area is of great importance to social psychology. First, although the tools that they offer are new, many key insights of these models are neither new nor foreign to social psychology. There are remarkable parallels between key aspects of connectionist models and Gestalt principles—those principles that guided many of the founders of modern social psychology. Second, these new connectionist modeling tools and accompanying research can dramatically enhance our ability to examine the dynamic and wholistic aspects of social phenomena.

Let us begin by focusing on the historical legacy of Gestalt psychology for social psychology. What was its role, and why did its influence wane? In so doing, we briefly discuss why concepts such as connectionism, parallel constraint satisfaction process, and neural networks are relevant to theory and research in social psychology. Then we turn to a more detailed treatment of parallel constraint satisfaction processes and connectionist models, and follow this with an examination of the commonalties between the various Gestalt concepts and parallel constraint satisfaction processes. Following that, we discuss the applications of parallel constraint satisfaction processes to three key areas in social psychology: impression formation and causal reasoning, cognitive consistency (balance theory and cognitive dissonance), and goal-directed behavior. One thick thread that winds throughout our argument is that parallel constraint satisfaction processes provide an integrative framework for thinking about a range of processes and phenomena that have been treated separately by social psychologists.

The Historical Legacy of Gestalt Psychology

Gestalt psychology was the foundation upon which much of modern social psychology was built. Consider the work of Asch, Lewin, Heider, and Festinger: Gestalt principles were central to each. Asch's impression formation work (Asch, 1946; Asch & Zukier, 1984) argued that the processing of social stimuli was wholistic.

Our thanks to David Walsh, Eliot Smith, Yoshi Kashima, Bernadette Park, Paul Robert Appleby, Kim Guster, Neal Lalwani, Joe Mancuso, George Montoya, Sadina Rothspan, Darren Urada, and Lynn Urban for their comments on earlier versions of this article.

Requests for reprints should be sent to Stephen J. Read, Department of Psychology, University of Southern California, Los Angeles, CA 90089–1061. E-mail: read@almaak.usc.edu.

Heider (1958) relied heavily on Gestalt principles of structure and organization (e.g., good form and equilibrium) in his classic work on attribution theory and causal perception. Lewin's (1947a, 1947b) model of group process and other early work on groups viewed group members' social interactions in terms of interacting fields of forces; group properties (e.g., group cohesiveness) were the result of interacting force fields (Festinger, 1950). Similarly, Lewin (1935), in articulating central issues in the dynamics of goal-directed behavior, proposed that person–situation interactions could be treated in terms of interacting force fields. And consider Heider's (1946, 1958) balance theory and Festinger's (1957) theory of cognitive dissonance. In discussing belief systems and their consistency, both relied heavily on Gestalt ideas of structural dynamics and Gestalt principles of organization and structure, such as good form and equilibrium. Finally, Gestalt principles provided the theoretical foundation for Krech and Crutchfield's (1948) classic textbook on social psychology.

Yet, despite their historical and theoretical importance, Gestalt principles are largely absent from current social psychological theorizing. Why? Gestalt principles stressed wholistic processing and interactions among fields of psychological forces. Such concepts may have struck many as too metaphoric and mystical. A second difficulty was simply grappling with the overwhelming richness and complexity of social interaction implied by basic Gestalt principles. For example, the construction of social meaning is the result of multiple, mutually interacting influences among numerous pieces of information (Asch, 1946; Heider, 1958). Isolated social behaviors rarely have a clear meaning separate from the context in which they occur. They can only be understood when integrated with a range of other information, such as other behaviors, the situation, the individual's personality, and so forth. Furthermore, individuals in social interaction must integrate large amounts of information in a short time, while concurrently planning, enacting, and monitoring their own behavior. Initial attempts to address this complexity can be found in Gestalt theorizing. Gestalt processes provided a mechanism by which multiple interacting pieces of information could be integrated within the narrow time frame of social interaction. However, capturing such dynamics—and studying such processes—may have seemed beyond the reach of the empirical and theoretical tools of the day. It is only recently, after a long hiatus, that current theories have begun to re-address these issues. Today, these obstacles may well be surmountable.

Recent work in connectionism, neural networks, and parallel distributed processing models suggests that seemingly metaphorical Gestalt processes can be given a concrete implementation (Holyoak & Spellman, 1993; Spellman & Holyoak, 1992). Work in this area seeks to model thought as occurring in networks of simple neuron-like units, wherein processing occurs by the passage of activation, in parallel, among those nodes. In one class of models this processing takes the form of a parallel constraint satisfaction process that simultaneously solves for a set of constraints among a set of concepts. In this article, we focus on this class of models because of their clear parallels to Gestalt principles and many of the issues addressed by early social psychologists.

Thus, the set of associated cognitive elements discussed by Gestalt theorists could be represented as a network consisting of nodes (representing concepts) and the links among those nodes, whereas Gestalt processes can be given a computational implementation as a parallel constraint satisfaction process applied to this network of nodes and the links among them. Because parallel constraint satisfaction processes have a concrete, computational implementation and have an increasingly well understood mathematical foundation (e.g., Amit, 1989; Hertz, Krogh, & Palmer, 1991), they are not subject to the claims of vagueness or abstractness that proved so damaging to Gestalt theory.

Moreover, given that Gestalt psychologists argued that psychological processing involves interactions among fields of forces, it is interesting that mathematicians and physicists (e.g., Amit, 1989; Hertz et al., 1991; Hopfield, 1982, 1984) are finding fruitful parallels between parallel constraint satisfaction models and models of interacting magnetic and electrical fields. As a result, they have been able to bring to bear a large body of existing results and mathematical tools from physics and have used them to greatly expand our understanding of the behavior and capabilities of neural networks.

Thus, this work on parallel constraint satisfaction processes suggests it may be time for social psychologists to use these emerging tools to push Gestalt ideas beyond the insights of Lewin, Asch, Heider, and Festinger. We turn now to a general description of connectionist models, with particular emphasis on parallel constraint satisfaction processes.

Connectionist Models

Connectionist modeling (e.g., Hertz et al., 1991; McClelland & Rumelhart, 1986; Miikkulainen, 1993; Murre, 1992; Rumelhart & McClelland, 1986a) treats the processing involved in perceptual and cognitive tasks in terms of the passage of activation, in parallel, among neuron-like units. The most important components of these models are (a) simple processing units or nodes—which sum the incoming activation following a specified equation, and then send the resulting activation to the nodes to which they are connected; (b) equations that determine the activation of each node at each point in time, based on the incoming activation from other nodes, previous activation, and the decay rate; (c) weighted connections between the nodes, where the weights affect how activation is spread; and (d) a learning rule that specifies how the weights change in response to experience (Bechtel & Abrahamsen, 1991). Processing in a connectionist model proceeds solely by the spread of activation among nodes, with the pattern of connections affecting how activation spreads. There is no higher order executive or control process. Moreover, knowledge in a connectionist model is represented entirely in the pattern of weights among nodes.

The possible activations of the nodes may be discrete—typically binary values such as 0 or 1—or they may vary continuously, such as between –1 and 1. The resulting activation may be a linear function of incoming activation, or a nonlinear function—such as a sigmoid or S-shaped rule—where the activation asymptotes at some minimum or maximum value. (As we discuss later, models with nonlinear functions have major advantages over linear models.)

One important difference among connectionist networks is whether there are feedback relations among the nodes. In feed-forward networks, units have unidirectional connections, with no feedback relations. The network is organized in layers, with inputs fed into the input layer and outputs generated at the top layer as a result of a single forward sweep of activation. The simplest such network has two layers, an input and an output layer, although more complicated networks may have intervening or *hidden* layers (so-called because they have no direct connections to the environment.). Networks with hidden layers, such as the well-known back propagation network, have greater computational power. A prototypical example of a feed-forward network is the pattern associator, in which the system learns an arbitrary association between an input represented as a pattern of activation on the input layer and a pattern represented on the output layer. Such networks can learn to categorize objects or assign names to objects.

By contrast, in interactive, or feedback networks, at least some connections are bidirectional—resulting in feedback relations—and processing occurs dynamically across a large number of cycles. Nodes in these networks have a minimum and maximum possible activation (typically ranging from 0 to 1, or from –1 to 1). The activation of the nodes is updated many times as the activation of the units moves towards asymptote, and as the system works toward settling into a solution to a particular input. In contrast, in feed-forward networks, activation is updated only once.

Because of the feedback relations, interactive or feedback networks are dynamic systems whose behavior evolves over time. As a result, they have interesting and useful properties that are not characteristic of feed-forward networks. One of the most useful properties of such networks is that they function as parallel constraint satisfaction systems, acting to satisfy multiple simultaneous constraints among elements in a network. In the current article we focus on such feedback networks and their ability to perform parallel satisfaction of multiple constraints. We do so because we believe that parallel constraint satisfaction processes have tremendous implications for addressing a number of classic and contemporary issues in social psychology.

Parallel Constraint Satisfaction Processes

In a feedback or parallel constraint satisfaction network, activation passes around symmetrically connected nodes until the activation of all the nodes asymptotes or "relaxes" into a state that satisfies the constraints among the nodes. This process allows for the integration of a number of different sources of information in parallel.

Constraint satisfaction networks have interesting dynamical properties with the activation of the nodes evolving over time in quite interesting and useful ways. As we discuss later, this gives them a capacitiy for content addressable memory as well as the ability to complete patterns, construct schemas, and solve global optimization problems. Because these capabilities depend on the feedback nature of the network, they are not found in nondynamic or feed-forward networks such as the well-known back propagation architecture.

The nodes in these networks represent hypotheses about the presence or absence of various features; the hypotheses can vary from microfeatures, such as color or lines, to concepts or entire propositions, such as traits or a behavior. Links among nodes represent the extent to which the hypotheses are consistent with and support one another or are inconsistent with and contradict one another. Thus, links can be thought of as representing constraints among the hypotheses. Hypotheses with positive links are mutually supportive; if one node is activated, it will try to activate the other. In contrast, hypotheses with negative links are contradictory, or compete; therefore, if one node is positively activated, it will try to deactivate the other. Weights on the links can vary, indicating the strength of the constraint between the nodes.[1]

[1]Neural network models are best viewed as neurally inspired, as McClelland and Rumelhart (1986) noted, rather than as neurally plausible. Few, if any, individuals who model cognitive processes try to implement their model at the level where nodes correspond to individual neurons in the brain. For example, the feature representations that are used, say, in a model of word recognition are many levels up from raw perceptions. A further salient difference between actual neural networks and neural network models is that—as many researchers have noted—real neurons signal information by their frequency of firing, whereas the nodes in the typical neural network model signal information by their level of activation. Thus, although strongly inspired by actual neural networks, very few current neural network models are strictly neurally plausible.

What the nodes and links represent depends on the theoretical assumptions of a specific model. For instance, in a model of vision the nodes might correspond to features such as lines with particular orientations, whereas in a model of discourse comprehension the nodes might correspond to actions. The links in a model of vision might correspond to spatial relations, whereas the links in a model of discourse comprehension might correspond to causal or inferential relations. Thus, in any specific model, parallel constraint satisfaction principles work hand in hand with theoretical assumptions about such things as representation that are specific to the particular phenomena being addressed. As J. R. Anderson (1978) and others have noted, any cognitive model consists of a representation–process pair, where the theorist must make assumptions both about how information is represented and how that information is processed. Parallel constraint satisfaction models are no different.

The set of constraints among the nodes and their eventual resolution is evaluated by spreading activation among the nodes in parallel. Because the nodes are symmetrically connected and thus have feedback relations, the activation of the nodes evolves over time. As node *a* is sending activation to node *b* it is also receiving activation from node *b,* as well as from other nodes. Thus, immediately after node *a* sends activation, its current state is likely to have changed and the amount of activation it can send has also changed. Thus, the activations of the nodes in such a network are continually evolving. However, considerable research and analysis has shown that, except in rare cases, the activation of the nodes eventually reaches asymptotic values and the network stabilizes and stops changing.

In such models the resulting activation of the node is a nonlinear function of the sum of the inputs, where the form is sigmoid-shaped as in Figure 1. One possible nonlinear activation function that has been frequently used in the following or a slightly modified form (e.g., McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986a; Thagard, 1989) is:

$$a_j(t+1) = a_j(t)(1-d) + enet_j(max\text{-}a_j(t)) + inet_j(a_j(t) - min) \quad (1)$$

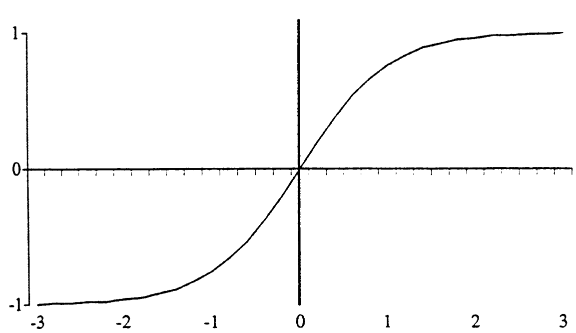

Figure 1. Graph of the general form of a nonlinear, sigmoid-shaped function.

where $a_j(t+1)$ is the new activation of the node and $a_j(t)$ is the activation of the node on the previous time step; d is a decay parameter; $enet_j$ is the net excitatory input; and $inet_j$ is the net inhibitory input. Here $enet_j$ is equal to $\sum_i w_{ji}a_i(t)$ for $w_{ji}a_i(t) > 0$, where w_{ji} is the weight from node i to j and $inet_j$ is equal to $\sum_i w_{ji}a_i(t)$ for $w_{ji}a_i(t) < 0$. Also, *min* is the minimum activation value possible, –1.0, and *max* is the maximum activation value, 1.0. Note that this activation rule is nonlinear because the possible amount of change in activation of a node is proportional to the difference between its possible maximum or minimum and its current activation. As the current activation approaches the maximum or minimum, this difference decreases and thus the amount of change possible also decreases, resulting in an asymptotic approach to the maximum or minimum. However, because the asymptotic value is a function of both the incoming activation and the decay, the final asymptotic value typically does not reach the max or min activation value. Updating stops when the activation of all nodes reaches asymptote. As can be seen from the equation, the amount of activation sent to a node is a function of the number of nodes connected to it, the strength of the links, whether the link is positive or negative, and the activation of the connected nodes.

One implication of this nonlinear form is that the impact of an individual input strongly depends on the other input activations. When the other inputs are fairly weak, then a strong input can have a major impact because the activation function will be in the strongly accelerating part of the curve. However, if all the inputs are strong, then the input will have a much smaller effect because the activation function will be near the asymptotic part of the curve. The nonlinear form of the activation function plays an important role in the behavior of most connectionist models and is important for understanding the similarities between Gestalt principles and parallel constraint satisfaction processes.

Thus, activation is spread in parallel among all the nodes until the activation of each node asymptotes. When activation spreads through such a network, nodes with positive links will tend to activate each other and nodes with negative links will inhibit each other. Because the activation of a node is a result of all of its positive and negative links to other nodes, the final activation of the node can be thought of as a solution to all the constraints represented by the links. Moreover, because activation is spread in parallel among all the connected nodes, this process results in a global solution to the constraints among the entire set of nodes.

Interestingly, Hopfield (1982, 1984; see also Amit, 1989; Hertz et al., 1991; Rumelhart, Smolensky, McClelland, & Hinton, 1986), who relied on extensive work in physics on thermodynamic systems, showed that such a system—with symmetric connections—can be treated as having energy, where the energy of the system is the sum of the product of the activation of all possible pairs of nodes times the weight between them: that is, $w_{ij} * a_i * a_j$. Specifically, the "energy" of the system is $E = -\sum_i \sum_j w_{ij} a_i(t) a_j(t)$. This equation specifies that the energy of the system will decrease when the sign of the product of the activations is consistent with the sign of the weight between them, but will increase when the sign of the product of the activations differs from the sign of the weight between them. That is, if the product of the activation of two nodes is consistent with the constraint between them, energy decreases; whereas, if the activation of two nodes is inconsistent with the constraint between them, energy increases. Thus, this energy function essentially measures the extent to which the pattern of activations of the nodes is consistent with the relations between them. Hopfield (1982, 1984) demonstrated that neural network systems of this form act so as to minimize the energy function, essentially minimizing the energy of the system.

Further, Hopfield (1982, 1984) noted that the energy of the system can be plotted in an N-dimensional space, resulting in an energy surface that represents the various possible energy states of the system. This idea is quite powerful in informing intuitions about the behavior of such systems. An energy surface is a multidimensional representation in which the possible range of activation of each node in the network defines one dimension in the representation, and the shape of the energy surface is defined by the amount of energy (or degree of organization) of the system at each of the possible combinations of activations of all the nodes in the network (see Amit, 1989; Hertz et al., 1991; Rumelhart et al., 1986). Thus, the energy of a network with N nodes can be represented by an energy surface in an $N + 1$ dimensional space, where the activation of each of the N nodes defines a dimension and the $N + 1$ dimension represents the energy of the system. Figure 2 provides an example of what such an energy surface might look like for a simple network with two nodes and therefore three dimensions.

Given the idea of an energy surface, solving for the constraints can be viewed as a gradient descent process—moving toward a minimum (or valley) in an energy surface. Thus over time the system moves down a gradient or slope until a minimum is reached. A system that has settled or relaxed can be viewed as having reached a valley in the energy surface, representing a minimum state of energy. However, such minima are not guaranteed to be global minima of the entire system, but may instead be local. That is, the system can settle into or be "trapped" in energy states that are higher than the global minimum of the system.

Equivalently, minimizing the energy of the system can be thought of as moving from a state where fewer constraints are satisfied to one where more constraints are satisfied (Rumelhart et al., 1986). Essentially, the energy of the system corresponds to its degree of organization. High energy in the system corresponds to less organization, and low energy corresponds to greater organization. Thus, a parallel constraint satisfaction process can be viewed as attempting to find the maximal degree of organization consistent with the constraints imposed by the relations among the nodes. That is, the system is maximizing the goodness of fit of the network. However, because the system is not guaranteed to find the global minima, it will not necessarily find the state representing the maximum degree of organization.

Because lower energy essentially corresponds to higher organization, some researchers have removed the minus sign from the energy equation and treated the result as a measure of the goodness of fit of the network (McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986a). Smolensky (1986) developed a similar measure, which he refers to as the *harmony* of the system (see also Thagard, 1989).

The minima in such systems are called *attractors* and can be viewed as having a basin of attraction that corresponds to the valleys or wells in the energy surface. If we probe such a network with a pattern of activations that falls within the basin of attraction of a particular attractor (i.e., it falls on the slopes of the basin), then the network will evolve toward that attractor. This explains the name attractors: Attractors are states of the system that act as if they "attract" nearby states. The systems are called *attractor systems.*

Such systems naturally move toward a greater degree of organization or consistency. There is no need to postulate any kind of need or motive for consistency as the system naturally moves in that direction as a result of its internal dynamics.

To make this more concrete, consider an individual trying to arrive at a coherent set of beliefs about a particular issue. Suppose Tom has been spending a good deal of time studying the various arguments for and against abortion and is developing a coherent position on the issue. Presumably, there are at least two relatively coherent positions on abortion: *pro-life* and *pro-choice* (to use the proponents' preferred terms). If we view the arguments and their relationships as a network in which the nodes represent the individual arguments and the links the relations among the arguments, then each coherent position can be viewed as a valley or attractor in the energy surface that represents the possible states of the individual beliefs. Once Tom's beliefs have developed to the point that they fall in the basin of attraction of one of the attractors of the two coherent positions, then the state of Tom's beliefs should continue evolving toward the attractor—where the attractor represents a minimum degree of energy or

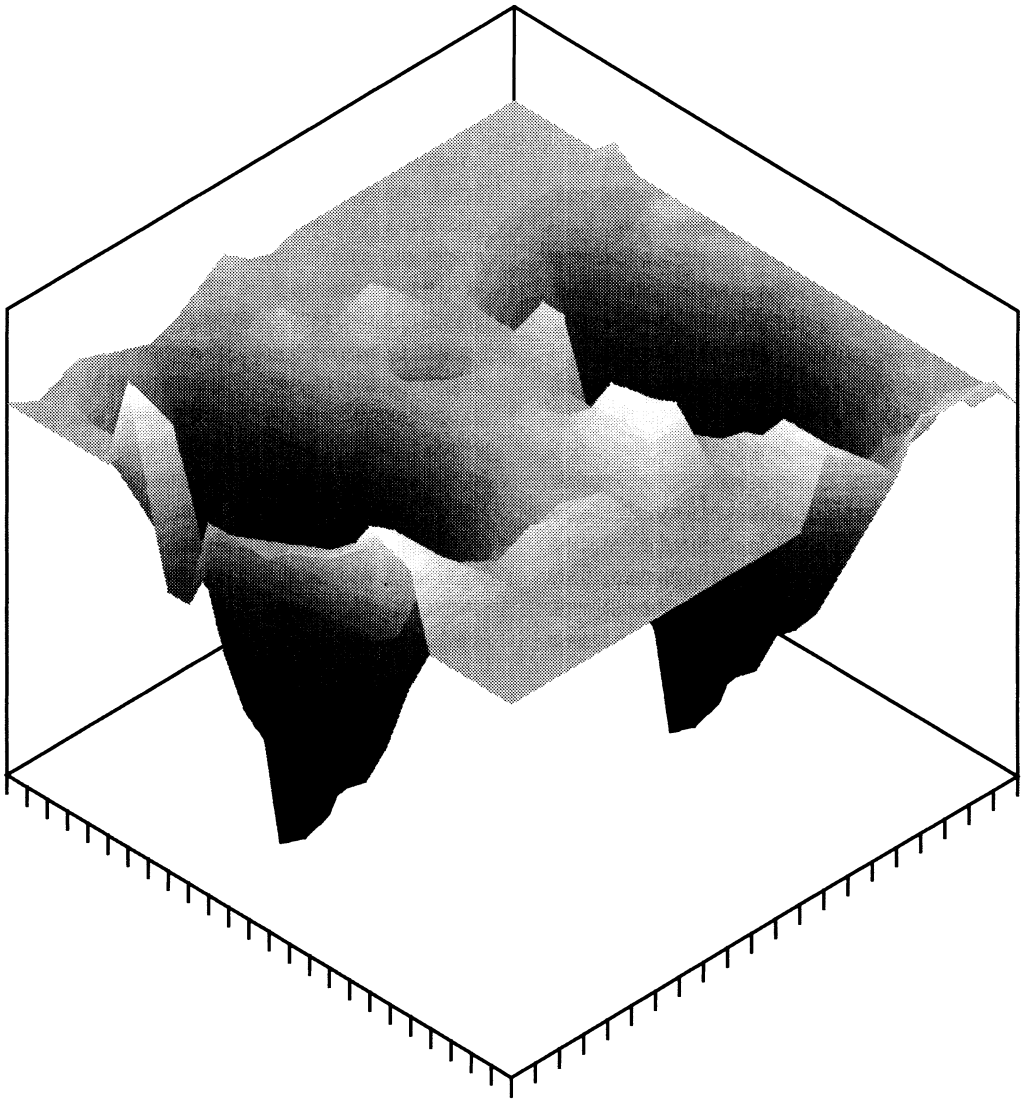

Figure 2. Example of a possible energy surface for a two-node network.

maximum degree of organization. Because the different attractors in such systems are separated by "hills," once Tom has arrived at the attractor corresponding to one position, it may be quite difficult to move to an alternative position, and far more difficult to move to the alternative attractor than it was before the system "settled." This points up the similarity between parallel constraint satisfaction processes and models of cognitive consistency, an issue to which we will return.

These kinds of systems can serve a variety of functions. First, as already mentioned, they can evaluate a set of constraints among a set of nodes. For example, Thagard (1989, 1992) used them to evaluate the goodness of causal explanations for such things as scientific phenomena and murder trials; Read and Marcus-Newhall (1993) used them to evaluate the goodness of social explanations; and Kunda and Thagard (1996) used them to model the integration of stereotypes with individuating information.

Another important aspect of these kinds of systems is that the minima or attractors can represent learned patterns of associations among features. Hopfield (1982, 1984) showed that learning rules that encode the patterns of associations or correlations among activations of different nodes in the network can be viewed as "digging" valleys or minima in the energy surface and that reaching a minima is equivalent to retrieving the pattern of activation of the nodes that corresponds to the position of the minima in a multidimensional space.

Thus, these systems can function as pattern completion devices. If the network has learned a particular pattern and has dug a corresponding attractor, and if the system is then given a partial pattern that places the system within the basin of attraction of the attractor, the system will evolve toward the attractor and fill in the rest of the pattern. For example, if given the features whiskers and purrs, the system should evolve toward an attractor with the remaining features and thus fill in the rest of the pattern for cat, including the name.

Because of this pattern completion ability, such systems can function as content addressable memories, categorize objects, and construct schemas. First, consider content addressable memories. As noted, learning a particular memory corresponds to learning the associations among the elements of the memory and constructing an attractor that corresponds to that pattern of associations. Thus, if a particular set of features—such as Bob's 30th birthday party—places us within the basin of attraction of that memory, the system will evolve toward the attractor and result in the retrieval of the other associated features of that memory.

Second, Rumelhart et al. (1986) showed how such systems can implement schemas. For example, they taught a system the associations among the objects, pieces of furniture, and appliances in a typical house and then showed that such a system seemed to have implicitly encoded the schemas for different rooms. Thus, when the system was given one or two appropriate cues, it could construct the schema for the appropriate room. For example, if the nodes corresponding to stove and walls were turned on, the remaining parts of a kitchen—such as refrigerator and sink—were also activated, whereas nodes for sofa or bed were not.

Third, such systems can be used to categorize, learning both exemplars and prototypes for a group. For example, if an individual is exposed to members of a category, such as members of an ethnic group or occupation, then attractors that correspond to those exemplars will be constructed. However, if the attractors for the exemplars are close enough together to overlap, the result is an attractor that corresponds to the prototype of the group. Thus, the extent to which categories are represented by exemplars or prototypes is partially a function of the similarity of exemplars to each other.

In such a system, if some distinguishing features of the group are presented, the system can evolve toward the attractor for that group and retrieve the group label. Or, if the group label is presented, it should then evolve toward the attractor and allow retrieval of prototypical features of the group. Moreover, if the system has attractors for both exemplars and prototypes, it is possible to retrieve both, depending on the similarity of the retrieval cues to the features that define the location of the attractor.

Localist, or Symbolic, Versus Distributed Models

Parallel constraint satisfaction processes can be applied to networks in which the nodes represent everything from low-level perceptual features to higher level constructs such as concepts or propositions. For example, in *localist,* or *symbolic,* connectionist models (e.g., Holyoak & Thagard, 1989; Thagard, 1989), nodes represent entire concepts or propositions, whereas in *distributed* models, nodes represent features—and higher level concepts are represented by patterns of activation distributed across those nodes. Distributed models have some important advantages, such as greater resistance to loss of information when a node is damaged—or when one is trying to model how the cognitive system moves from perception of features to higher level concepts, as in word recognition or categorization; however, they also have an important cost. They can be much harder to implement and to interpret than localist models. And there are certain kinds of problems for which this additional cost brings little benefit (Bechtel & Abrahamsen, 1991). Because the examples we discuss later focus on the integration of fairly high-level concepts, such as traits or beliefs, we will focus almost exclusively on localist models, because they address the central problem in that context without the cost and complexity of a distributed representation.

Now that we have seen what parallel constraint satisfaction processes are and what they can do, let us turn to a more general consideration of the various assumptions of Gestalt psychology and their parallels to parallel constraint satisfaction processes.

Principles of Gestalt Psychology: Their Relation to Principles of Parallel Constraint Satisfaction Processes

As we already noted, much classic work in social psychology was grounded in Gestalt principles. We now examine some of the basic assumptions of Gestalt psychology, many of which differ considerably from the assumptions of much of the psychology of the time and also differ quite a bit from assumptions of contemporary psychology. However, as we will show, many of the fundamental assumptions of Gestalt psychology have very close parallels with the characteristics of parallel constraint satisfaction systems. In the following, we focus on five key assumptions of Gestalt psychology:

1. That psychological processing can be thought of in terms of interactions in fields of forces.
2. That processing is wholistic rather than atomistic or elementalistic.
3. That the whole of the perception or concept is greater than the sum of its parts.
4. That the structure of a stimulus, how its components are connected and related, plays a critical role in how it is perceived or thought about.
5. That the psychological field is a dynamic system in which elements continually and mutually

influence each other, and the state of the system is always changing.

Psychological Processing Conceptualized as Interactions in Fields of Forces

When Wertheimer, Koffka, and Köhler began to develop their theoretical framework of Gestalt psychology, they viewed traditional psychology as being essentially atomistic and mechanistic, and proposed instead the adoption of physical field theory as their model (Henle, 1986). Building on Faraday's ideas about fields in physics, the Gestalt psychologists proposed that such fields also exist in the psychological realm and included forces, tensions, and states of equilibria as did their counterparts in physics. According to Koffka (1935), "If the locus of behavior is the physical world, then the field concept which is so powerful a tool in physics must be applied to behavior" (p. 49). Koffka argued that these fields were no less real than those of physics, and it was the goal of psychology to study the causal relation of behavior to these fields and to identify and understand the forces that caused behavior to occur. Just as the arrangement of electrostatic forces in an electrical field determines the flow of current, the arrangement of psychological forces in what Lewin (1935) called the "life space" or what Koffka (1935) called the "behavioral field" determines behavior, perception, and other psychological processes. They believed that from a scientific point of view psychological fields of forces were every bit as real as the forces studied in physics, although they clearly did not view a psychological field as being the same kind of field as a magnetic or electrical field.

In social psychology this assumption of fields of forces found its way into several areas. In group dynamics, the *group* was defined as a field in which the members acted as forces that affected one another (Krech & Crutchfield, 1948). Similarly, *group cohesiveness* was defined as "the resultant of all the forces acting on the members to remain in the group" (Festinger, 1950, p. 274). Heider's (1946, 1958) balance theory proposed field-like relations or bonds among social objects such as persons and ideas that were largely determined by the configuration of forces associated with attitudes, values, and sentiments. Certainly, Lewin's (1935) conceptualization of personality and motivation follows a force field notion. He considered tension, which energized behavior, to be a scalar (i.e., a magnitude or quantity without direction). Force, or valence, provided the direction. For example, a person's needs (e.g., hunger) create tensions that induce valences (e.g., attractive food) in the environment. Behavior (and personality) is thus dictated by the relative position of the person within a field of such forces. Lewin (1935) contrasted this theory to one in which the direction of behavior derives solely from tensions in the person.

Given this insistence on understanding psychological processing in terms of interactions among fields of forces, it is of considerable interest that research on connectionist models and parallel constraint satisfaction processes has shown that there are precise mathematical parallels between the behavior of neural networks described by these models and the behavior of various kinds of physical systems, such as interacting magnetic fields. For example, following work by others, Hertz et al. (1991) pointed out that one important kind of neural network model, a Hopfield net (Hopfield, 1982, 1984), is precisely equivalent, mathematically, to certain kinds of simplified (but highly useful) models describing the interactions of the magnetic fields of individual atoms in a magnetic material. The patterns of influence among atoms in this material precisely correspond to the patterns of influence among neurons in a Hopfield network. Further, researchers have applied a whole host of ideas from statistical mechanics and thermodynamics to the analysis of neural network models (e.g., Amit, 1989; Hertz et al., 1991; Hinton & Sejnowski, 1986). This work suggests that at an abstract, conceptual level, the behavior of psychological systems and processes is similar to—or maybe even isomorphic with—the behavior of certain kinds of dynamic physical systems that can be treated in terms of interacting force fields.

Psychological Processing Is Wholistic

Koffka (1935) pointed out that at the time of Wertheimer's (1912) first experiments in perception, psychologists and physiologists considered nervous system processes to be composed of the excitations of individual receptor cells that then moved along an independent or isolated nerve to the brain where it activated a corresponding independent or isolated brain region. Perception (or consciousness) was somehow the sum of all of these excitations. Koffka (1935) wrote:

> The enormous complexity of behavior was not explained by an equal complexity of processes as such, but only by an equal complexity of a host of separate processes, all of the same general kind but occurring in different places. (p. 54)

The Gestalt psychologists proposed an alternative: "Instead of reacting to local stimuli by local and mutually independent events, the organism responds to the *pattern* of stimuli to which it is exposed ... a unitary process, a functional whole" (Köhler, 1929, p. 103). Further, the change in any single piece of information could directly influence the perception of the whole.

Wholistic processing has been demonstrated by the use of several familiar visual perception examples, such as figures that are perceived in an apparently random configuration of dots, or the perception of an object that completely changes with the slightest change of a single element. In social psychology, Asch (1952) theorized that person perception works in much the same fashion; we perceive other individuals as whole units. Like one of the visual illusions studied by Gestalt psychologists, the perception of personality traits is wholistic. "Each trait possesses the property of a part in a whole. The introduction or omission of a single trait may alter the entire impression" (Asch, 1952, p. 216). Asch also suggested that group behavior was wholistic, that we could not understand groups by treating them as the sum of the behavior of individuals. Again, adding or removing one individual could potentially cause a tremendous change in the behavior of the group.

Wholistic processing of information is precisely what happens in feedback neural network models. Items simultaneously send activation to and receive activation from all the items to which they are connected. As a result, the activation of each item depends on the activation of all other items. Thus, there is no way to separate the interpretation of any individual item from the interpretation of the other items to which it is related, because the activation of each element in the network depends upon the activation of all the other elements in the network.

Moreover, these systems can be seen as a realization of the kinds of processing that Wertheimer (1912) and Koffka (1935) argued were characteristic of the brain. Rather than having the perception of an object be due only to "local and mutually independent events" (Köhler, 1929, p. 103), processing takes place in the interaction among a large number of neurons, and the perception of a stimulus corresponds to a pattern of activation across these neurons.

The Whole Is Greater Than the Sum of Its Parts

This may well be the signature assumption of Gestalt psychology, a phrase so well known as to be a cliché. As a result of their rejection of the atomistic view of psychology, the Gestalt psychologists compared their approach to the molar science of physics. Köhler (1920) showed that the physicist does not try to understand water solely by conducting a molecular analysis of its constituent atoms, hydrogen and oxygen. Why? A completely new system is formed by the combination of these atoms that has properties that cannot be derived by adding the individual properties of each. In the same way, perceptions of the world or of people cannot be derived simply by adding together individual points of stimulation in the perceptual apparatus or by adding together individual features. Rather the combination of perceptual elements leads to new properties that are not simply the sum of the elements.

Although the assumption that the whole is greater than the sum of its parts was widely shared in social psychology, not everyone agreed. For example, the study of group behavior has often been reduced to a study of individuals, consistent with Floyd Allport's (1924) proclamation that "there is no psychology of groups which is not essentially and entirely a psychology of individuals" (p. 4). In contrast, Asch (1952) posited that group action has laws that are not reducible to those pertaining to individuals in isolation.

One problem Gestalt psychology always had was that as good as this phrase sounded and as much as it seemed to fit with many people's intuitions, it was never quite clear how it could be implemented in an explicit psychological process model. However, neural network models can provide a computational implementation of this assumption. Because most kinds of neural network models are nonlinear rather than linear systems, they can model situations in which the addition of small amounts of information or the change of state of a small part of the network can lead to radically different states of the system and therefore quite different meanings. For example, if we think of the representation of the possible states of a neural network in terms of the energy surface discussed earlier, then the addition of only a few elements or only a small change in one part of the network is sometimes sufficient to ensure that the system will settle in a very different energy minima. That is, the network will arrive at a very different final state.

Such systems often have emergent characteristics, with properties that cannot be predicted from any kind of sum or average of its components. Rumelhart and McClelland (1986b) stated this point quite eloquently:

> We certainly believe in emergent phenomena in the sense of phenomena which could never be understood or predicted by a study of the lower level elements in isolation. These phenomena are functions of the particular kinds of groupings of the elementary units. ... For example, we could not know about diamonds through the study of isolated atoms; we can't understand the nature of social systems through the study of isolated individuals; and we can't understand the behavior of networks of neurons from the study of isolated neurons. Features such as the hardness of the diamond is [*sic*] understandable through the interaction of the carbon atoms and the way they line up. The whole is different than the *sum* of the parts. There are nonlinear interactions among the parts. (p. 128)

The role and importance of nonlinearity in connectionist networks. One of the reasons why earlier work on neural networks (which began in the 1940s and continued until the late 1960s) largely stopped is because Minsky and Papert's (1969) critique of one kind of neural network, the *perceptron*, demonstrated that these early networks could handle only linear problems. Yet, many researchers recognized that many kinds of psychological processes required nonlinear processing, in which the end result of processing a set of elements was not based on a linear function of the individual elements. Partially in response to this issue, many current models use a nonlinear activation function, such as the kind we discussed earlier, where the activation of a node is a nonlinear function of its inputs.

One way to interpret what the Gestalt psychologists were claiming is that the meaning of a stimulus configuration (the whole) cannot be calculated using any kind of linear integration rule, such as averaging or summing a set of stimulus elements (the sum of its parts). Currently, there are several areas in psychology in which it is clear that the processing of stimulus configurations cannot be modeled by a linear function.

One important example of this is in work on human categorization. A central question has been whether category membership for objects can be characterized in terms of linear rules, such as weighted averages or sums of the features. Categories that can be defined by such a linear rule are termed *linearly separable*, whereas categories that cannot be defined by a linear rule are termed *nonlinearly separable*. Research has demonstrated that oftentimes human categories are not linearly separable; that is, there is no linear function that can be used to calculate category membership (see Medin & Wattenmaker, 1987, for a discussion). Instead, nonlinear rules must be used. An example of a nonlinearly separable category is the following: An individual who works outside in the summer could be a house painter, as could an individual who works inside during the winter; however, an individual who works outside during the winter would not be a house painter. Obviously, adding features is not sufficient to define membership in this category.

Further, there is evidence that linearly separable categories are no easier to learn than are nonlinearly separable categories (Kemler-Nelson, 1984; Medin & Schwanenflugel, 1981; Wattenmaker, Dewey, Murphy, & Medin, 1986). Medin and Wattenmaker (1987) argued that linear separability may not be an important constraint on human categories because people's categories "typically have more internal structure than can be captured by an independent summing of evidence or by similarity to a prototype" (p. 37). Thus, category membership judgments are often greater than the sum of their parts. The work on category judgments demonstrates that whether a system can compute a nonlinear function is not simply academic. Rather, there are important aspects of human cognition that are clearly dependent on the ability to compute nonlinear functions.

Actually, there are two ways in which such systems are nonlinear. First, as noted, the activation rule for individual units is frequently a nonlinear, sigmoid-shaped function that asymptotes at some maximum value. This means that the impact of incoming activation on a single node is strongly dependent on the current activation of the node. If the current activation is low, then strong incoming activation can make a large change; whereas, if the current activation is nearing asymptote, then strong incoming activation will make only a small change. Further, as a number of individuals have noted, systems with such nonlinear functions can compute functions or rules that cannot be computed by systems using linear activation functions. For instance, systems with nonlinear activation functions can make nonlinearly separable categorizations, as in the house painter example. Thus, systems with nonlinear functions have greater computational power than linear systems.

Further, nonlinear activation functions are fundamental to feedback or attractor models. If there were no maximum value, as in a linear function, then the activation of the system would increase indefinitely. However, because these systems have nonlinear, asymptotic activation functions, they will reach a stable state rather than "blowing up." And, as argued earlier, feedback networks can perform tasks—such as schema construction, pattern completion, and optimization—that cannot be done by nonfeedback networks.

Second, and perhaps more important, the behavior of the entire system is nonlinear. If we think about attractor or feedback systems in terms of the energy surface that represents the state of the system, it is clear that small changes in the initial value of the system can lead to large changes in the final value (or vice versa) by affecting how close the initial state of the system is to different attractors. Small changes that move the initial state from being more similar to attractor *A* to being more similar to attractor *B* can make large changes in the final state of the system.

There seem to be a number of phenomena that have this flavor, which we discuss in more detail later. For example, Asch's (1944) work on change of meaning and related work on the generation of emergent attributes from novel combinations of social concepts (Asch & Zukier, 1984; Hastie, Schroeder, & Weber, 1990; Kunda, D. T. Miller, & Claire, 1990) seems to be the result of a nonlinear system as the emergent attributes cannot be predicted by a linear rule. Another possible example of the operation of a nonlinear, dynamic system is Vallacher, Nowak, and Kaufman's (1994) recent work on the dynamics of human judgment. They demonstrated that when individuals are asked to give their evaluations of a target about whom they are of two minds—both positive and negative—the evaluation continually oscillates, rather than quickly settling at a value that is the average of the target's attributes.

There are also phenomena in which small additions of information can lead to large differences in the final state. Conversion experiences or other situations where one suddenly changes one's mind provide one example of this, where the addition of seemingly small amounts of information can play a major role in the final state of a disordered system.

Finally, belief perseverance (e.g., C. A. Anderson, Lepper, & Ross, 1980; Ross, Lepper, & Hubbard, 1975) may provide another example of a nonlinear system. Here, once individuals make an initial judgment, it becomes remarkably hard to change. Information that would have strongly affected the final judgment if it had been received before the judgment was made has little effect once the individual has made up his or her mind. Again, this can be visualized in terms of the energy surface. Before an individual forms a judgment, the information can be viewed as defining the initial state of the system. The location of the initial state will affect which attractor or final state the system is likely to move towards. However, once the system has reached the bottom of an attractor or minimum, it is quite difficult to climb out of the valley or well. Thus, information that would have affected the initial state of the system, and therefore which attractor would be reached, may well be far too weak to move the system out of the attractor once it has been reached. However, in a linear system, the point in time the information is received should not have such a dramatic impact.

Emphasis on Structure: How Things Are Connected and Related

Gestalt psychologists proposed that our perceptions of the world are guided by organizational principles such as good form, proximity, and similarity. Thus, even given an incomplete figure we perceive a circle rather than a set of curved lines, and a triangle rather than three dots. We perceive alternating rows of roses and tulips, rather than an undifferentiated field of flowers. These principles apply not only to spatial relations but to temporal ones as well. Temporal organization enables our perception of causality. Without it, Koffka (1935) wrote, "One billiard ball would run, come in contact with another, stop, and the other would begin to roll. Two trains would collide, leave the tracks, and cars turn turtle and become wrecked; another mere consequence" (p. 383).

Heider (1944) incorporated these Gestalt principles into his analysis of causality. Viewing cause and effect as parts of a single unit, he demonstrated how similarity and proximity influence the creation of causal attributions. Later, he extended this analysis in balance theory (Heider, 1946). For interpersonal perception, the parts of the units are considered to be persons and objects, as well as the relations of these to one another. People are said to perceive these interpersonal and attitudinal bonds as units. The bonds themselves follow the same Gestalt organizational principles. For example, similarity creates a balanced state if "all parts of a unit have the same dynamic character (i.e., if all are positive, or all are negative), and if entities with different dynamic character are segregated from each other" (Heider, 1946, p. 107).

Thus, Gestalt psychologists argued that structure played a central role in the interpretation of stimuli. One could not just sum up all the elements; one had to know how the elements were organized, what was related to what, and how they were related. The same kind of argument has been made for the importance of schema-type representations, in which the organization of attributes plays a central role.

Again this is a key part of parallel constraint satisfaction models. Elements in a parallel constraint satisfaction network are connected to other elements in the network. These connections may be positive and negative, and the size of the weights may differ considerably. The activation (and thus interpretation) of the elements in the network depends on the nature of the connections among the elements. The activation of any element is not simply a function of the activation of the other elements, but also critically depends on the nature of the connections among them. Put another way, the final state of the system depends on the pattern of constraints among the elements of the system. The final state depends on the structure of the system. Different patterns of constraints among precisely the same elements will lead to very different states of the system.

Emphasis on Dynamics: Change, Equilibrium, Tension

Finally, by adopting physical field theory as their model, Gestalt psychologists emphasized the dynamics produced by their fields of forces. Opposing forces create tensions, which in turn cause change to occur so as to reach some end-state. Terms such as *balance*, *consistency*, *equilibrium*, and *harmony* refer to the preferred state of a dynamical system in which the degree of tension is at a minimum. Whether it is a perceptual, motivational, or behavioral process, a dynamic striving for the end-state always underlies the process itself. Thus, the individual is conceived of as an equilibrium-maintaining system that in psychology translates into "an interest in the processes by which equilibrium is restored once it is disturbed" (Deutsch, 1968, p. 421).

The dynamical system was a central feature of Lewin's (1935) conception of life space and theory of motivation. Tensions arising in regions of the life space create a state in which a person strives for goals that ultimately lead to tension reduction. Heider (1946) also incorporated a dynamical approach by proposing that attitude and interpersonal bonds are driven by forces toward balance or consistency. He stated, "If no balanced state exists, then forces towards this state will arise. Either the dynamic characters will change, or the unit relations will be changed through action or through cognitive reorganization" (Heider, 1946, p. 341). Likewise, in Festinger's (1957) theory of cognitive dissonance, dissonance after a decision reflects a change from a previous state of equilibrium. Having chosen an alternative inconsistent with one's prior beliefs, dissonance pressures one to change his or her cognitions or to develop new ones that will eventually lead to a new state of equilibrium.

Thus, the idea of tension within a field of forces, and the resulting attempts to reduce that tension, played a central explanatory role in Gestalt psychology. Systems under tension would evolve towards a state that minimized that tension. The evolution of the system toward reduced tension was responsible for the movement of the individual through psychological or physical space, resulting in psychological or behavior change.

This idea is remarkably similar to a parallel constraint satisfaction system. As noted earlier, the positive and negative relations among the nodes in such systems represent the constraints among the possible states of activation of the nodes and essentially characterize the degree of tension in the system. As parallel constraint satisfaction processes work to satisfy the constraints imposed by the positive and negative relationships and minimize the energy of the system, one way to view what is happening is that this is an attempt to minimize the degree of tension or conflict. One is trying to find the minimum level of tension possible, given the constraints imposed by the actual set of relations among the cognitive elements. As many researchers have noted, neural networks can be viewed as trying to find the minimum energy or maximum degree of organization of the system. Alternatively, one can also think of this as trying to find the maximum degree of organization or minimum degree of disorganization of the system.

Note that parallel constraint satisfaction models do not propose that one will necessarily ever reach an absolute minimum, but rather that one seeks a minimum given the current set of constraints. This also seems to be the most reasonable interpretation of how the Gestalt psychologists thought about this issue. Lewin (1935), for example, talked about the minimum tension given the current state of the system.

Further, many neural network models can be explicitly characterized as dynamic systems in which the state of the system changes over time. For example, one can look at how, after initial input, the system evolves over time to an increasing degree of organization, and examine the trajectory it follows. Or, once a system has reached a minimum or equilibrium state, one can examine how new stimuli first reduce the organization of the system and then how the system evolves to a new state. Further, feedback networks, such as Hopfield nets (Hopfield, 1982, 1984) and Boltzmann machines (e.g., Hinton & Sejnowski, 1986), have been explicitly characterized as a kind of dynamic system called *attractor systems*, in which the minimums in the energy surface are attractors toward which the state of the system tends or is "pulled" (Hertz et al., 1991).

Applications of Parallel Constraint Satisfaction Processes

Having outlined the parallels between Gestalt ideas and basic aspects of parallel constraint satisfaction processes, we now take three key areas in social psychology in which Gestalt ideas played a central historical role and examine how parallel constraint satisfaction processes have been applied in each domain. The three areas are: impression formation and causal attribution, cognitive consistency, and goal-directed behavior. In the process, it should become clear how these applications realize many of the fundamental assumptions of Gestalt theory.

Our focus on these areas is not purely academic or historical. We hope to convey that parallel constraint satisfaction models provide extremely useful conceptual and methodological tools for advancing theory and research in social psychology. First, these models provide concrete, computational implementations for many ideas such as wholistic processing of information or the structural dynamics that underlie balance theory and cognitive dissonance. Second, these models provide one way to think about how individuals, within a brief time frame, can integrate the wide array of information available in a social interaction and so behave competently. Third, these models provide a way to capture the dynamics of social interaction, the ebb and flow of social thought, motivation, and action (see also Vallacher & Nowak, 1994; Vallacher et al., 1994).

Before proceeding, we offer two caveats. The first is that in each of the domains discussed, parallel constraint satisfaction processes provide only some of the pieces to the theoretical puzzle (although we think the pieces are major ones). Any cognitive model contains assumptions about both process and representation (J. R. Anderson, 1978); the two are inextricably linked. Parallel constraint satisfaction processes have much to say about how processing takes place, but little to say about how information is represented. For instance, parallel constraint satisfaction processes can be applied equally well to both distributed and localist representations; or the features in a representation can be of many different kinds; or the links among elements can be given a variety of different interpretations, such as causal or associative ones. Moreover, the notion of a parallel constraint satisfaction process may capture only part of cognitive processing. Additional processing assumptions may also be made. For example, in our models of impression formation (Read & L. C. Miller,

1993) and cognitive consistency (Read & L. C. Miller, 1994), we argued that parallel constraint satisfaction processes are applied only to items that are attended to.

Thus, in each case we must make theoretical assumptions about such things as representation and how nodes and links should be interpreted in a given model. For example, when we discuss impression formation and the integration of multiple sources of information, we make specific assumptions about how traits and situational information are represented and how those representational assumptions relate to the parallel constraint satisfaction processes that operate on those representations. In the section on goal-directed behavior we see that various authors have assumed that nodes correspond to such entities as goals, actions, and situational features and that links among nodes represent such things as whether goals are compatible or incompatible with one another—whether a particular action helps to achieve or block a particular goal and whether particular situational features help or hinder the attainment of a goal. Thus, in each case parallel constraint satisfaction processes are inextricably intertwined with additional theoretical assumptions about such things as representation or other aspects of process. This is no different from any other kind of cognitive model, in which one must simultaneously make assumptions about both representation and process.

However, one major advantage of neural network models is that they provide a computational implementation of one's assumptions. Typically, our theories and models are strictly verbal ones. We verbally specify the variables, their relations, and the hypothesized processes involved. But, especially when our theoretical models and assumptions are complex and involve changing dynamics over time, we cannot tell what patterns and behaviors would be the emergent product of this complexity. Neural network models provide us with a new set of tools for testing our theoretical assumptions and hypotheses. To develop a viable theory using this conceptual language, we must embody our theoretical assumptions in an explicit neural network model. This results in a computational implementation of our theory and its assumptions, the outcome of which can be examined by running these models. Unlike verbal models alone, this output can then be compared against other known phenomena or behavioral outcomes.

The second caveat we offer is that our focus is on how parallel constraint satisfaction processes integrate information into a coherent whole, once the information has been activated. We do not examine how concepts are initially activated. Although this is clearly an important problem that has to be addressed as the application of connectionist models to social phenomena develops, our focus here is on the integration of information rather than its initial activation. Nevertheless, it is important that researchers who wish to apply the kinds of models we discuss utilize procedures to explicitly identify what information has been activated.

A variety of different procedures are possible. For example, if one wished to predict a priori which concepts are linked and are likely to be activated, one could use the free association procedure used by Kintsch (Kintsch, 1988; Mannes & Kintsch, 1991), in which participants are given the concepts that are part of explicit problem statements and then asked to free associate to these concepts. Another possibility is Graesser and Clark's (1985) question-answering procedure that has been used to map out the structure of a variety of different concepts by extensively questioning participants about the features of concepts and the relations among them. Or, if one wished to assess the concepts that were actually activated and the links among them, one could gather think alouds or use the kind of thought-listing procedures that are often used in attitude change studies. Regardless of whether these or other procedures are used, it is important that researchers map out which concepts are activated. We now address how parallel constraint satisfaction processes have been applied in each of three key areas.

Impression Formation and Causal Attribution

Many of the classic contributions concerned with impression formation and causal attribution were based on aspects of the Gestalt ideas we discussed earlier. For instance, Asch, both in his classic work on impression formation (Asch, 1946), as well as in his more recent work (Asch & Zukier, 1984), argued that the processing of information in forming an impression is wholistic. Each piece of information influences the interpretation of all the other pieces. Moreover, the processing of the information relies on Gestalt principles of organization and the dynamics of meaning. Further, Heider's (1944, 1958) work on causal attribution and perception was based on Gestalt ideas about organization and structure—such as principles of unit formation—and he focused on the dynamics of causal perception and attribution.

In many respects these ideas find their parallels in the recent use of parallel constraint satisfaction processes to elucidate both impression formation processes and the processes involved in causal attribution. Before discussing the application of parallel constraint satisfaction processes here, however, let us remind ourselves first of the perceiver's task in forming impressions and in making causal attributions. Following a delineation of the perceiver's task, we ask these questions: What are the problems, familiar to some Gestalt theorists, that the perceiver must grapple with, and how do parallel constraint satisfaction processes address these problems?

Mundane social perception requires the integration of numerous pieces of information to form a coherent model of the interaction and the parties to it. In a typical social interaction we have information activated about a universe of different things: (a) characteristics of the individuals with whom we are interacting—such as gender, race, personality, role, age, and so forth, (b) our relationship with them and past interactions with them (and perhaps others), (c) information about ourselves, and (d) considerable information about the situation. All of this information must be integrated online, during the process of social interaction, if we are to competently perform our part. Although there are frequently similarities between different interactions, the precise configuration of information is almost always unique. Therefore, no two interactions are precisely the same, and as a result, there is almost never a preexisting representation for the interaction. The representation of each interaction is in some ways novel and must be built anew each time.

The perceiver's task points to several problems that theories of causal attribution and impression formation need to address. First, how do we create novel representations—for example, of combinations of various traits? And, second, how do we integrate and interweave information about persons, situations, and relationships in understanding social interaction?

The problem of creating novel representations: Using parallel constraint satisfaction processes. Social perception and impression formation can be viewed as the combination of numerous concepts to create novel representations. For instance, Barsalou (1992) noted that the comprehension of narrative text—which involves the use of considerable social knowledge—or the planning of novel activities, were problems of conceptual combination in which an individual must combine a variety of concepts to create a novel representation. We made a similar point about social perception and causal reasoning (L. C. Miller & Read, 1991; Read & L. C. Miller, 1993).

Parallel constraint satisfaction processes can be used to explain how individuals combine social concepts to create novel representations that are more than the sum of their parts. For example, they could be applied to Asch's (1946) classic findings that the impression of an individual can sometimes be dramatically changed by the addition or subtraction of a single piece of information, or Asch and Zukier's (1984) work showing how people combine inconsistent trait pairs to form integrated impressions. They could also be used to elucidate how people combine other kinds of inconsistent social concepts, such as inconsistent role concepts (Hastie et al., 1990; Kunda et al., 1990), stereotypes, and individuating information (Kunda & Thagard,1996); or they could be used to illuminate the processes underlying other work on "change of meaning" done by researchers such as Hamilton and Zanna (Hamilton & Zanna, 1972; Zanna & Hamilton, 1977) and Wyer (1974).

Here we concentrate on how parallel constraint satisfaction processes might play a role in how individuals combine inconsistent traits when forming an impression. We propose that parallel constraint satisfaction processes operate upon the underlying components or structure of trait concepts so as to form a new conceptual structure from the components of the combined traits. To understand how the process works, it is necessary to make explicit assumptions about the structure of traits and how they are represented. (This is an example of the necessity of making joint representation–process assumptions.) We propose that traits can be viewed as frame-based or schema-based representations (John, 1986; L. C. Miller & Read, 1987, 1991; Read & L. C. Miller, 1989, 1993) with slots for the various components of the trait—such as goals, conditions that would instigate those goals, and the behaviors that would be enacted to achieve those goals. Further, these slots contain information about the range or distribution of attributes that can fill those slots. When individuals attempt to integrate a set of traits to form a coherent impression, links will be formed among the appropriate slots of the various traits.

For example, consider the trait pair "generous–vindictive," used by Asch and Zukier (1984). Read and L. C. Miller (1993) argued that generous and vindictive each have, as part of their representation, slots for the goals associated with that trait and slots (with default and possible values) for the behaviors that can achieve those goals. Vindictive activates goals such as hurting others and gaining revenge, whereas generous activates goals ranging from helping others to self-presentational goals. The self-presentational goals of generous are positively linked to the goals of hurting others and getting revenge, because presenting a false positive front can be used in the service of hurting others. A partial diagram of this conceptual structure is given in Figure 3. Further, the plans and behaviors associated with generous have a positive link to the self-presentational goal. As a result, because vindictive goals can explain self-presentation, then the plans associated with generous will be positively linked to the goals associated with vindictive.

Further, the behaviors associated with vindictive will have an inhibitory link to some of the default goals of generous (e.g., helping others) and a positive link to the goals of hurting others and revenge. A likely resolution of this trait discordance is that an individual appears generous to enable him to be vindictive, because such a combination of slots and their values is the most coherent.

Thus, two traits that have contradictory goals may have inhibitory links between the goal slots of the two traits, whereas when the behaviors associated with one trait can achieve the goals associated with another trait,

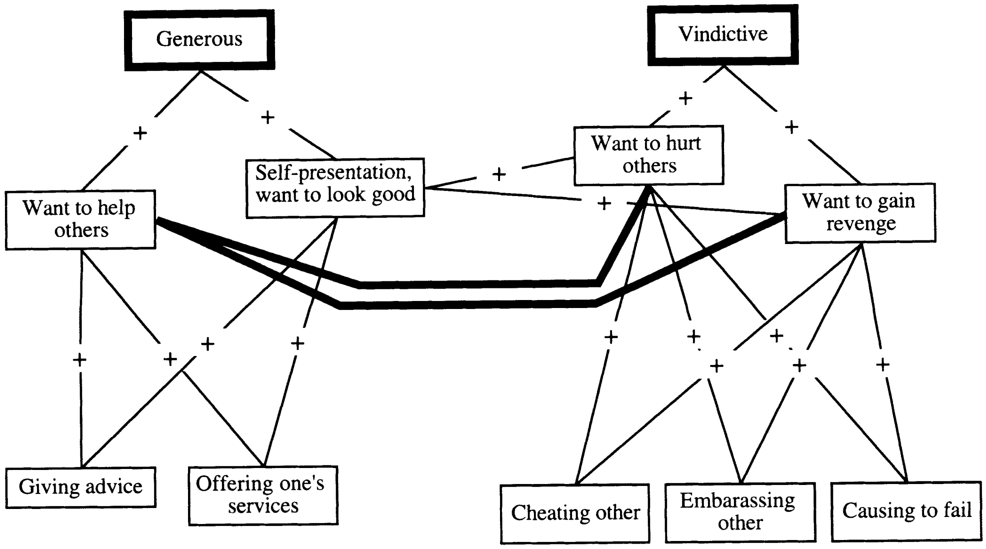

Figure 3. Partial network of the possible conceptual structure involved in the integration of the incongruent traits generous and vindictive. Thin lines are positive links, and thick lines are negative.

there may be a positive or excitatory link between these two slots. Moreover, sometimes the slot of a trait may specify several alternative features that may fill that slot. Some of those alternatives may be consistent with the features in the slot of another trait and thus have a positive link to them, whereas other alternatives may be inconsistent and have a negative link to the alternatives in the slot of the other trait.

The positive and negative links among the slots of the traits provide a set of constraints on the final impression formed from the initial ensemble of traits. Once this network of links is formed, a parallel constraint satisfaction process is applied to the network until the activation of the nodes asymptotes. The final activation of the various nodes in the network is a solution to the constraints specified by the links among the nodes and can be taken as the impression that is formed.

Understanding the interwoven nature of structures underlying social meaning: Using parallel constraint satisfaction processes. A second problem raised in considering mundane social perception is that individuals—in understanding social interaction—do not just integrate information that is all of the same type, such as traits. Rather, they must integrate a variety of different kinds of information, such as that about persons, situations, and relationships. How might they process these different kinds of information wholistically? This information is unlikely to be neatly isolated as person, situation, or relational information. Rather, these concepts are apt to activate one another and be highly interwoven. Lewin (1935) argued that we needed to understand the person-in-the-situation as if these concepts could not logically be disconnected. A similar, updated version of this perspective can be found in the ground-breaking work by Cantor, Mischel, and Schwartz (1982; see also Mischel & Shoda, 1995). How might we study how individuals naturalistically understand the relations among these concepts?

We have argued that there may be common units underlying our representations of persons, situations, relationships, and other social concepts (Read & L. C. Miller, 1989, 1993). For instance, just as traits may have underlying script structures, so too may other social concepts, such as situations (L. C. Miller, Cody, & McLaughlin, 1994) and emotions (Lutz, 1988; Roseman, Spindel, & Jose, 1990; Shaver, Schwartz, Kirson, & O'Connor, 1987; Wierzbicka, 1994). These script structures are likely to be interwoven rather than sitting neatly in isolation.

For example, in discussing the William Kennedy Smith rape trial, L. C. Miller et al. (1994) argued that the same pieces of information crucial to the activation of a situational construction (e.g., this is a rape situation in which he forced her to have sex against her will) may be part of the representation of an emotional appraisal (e.g., she was angry and upset because he forced her to do something she didn't want to do) and may lead to a person attribution (e.g., he's a Dr. Jekyll and Mr. Hyde). Because these different concepts may involve many of the same activated script components, shifts in our inferences about a character or an emotion may simultaneously influence inferences about this character's other traits—as well as the situation, the other character, the relationship among characters, and so forth. These complex mutual influences—so difficult to capture using standard social methods—can be modeled using parallel constraint satisfaction processing models.

Furthermore, social concepts such as persons (Read & L. C. Miller, 1989), emotions (Markus & Kitayama, 1994), and situations (L. C. Miller et al., 1994) may be tremendously fluid. Fill in one of the components of the underlying script differently—or alter the script slightly—and the meaning may shift, perhaps dramatically, in a rather wholistic fashion. Also, these concepts—and their underlying structures—are apt to mutually influence one another and combine with other concepts to create new emergent wholes. Attributions about persons, situations, emotions, and so forth are apt to depend on the total array of activated information and inference in the network of linked concepts. Our attributions are an emergent product of the changing relationships among these links. Parallel distributed processing models provide the tools to allow us to study such wholistic processing of complex and interwoven social information and inference.

Integration of information over time, as in social interaction. More ambitiously, these processes can be used to understand and explain how parties to a social interaction can comprehend each of the steps in a social interaction, create a representation of that interaction—including the role of dispositional and situational forces—and make inferences about the characteristics of the individuals involved (e.g., their goals, plans, traits, and social roles).

In any social interaction, as the interaction proceeds, we continually receive new information. How is that information integrated to form a coherent representation of the interaction and the individuals in it? We (L. C. Miller & Read, 1991; Read & L. C. Miller, 1993) argued that this can be analyzed as a two-step process in which a parallel constraint satisfaction process is repeatedly applied to a network of activated concepts as new information is received. In the first step, as new information is received, it activates a whole host of related concepts. These newly activated concepts are activated promiscuously and may or may not be relevant to the final interpretation of the interaction.

How is this "cloud" of concepts organized into a coherent representation? The new concepts and the concepts that are still activated from previous steps have positive and negative links to each other, depending on whether the concepts support or contradict each other. In the second step, this heterogeneous network of activated concepts and the positive and negative links among them is organized into a coherent interpretation by the application of parallel constraint satisfaction processes that implement Thagard's (1989, 1992) model of explanatory coherence. Concepts that are supported by other concepts become more activated, whereas concepts that receive no support, or are contradicted by other concepts, are deactivated. Highly activated concepts are taken as the interpretation of the interaction to that step and are stored in long-term memory.

Further, some of the most highly activated concepts remain in working memory and are carried over to the interpretation of subsequent information, where new information activates additional concepts. The resulting heterogeneous network is then organized into a coherent representation and the process continues. (For a related account applied to text comprehension, see Kintsch, 1988.)

This model has been used to analyze how members of couples may form mental models of each other over time in close relationships (L. C. Miller & Read, 1991) and how people understand the interactions they observe in everyday life and explain the behavior of the individuals in them (L. C. Miller & Burns, 1992; Read & L. C. Miller, 1993; Seiter, 1993). In particular, we (Read & L. C. Miller, 1989; Read, 1987) and others (Pennington & Hastie, 1986, 1988, 1992) have argued that people understand social action by constructing causal scenarios or narratives from the events. For example, Pennington and Hastie (1986) showed that jurors organize the evidence in a jury trial in the form of a story, or narrative, and that the particular story jurors create predicts their verdict in the trial. In such situations, different stories can be (and often are) constructed from the same facts. Pennington and Hastie argued that the coherence of different stories plays a central role in jurors' choice among alternatives. Thagard (1989) argued that the explanatory coherence or goodness of the alternatives—operationalized as a parallel constraint satisfaction process—will play a major role in which is chosen. Consistent with this proposal, Thagard (1989) used his ECHO program—which implements several principles of explanatory coherence as a parallel constraint satisfaction system—to successfully simulate the decisions made in several famous jury trials. Further, Read and Lincer (1994) and Read and Marcus-Newhall (1993) empirically demonstrated that principles of explanatory coherence play an important role in social inference.

Read and L. C. Miller (1993) used this model to argue that stages in the dispositional inference process that have been proposed to be serial and consciously controlled, may actually occur in parallel and require little conscious control. Specifically, Gilbert (1989; Gilbert, Pelham, & Krull, 1988) suggested that the dispositional inference process consists of three sequential stages: (a) identification of the behavior, (b) dispositional inference from the behavior, and (c) correction of the initial dispositional inference for the impact of situational forces. Gilbert argued that because high cognitive load tends to reduce the impact of situational factors on dispositional inference, this therefore demonstrates that the inferential and the correction stages are separate, sequential stages and that the correction stage is a conscious, controlled process.

However, Read and L. C. Miller (1993) argued that the integration of dispositional and situational constraint information could actually occur in parallel and that the use of neither of these two types of information is necessarily more conscious or controlled than the other. They also used Thagard's (1989) ECHO computer simulation to successfully simulate such a parallel process model. The simulation assumed that (a) dispositional and situational explanations are competing explanations and have an inhibitory link; (b) the extent to which nodes can give and receive activation is partially a function of the amount of attention they receive; and (c) under high load, perceivers may manage resources by preferentially withdrawing attention from some of the concepts—in Gilbert's case, the situational information. As a result, under high cognitive load the situational explanation has a weaker inhibitory impact on the dispositional explanation, and therefore the dispositional explanation was stronger. This was precisely Gilbert's pattern of results. Thus, Gilbert's results did not strongly argue that the integration of dispositional and situational information were separate, serial stages.

Causal unit formation. Interestingly, this approach to causal reasoning has some strong similarities to Heider's (1944) ideas on causal reasoning, particularly his ideas on the role of causal unit formation. Heider proposed that one could analyze causal reasoning in terms of principles of causal unit formation, where various factors such as temporal or spatial proximity, salience of a cause, or causal expectancies affected the strength of causal unit formation between a potential cause and an effect and thus affected the perception of causality.

In our approach the activation of concepts is a function of the strength and sign of the links among these concepts as well as the activation of the other linked concepts. Causal unit formation factors, such as temporal and spatial proximity, or causal expectancies, should affect the strength of these links and therefore the degree of activation that can be sent. Further, we argue that the potential for a concept to send and receive activation is a function of the degree of attention it receives. Thus, the salience of a concept, which should affect the degree of attention to it, should affect its ability to send and receive activation. We are thus in the interesting position that Heider's ideas on causal unit formation, which are at least 50 years old and strongly based on Gestalt principles, can be naturally integrated with what are regarded as the newest and most novel models of cognitive processes.

Wholistic processing versus averaging of information. There is a controversy in social psychology that at the least dates back to Asch (1946) concerning how information is integrated in forming an impression. Asch argued that the integration of information was a wholistic process in which the meaning of each element of the impression influenced the meaning of all the other elements. One result of this wholistic process was that the meanings of different elements could change radically depending on the other elements. For example, Asch (1952) reported the results of a study in which he gave participants the following statement, originally made by Thomas Jefferson: "I hold it that a little rebellion, now and then, is a good thing, and as necessary in the political world as storms are in the physical" (p. 421). Half of the time the statement was accurately attributed to Jefferson, and half of the time participants were told it was made by Lenin. Participants' written responses to this statement portrayed very different understandings of its meaning, depending upon whether it was ostensibly made by Jefferson or Lenin. For example, participants who thought the statement was made by Jefferson talked about it in terms of peaceful change of political control, such as change of party control, establishing a third party, or agitation intended to keep the politicians on their toes; when it was attributed to Lenin, participants were more likely to talk about outright revolution, the release of pent-up frustrations, and overthrowing the old order.

According to Asch, the reason people's evaluations are quite different in the two conditions is that the actual thing being judged, the object of judgment, has changed dramatically: "The fundamental fact involves a change in the object of judgment, rather than in the judgment of the object" (Asch, 1952, p. 424).

Asch contrasted this change-of-meaning approach with an algebraic model. Although Asch thought he had convincingly defeated the algebraic model, this was only the first shot in a long battle. Probably the best representative of the other side is the work of N. H. Anderson (1981). Anderson and others argued that the integration of information could be understood as an algebraic, weighted averaging process rather than a wholistic, change-of-meaning process.

Although we suspect that many social psychologists have a sneaking sympathy for Asch's wholistic idea, the averaging side of the argument was greatly aided by N. H. Anderson's (1981) ability to provide an explicit information integration rule, while those arguing for change of meaning (e.g., Asch, 1946; Hamilton & Zanna, 1972; Wyer, 1974; Zanna & Hamilton, 1977) were hampered by the absence of any comparably explicit rule. However, the change-of-meaning side need no longer labor under such a disadvantage. As noted earlier, parallel constraint satisfaction processes provide an explicit computational model of a wholistic, change-of-meaning process. Because parallel constraint satisfaction processes are nonlinear systems and because the different elements in the network can interact with many others, a situation can be modeled in which the whole is truly greater than the sum of the parts and the final impression cannot be predicted from a weighted average or any other kind of monotonic function. For example, under some circumstances the addition of a single element can radically change the pattern of activation of all the other elements.

Note that whether you get change of meaning or what looks like averaging depends on the elements that are integrated, the relations among them, and the relation of the new element to the other elements. Although in most instances the results of information integration using a parallel constraint satisfaction network may look like an averaging process, under some circumstances the nonlinear nature of the network will lead to a change of meaning.

Something like change of meaning is not a property of all connectionist models. Models that assume linear, rather than nonlinear, activation functions do not exhibit the dramatic shifts that would be necessary to model change of meaning (although most current models do use nonlinear activation functions). For example, Kashima and Kerekes (1994) presented a model of impression formation—based on J. A. Anderson, Silverstein, Ritz, and Jones's (1977) distributed memory model—that successfully simulated several aspects of N. H. Anderson's (1981) weighted averaging model. However, because this model uses a linear activation rule, it does not exhibit the Gestalt-like properties of a parallel constraint satisfaction system.

Cognitive Consistency

Gestalt processes played a fundamental role in work on cognitive consistency. For instance, in Heider's (1946, 1958) balance theory, the whole notion of a balanced or unbalanced system relied on Gestalt notions of structural dynamics. Heider believed that the relationships among people could be understood in terms of the structure of the relations among people and the objects in their environment. Aspects of this structure, such as good form, drove the dynamics of the system. Systems that were unbalanced (lacking good form) experienced a tension or dynamic that attempted to move the system to one of greater equilibrium or balance. Other models of cognitive consistency, such as cognitive dissonance (Festinger, 1957) and many others (for a review, see Abelson et al., 1968) also relied heavily on Gestalt ideas about structural dynamics and organization.

Thus, there seems an obvious mapping between parallel constraint satisfaction processes and the Gestalt processes that were the basis for balance theory (Heider, 1958) and cognitive dissonance theory (Festinger, 1957). It is particularly interesting that some researchers in connectionist modeling (e.g., Smolensky, 1986; Thagard, 1989) talked of the harmony of the system, as Jordan (1968) noted that Heider, in discussing balance principles, had begun to talk of harmony of relations rather than balance or imbalance.

Recently, several authors have begun to investigate this mapping. Gabrys (1989) and Read and L. C. Miller (1994) modeled balance principles as a parallel constraint satisfaction process, whereas Shultz and Lepper (1992, 1996) and Read and L. C. Miller (1994) successfully modeled the results of several classic cognitive dissonance experiments. Other authors have looked at more general notions of cognitive consistency. Spellman, Ullman, and Holyoak (1993) recently showed that a parallel constraint satisfaction system can be used to model the changes that occur in political beliefs, given new information. And Seiter (1993) examined how beliefs shift and affect judgments about whether someone is lying or telling the truth.

General account of parallel constraint satisfaction processes and cognitive consistency. Belief systems can be represented as networks of nodes and the links among them, where each node represents a belief proposition and the links represent the relations among the beliefs. Links may be either positive or negative, and they may differ in strength. A positive link indicates that two beliefs are consistent with or support each other, whereas a negative link indicates that they are inconsistent with or contradict each other—with the strength of the link indicating the degree of support or contradiction. In line with our earlier discussion, these links represent the constraints among the various beliefs of the individual. Further, the nodes representing the propositions may differ in activation, indicating the initial strength with which the corresponding belief is held.

The consistency of this network can then be evaluated by applying a parallel constraint satisfaction process to the network. Activation is spread in parallel through the network until the activation of all the nodes asymptotes. The final activation of each node is a function of its initial activation, the strength of the positive and negative links to other nodes and the activation of those nodes to which it is linked. Thus, the links among the nodes act as constraints on the final activations of each node in the network.

The activation of each node can be viewed as its degree of acceptability or belief. Thus, belief in a proposition is the result of a set of multiple constraints among the nodes in the belief system. Beliefs that are mostly supported by other beliefs will be positively activated and therefore acceptable, whereas beliefs that are contradicted by many other beliefs will be negatively activated and therefore not believed.

Consequently, a parallel constraint satisfaction network provides a way of evaluating how the strength of individual beliefs is affected by their place in an interconnected network of beliefs, as well as providing a mechanism for assessing the overall coherence of an entire belief system by using something like a measure of the energy (Hopfield, 1982, 1984) or goodness of fit of the network. Moreover, such a network can model how both individual beliefs and the entire system can change as the result of the interaction among multiple interacting elements. These are characteristics largely missing from current models.

Cognitive dissonance. Let us first examine the parallels between cognitive dissonance (Festinger, 1957) and parallel constraint satisfaction processes. As noted, dissonance theory relies heavily on Gestalt ideas about structural dynamics and organization. In dissonance theory, two cognitions are predicted to be dissonant when the obverse of one cognition follows from the other. It was proposed that individuals are motivated to reduce the dissonance, often by changing one of the cognitions.

In the initial description of dissonance theory, dissonance was often discussed as if only two cognitions were involved. However, as Aronson (1968) and others observed, what often determined whether the two central cognitions were dissonant was the set of cognitions within which they were embedded. For example, Aronson (1968) argued that what gave dissonance its "juice" was the implication of the cognitions for the individual's self concept.

Consider the original dissonance experiment by Festinger and Carlsmith (1959). Individuals in that study first spent almost an hour performing an excruciatingly boring task. Once they were finished they were asked to persuade the next participant that the task was actually quite interesting. Half the participants were offered $1 to do this and the others were offered $20 (this was over 35 years ago). Participants who were offered only $1 came to like the boring task, whereas those offered $20 did not change their opinion. Aronson (1968) argued that what made the $1 condition dissonant was not that the participant had lied. After all, the $20 participants also lied. Rather, the $1 was insufficient reason to have lied. The participants had violated their moral standards without good reason. In essence, they were fools for selling out too cheaply. This is a much more complicated set of cognitions than the simple realization that one had said something that was inconsistent with one's beliefs about the tasks.

Let us consider how to model the results of this study, using a parallel constraint satisfaction process (see also Read & L. C. Miller, 1994). Before doing so, we wish to be quite explicit that this simulation does not assume that inconsistency reduction is necessarily a motivated process. Rather, as discussed earlier, the evolution of parallel constraint satisfaction systems toward greater consistency or coherence follows directly from the organization of the network and the way in which activation is updated. In such a network the impact of any activated goal or motive is simply treated as another constraint on the evolution of the final state of the network.

We use a modification of Thagard's (1989, 1992) ECHO program. The program uses a standard parallel constraint satisfaction algorithm adopted from McClelland and Rumelhart's (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982) interactive activation model of reading. Activation is spread synchronously through the network using Equation 1, the equation for updating the activation of the nodes, which was introduced earlier.

The top diagram in Figure 4 represents the likely cognitions and their relations for an individual in Festinger and Carlsmith's (1959) study who agreed to lie for $20. (Because we obviously do not have access to their participants, we have used a plausible set of cognitions for illustrative purposes.) The boxes at the bottom represent three cognitions: that he had just told someone the task was interesting, that the task was boring, and its opposite, that the task was interesting. The middle boxes contain the two alternatives that he lied or that he told the truth. The cognition "I lied" is positively linked to "The task was boring" and "I told him the task was interesting" because believing one thing and saying another fits the definition of lying. The cognition "I told the truth" is positively linked to "The task was interesting" and "I told him the task was interesting" because saying what one believes fits the definition of telling the truth. Further, there is a negative link between "I lied" and "I told the truth" because they contradict each other.

The final two cognitions in this network are the recognition that the participant was paid $20 to say the task was interesting and the self-concept that the participant is not the kind of person who lies without good reason. These two cognitions would jointly explain lying, because the equivalent of $60 today would be sufficient reason to lie about something so minor; thus there is a positive link between these two cognitions and lying. (In Figure 4, the curved line between the self-concept and the money represent the fact that they jointly explain lying.) In contrast, the two cognitions together have a negative relation to telling the truth as they jointly contradict telling the truth.

Several other assumptions were made in doing this modeling. First, the cognitions "The task was boring," "I told him the task was interesting," "I was paid $20," and "I'm not the kind of person who lies without good reason" were given initial positive activations of .4 to indicate that they already had a certain degree of belief by the time the participant agreed to tell someone that the task was interesting. Second, the cognitions "I told him the task was interesting," "I was paid $20," and "I'm not the kind of person who lies without good reason" had positive links to a special evidence unit or

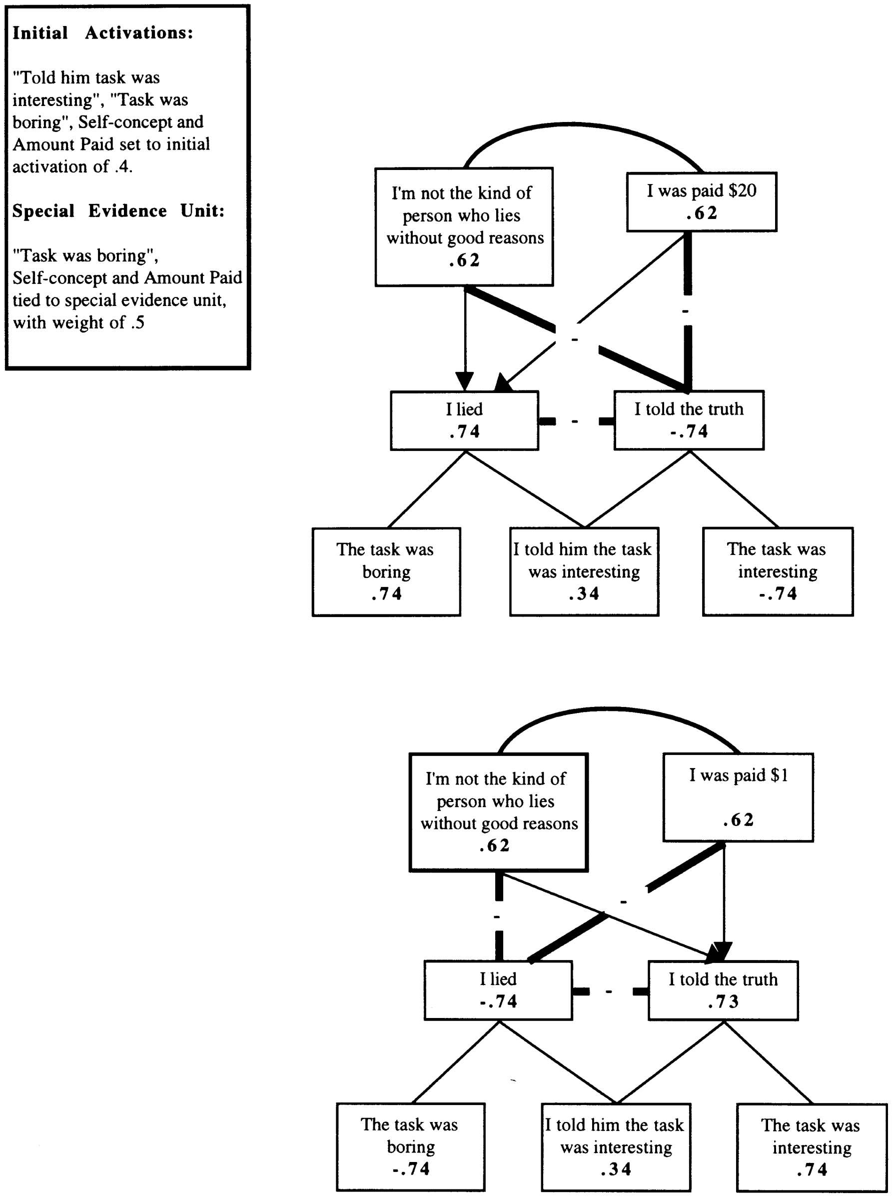

Figure 4. Diagrams of the results of a simulation of the $1 and $20 conditions of the Festinger and Carlsmith (1959) cognitive dissonance experiment.

bias unit (which is clamped at an activation of 1) to indicate that these cognitions had some degree of support or believability—based either on a publicly performed behavior or a well-supported self-concept.

When this simulation is run, we get the finding that we would expect and that Festinger and Carlsmith (1959) obtained. Participants who were paid $20 continued to find the task boring. The cognition "The task was boring" has a strong positive activation, indicating participants should find the task boring, and the cognition "The task was interesting" has a strong negative activation. Further, the cognition "I lied" has a strong positive activation—indicating participants should believe that they were lying—whereas the cognition "I told the truth" has a strong negative activation.

But what does the simulation look like when the participant is paid only $1? The bottom part of Figure 4 represents this situation. There are several changes in the structure as a result of the one key change in cognitions. First, given that $1 is insufficient reason for lying, the self-concept and the recognition of making the statement for $1 have negative relations to the cognition "I lied," because they jointly contradict it. Second, these same two cognitions now have a positive relation to the cognition "I told the truth" because they would jointly explain telling the truth. Everything else in the simulation is the same, including the initial activations of some concepts and their connection to special evidence units.

These changes make a radical difference in the results of the simulation. The cognition "The task was interesting" now has a positive activation, and the cognition "The task was boring" now has a negative activation. Further, the cognition "I lied" now has a negative activation, and the cognition "I told the truth" now has a positive activation. What makes this strong shift particularly interesting is that the simulation was run with the cognition "The task was boring" set to a moderate positive starting activation of .4. Thus, despite the initial positive activation for this cognition, the impact of the other cognitions was sufficient to override it and lead to a strong negative activation.

As noted, this process can be viewed as trying to find a minimum in an energy surface. One fascinating implication of such a view is that it provides a novel way of thinking about how certain kinds of belief change, such as conversion experiences, might come about. In such a system, holding a particular set of beliefs can be thought of as being in a particular minimum or valley—essentially a stable state. Although there might be other minimums in the surface, they are not easily reachable as there is effectively a large hill or mountain in the way. However, suppose we now increase the energy of the system—for example, by presenting some new information inconsistent with the current state of the system. This would raise the energy of the system, and if it was raised sufficiently, one might think of the new state of the system as being on the top of a hill or mountain. From the top of the mountain, other valleys in the energy surface are now more easily accessible and a slight push might direct the system to a different valley or minimum. Thus, increases in the energy of the system—or conversely, decreases in its organization or harmony—would increase the likelihood that the system will end up in a different state.

This might also provide a way of thinking about why dissonance manipulations or other attitude change manipulations do not always have an effect. They may not provide enough energy to push the system to a new state, where alternative minimas are relatively accessible; or, alternatively, a different stable minimum might not exist.

Balance theory. The account of a parallel constraint satisfaction system presented in this article may sound similar in some respects to earlier accounts of balance. Heider (1946, 1958) and those who followed him (Abelson & Rosenberg, 1958; Cartwright & Harary, 1956) viewed balance in terms of the structure or organization of networks of nodes and links, where nodes represented individuals or physical objects and links represented the relations among these objects. There were four types of relations: positive sentiment, negative sentiment, unit (meaning the two objects went or belonged together), and null (meaning no relation). One of the best-known examples of balance is the friendship triad. If all three members like one another, the system is balanced, whereas if one of the pairs dislikes one another, there is psychological tension or imbalance and there is a force toward change of the relations to reduce tension or to achieve a balanced state.

However, balance theory has several important weaknesses. First, there was no way to represent differences in the strength of a node or in the strength of a relationship between two nodes. Thus, for example, there was no way to represent differences in the strength of a friendship. However, as Abelson (1968) observed, these seem to be characteristics of any realistic belief system. Second, balance theory is essentially a formal, syntactic theory and fails to consider the role of conceptual structure in balance. However, as Abelson (1968, 1983) and others have noted, our conceptual structures—what we know about the world—play a central role in our intuitions about balance. One of the best-known examples of this problem with balance theory is the romantic triangle. Although a triangle in which two women love the same man would be balanced from a formal perspective if the two women liked each other (being formally equivalent to a balanced friendship triad), our intuitions tell us that it feels more consistent or balanced when the two women dislike one another. Abelson (1983) suggested that it makes more sense to think about balance in terms of the implications of various systems of relations for actual social interactions rather than as a formal model. Given what we know of friendship, if two of the three possible dyadic relationships are positive, indicating the individuals get along, then we expect that the other pair of individuals should also get along. However, contrary to friendship, romantic love is typically considered an exclusive relationship. Thus, two different women typically cannot love and have a romantic relationship with the same man at the same time. Trying to do so leads to a major goal conflict and resulting antagonism. Thus, given what we know about romantic love, a triangle in which the two women dislike one another would in some sense be more balanced than one in which two women competing for the same man like one another. Thus, we believe that whether a structure is balanced or not should be viewed in terms of the underlying conceptual structure, and not in purely formal terms. We showed (Read & L. C. Miller, 1994) that by taking into account the underlying conceptual structure, the dynamics of both the friendship triad and the romantic triangle can be simulated by a parallel constraint satisfaction process.

Thus far, most of our discussion has focused on the processes affecting how individuals construe others and form new and changing representations. However, once such representations activate goals for subsequent action, how does that lead to the activation and enactment of behavioral sequences? How can we think of social behavior more dynamically? (See also L. C. Miller, Bettencourt, DeBro, & Hoffman, 1993.) It is to this issue that we now turn.

Goal-Directed Behavior in Social Interaction

Social psychologists have long been interested in understanding goal-directed behavior. Those influenced by Gestalt psychology proposed that this process could be understood in terms of interacting fields of forces. For example, Asch (1952), Lewin (1935), and Krech and Crutchfield (1948) viewed social behavior by individuals, dyads, and groups as the result of interacting force fields, where both social actors and aspects of their environment (e.g., goal objects or other people) possessed force fields and behavior was the result of the interaction of such fields. Lewin's topological psychology was an attempt to develop a mathematical formalism for thinking about the relation among such psychological force fields. Although it is commonly agreed that topological psychology was a failure as a formal system, it may now be possible to place the intuitions of Lewin and other early social psychologists on a firm theoretical and mathematical basis as characteristics of parallel constraint satisfaction processes.

In the following we summarize and discuss two models of goal-directed behavior that are based on parallel constraint satisfaction processes. These models speak to a number of issues that were central to Gestalt models of social behavior. First, as parallel constraint satisfaction models, they provide possible computational implementations of the Gestalt claim that social interaction can be conceptualized in terms of interacting force fields. Second, they provide an online mechanism for the relatively rapid integration of the myriad kinds of information that are available in social interaction. For instance, they provide a way to integrate the mutual influence of person and situation that was so central to Lewin's theories. Third, they provide a mechanism that describes how the organism could mediate among the influence of multiple, salient, often conflicting goals and do so in a way that results in reasonable behavior that is sensitive both to the desires of the individual and the opportunities and constraints of the environment.

Models of goal-directed behavior based on parallel constraint satisfaction processes may also overcome some of the failings of standard models of planning. Researchers in artificial intelligence and cognitive science (e.g., G. A. Miller, Galanter, & Pribram, 1960; Newell & Simon, 1972; Wilensky, 1983) have developed quite detailed models of planning behavior in a number of domains. Also, recently researchers have begun to apply such models to social interaction and conversation (e.g., L. C. Miller & Read, 1987; Read & L. C. Miller, 1989). Unfortunately, as researchers have noted (e.g., Maes, 1990, 1991), these models seem ill-suited to the oftentimes rapidly changing social environment because they assume a highly serial and quite deliberative process. The image portrayed by the typical planning model is one in which an individual is working in a relatively stable environment and the planner has the time to develop a quite detailed plan with all the steps laid out. Unfortunately, individuals engaged in social interaction lack this luxury; they confront a highly fluid environment in the form of the changing behaviors of others and a possibly changing physical environment. Faced with such an environment, standard planners would either "break" or fail to keep up, being unable to integrate the myriads of available information in such a short time frame. For example, standard planners would be unable to coordinate the interaction between two people on a date.

In contrast, models based on the parallel processing of information should be better able to handle the demands of social interaction. They provide the mechanism necessary for the integration of a wide range of information in a narrow time frame. In the following, we examine two models that view goal-directed behavior as a parallel constraint satisfaction process.

Mannes and Kintsch's construction-integration model of planning. Mannes and Kintsch (1991) applied Kintsch's (1988) construction-integration model of text comprehension to relatively routine planning tasks, where planning is very much like a comprehension task. These are tasks for which an individual is somewhat expert in the domain, but the tasks are not totally routinized—something true of much of everyday social interaction. Mannes and Kintsch used as their domain routine computing tasks such as file manipulation (e.g., attaching a file to an E-mail message and then sending the message to the addressee). Based on a detailed analysis of how individuals dealt with these tasks, their model provided an explicit computational simulation of participants' planning behavior.

In this model, an initial representation of the plan is constructed using a set of crude and imprecise production rules which promiscuously activate a wide range of information that may or may not be relevant to the final representation. This initial step is data driven and fairly automatic, and results in a rough initial representation. The initial representation is then integrated into a coherent plan using a parallel constraint satisfaction process. Actions that fit together support each other and suppress inconsistent elements. As Kintsch (1988) pointed out, one of the advantages of separating the activation and integration of information—and assuming that initially activated information is integrated by a separate process—is that it eliminates the need to assume unrealistically precise rules that initially activate only the correct information.

Mannes and Kintsch's (1991) NETWORK program uses the initial task description to retrieve information from long-term memory to create a task description network. Long-term memory represents what the individual knows about the particular domain, such as the available plan elements. Each plan element has information about both its preconditions and its consequences, if executed. All elements in long-term memory are also linked in terms of their semantic and associative relations.

The resulting task description network includes the original request, the outcome that request would have if it was successfully carried out (the goal), various plan elements that are related to the task description, and associated elaborations that are probabilistically retrieved based on associations to elements of the task description. Plan elements are retrieved by binding variables of the element to objects in the task description. That is, if a file name is specified in the task description, it will retrieve any plan element that has a variable for file name.

The links among the propositions in this network are then determined on the basis of goal–plan relations and semantic associations. First, the desired outcome or goal is positively linked to all plan elements that would achieve that outcome and negatively linked to all plan elements that would bring about an outcome inconsistent with the desired outcome. Second, plan elements have positive links to other plan elements that achieve their preconditions and negative links to plan elements that would destroy their preconditions. (This creates a degree of causal chaining among related plan elements.)

In addition, propositions are linked if they are associatively or semantically related, and the original request is linked to all plan elements with the same name. Finally, to prevent the execution of redundant plan elements, if all of the outcomes of a plan element already exist in the world, there is a negative link from the outcomes to the plan element.

This network is then integrated by spreading activation so that relevant items, which are supported by other items in the network, receive increased activation, and irrelevant or contradictory items are deactivated. Activation is spread through the task description network until the activation levels asymptote (see Mannes & Kintsch, 1991, for operational details). Once the activations have reached a stable value, the system executes the most highly activated plan element from among those whose preconditions are all satisfied.

Once a plan element is executed, its consequences are added to the task network. If the consequences match the desired outcome, then the planning process stops. If the consequences do not match, then this cycle is repeated, adding a new plan element to the sequence on each cycle, until the desired outcome is achieved. Newly added consequences inhibit both the plan that just produced them and any other plan that would produce them.

Thagard and Millgram's DECO model. Thagard and Millgram (1995) presented a model of more deliberative decision making and planning. They focused on how decision makers decide among alternative plans that are already known and available rather than on how the components of a plan are assembled. In their model the decision maker already has several alternative plans and must decide which of those plans to pursue.

Further, they focused on situations where the decision maker has multiple goals—some compatible and some incompatible—and must somehow trade off or balance these goals in choosing among alternatives. For instance, they gave the example of a college professor who must decide whether to accept a job offer or stay at his current institution. Each choice is related to multiple goals and may enable or block subsequent actions. Thus, a major focus of this model is on how decision makers weigh and balance multiple, potentially conflicting goals—a concern that was central in the work of Lewin.

In contrast, Mannes and Kintsch (1991) focused on situations where there is one clear goal and the problem is how to assemble, online, a plan that would allow for the achievement of that goal. Thus, the two models focus on somewhat different kinds of planning and on different parts of the planning process.

Thagard and Millgram (1995) described their model as involving "inference to the best plan":

> When people make decisions, they do not simply choose an action to perform, but rather adopt complex plans on the basis of a holistic assessment of various competing actions and goals. Choosing a plan is in part a matter of evaluating goals as well as actions. Choice is made by arriving at a plan or plans that involves actions and goals that are coherent with other actions and goals to which one is committed. (p. 440)

Thagard and Millgram (1995) suggested that decision makers evaluate the desirability of alternatives by evaluating the coherence of a set of goals and actions and the relations among them. The plans and goals that best cohere will be chosen as the most desirable. In this model, goals and actions are represented as nodes in a network, with the links between nodes representing the extent to which these actions and goals facilitate each other or are incompatible with each other. There is no strong distinction made between goals and actions, as a given action can serve both as the goal of a preceding action and a subgoal that achieves a subsequent goal.

Actions that facilitate other actions or goals have excitatory links, thus representing goal–subgoal relations. Conversely, actions that are incompatible with other actions or goals have inhibitory links. For example, two actions that require the same limited resource would presumably have inhibitory links, or an action that undoes a condition necessary for the execution of another action would have an inhibitory link to that action. The strength of these inhibitory links varies depending on the degree of incompatibility.

Further, sets of actions that are jointly required to achieve a goal have excitatory links to each other, encouraging them to be treated as a set. However, as the number of actions necessary to achieve a goal increases, the strength of the excitatory links to the goal from each individual action decreases. This implements a preference for simpler plans.

Thagard and Millgram (1995) also made the intuitively plausible claim that goals can differ in priority, with some goals being intrinsically more desirable. This is represented in their model by linking intrinsically desirable or important goals to a special goal unit that can send activation to the goal, with the strength of the link representing the importance of the goal. Goals that are linked to the special goal unit start out with more activation and, as a result, will tend to have more influence on other nodes and will be more resistant to influence from other nodes. In addition to the importance of the relations among goals and actions, they argue that the facilitative and competitive relations may often depend on the coherence of the goals and actions with factual beliefs, which indicate the degree of facilitation or inhibition that is believed to be the case.

The coherence of the network of goals and actions is evaluated by passing activation among the nodes in parallel, using the algorithm from Thagard's (1989, 1992) ECHO model. Once the activations asymptote, the decision maker is predicted to choose the set of actions and goals that are most coherent and have the highest levels of activation. Actions and goals with high levels of activation are part of the plan to be performed. Note that in this model the importance of goals (indicated by their final activation) can change, indicating that decision makers may revise the importance of goals.

Similarities between the models of goal-directed behavior. Despite some important differences, these two models share important similarities. First, in both models, the activation of an action increases if some or all of its preconditions are satisfied. Second, the activation of an action increases if it can help achieve some currently active goal. Third, the activation of an action can decrease if it undoes the conditions necessary for another action.

One potentially important difference between Mannes and Kintsch's (1991) model and Thagard and Millgram's (1995) model is that in Mannes and Kintsch's, the activation of an action is not the only criterion for whether it is chosen. It is also necessary that all the preconditions of the action be satisfied. The most highly activated action will still not be executed if some of its preconditions remain unsatisfied. In contrast, Thagard and Millgram seem to suggest that an individual could choose a plan even if some of its preconditions were unsatisfied.

Summary and Conclusion

As we have proceeded through the article, we discussed many of the advantages of this approach. We have shown how parallel constraint satisfaction processes can be applied to three broad areas in social psychology that were central in the early development of our field and that remain at its heart today: impression formation and causal attribution, cognitive consistency, and goal-directed behavior.

In discussing impression formation and causal attribution, we did a number of things. First, we used parallel constraint satisfaction processes to provide an explicit computational model of how social concepts—such as traits, roles, and situations—can be combined to create novel social concepts as part of creating an integrated impression of an individual. Second, we outlined a model for the online construction of mental representations of social interactions, providing an explicit account of how observers of and participants in social interaction can create a model of the goals, plans, and characteristics of social actors. Third, as part of this model, we indicated how a parallel process can account for phenomena in dispositional inference that Gilbert and his colleagues (Gilbert, 1989; Gilbert, Pelham, & Krull, 1988; Krull, 1993) have argued quite strongly are evidence for a controlled serial process. Fourth, we demonstrated how several Gestalt principles of causal unit formation, such as salience and proximity, can be given a natural interpretation as part of a parallel constraint satisfaction model. Finally, we showed how parallel constraint satisfaction processes can provide an explicit computational model of how the wholistic processing of social information could be achieved.

Our examination of cognitive consistency demonstrated how parallel constraint satisfaction systems can be used to evaluate the consistency and structure of belief systems. How to do this is a question with a long history in social psychology, but one that has lacked an adequate mechanism. As part of this analysis, we discussed how parallel constraint satisfaction systems can provide a computational account of two classic models of cognitive consistency: balance theory and cognitive dissonance theory.

Finally, in our discussion of goal-directed behavior, we elucidated how parallel constraint satisfaction systems can realize goal-directed behavior as the result of the simultaneous solution of a large number of constraints involving characteristics of the actor, other people, the situation, and so forth. Further, Mannes and Kintsch's (1991) work demonstrates how such a system can provide one possible model of how social interaction can be constructed in real-time. Thagard and Milgram's DECO model provides an account of how people can solve for a set of constraints among a set of interacting goals and the actions that might achieve them. Thus, parallel constraint satisfaction processes, in conjunction with other relevant theoretical assumptions, illuminate a number of central issues in social psychology.

It is interesting that in each of these areas we found ourselves relying on principles and insights that are remarkably similar to those driving the classic work in these areas. And each of these three broad areas can be understood in terms of the operation of the same kind of process. Rather than having to specify quite different models for each domain, we can view the phenomena within each domain as different manifestations of a common underlying mechanism. Aside from benefits of parsimony and generality, the common framework makes it easier to see how these and other phenomena can be integrated.

In addition, from the perspective of this approach, phenomena that seem unexpected or counterintuitive are revealed as being due to the normal functioning of the cognitive system. For example, Shultz and Lepper (1996) noted that in thinking about cognitive dissonance in terms of parallel constraint satisfaction processes, it becomes clear that cognitive consistency phenomena—such as those studied by dissonance researchers—are not the result of atypical or unusual cognitive processes but rather are the direct result of normal cognitive functioning.

Toward a More Dynamic Social Psychology

Many social psychologists undoubtedly see the application of parallel distributed processing models to social reasoning and behavior as the importation of foreign, incomprehensible concepts. We hope that we have shown this is far from the case. Instead, concepts similar to many of those from connectionism and parallel distributed processing models have deep roots in our field. If Asch, Lewin, and Heider were working today, they might well be investigating the applications of parallel constraint satisfaction processes to social reasoning and behavior.

Consider the promise of parallel constraint satisfaction processes for addressing the dynamics of social behavior and meaning. Since Asch (1946), social psychologists have argued the question of whether the construction of social meaning is wholistic and more than the sum of its parts, or whether it can be adequately modeled by some kind of linear rule. The advocates of a linear model always had the advantage of an explicit computational rule. They no longer have this advantage. Parallel constraint satisfaction processes provide an explicit mathematical model of wholistic processing and make plain how the whole can truly be more than the sum of its parts. We have waited a long time in this field for tools that make an explicit model of wholistic processing possible. Moreover, it seems within our reach to better understand how to integrate intertwined information about persons, situations, emotions, and relationships—while allowing for the fluidity of social meaning and behavior.

And consider the complexity of social interaction: Parallel constraint satisfaction processes make understanding and exploring the richness of social dynamics possible. As we argued at several points, the successful coordination of social interaction requires the integration of a considerable amount of information within a small window of time. We must simultaneously perceive the other person, figure out what he or she is doing, and plan our own behavior—all in a brief moment. This clearly requires a high degree of parallel processing. Parallel constraint satisfaction processes provide one account of how this might be done.

Finally, parallel constraint satisfaction processes make possible the rebirth of cognitive dynamics. The founders of modern social psychology were intensely interested in cognitive dynamics, the mechanisms underlying the ebb and flow of thought. Although interest in such topics has largely disappeared, there are signs of renewed interest in the dynamics of thought and meaning. In several different articles (L. C. Miller et al., 1993; L. C. Miller & Read, 1991; Read & L. C. Miller, 1989), we have discussed how to integrate the dynamics of thought and behavior. And there is considerable other movement afoot suggesting a new zeitgeist and the emergence of a truly dynamic paradigm (e.g., Eiser, 1994; L. C. Miller et al., 1993; Nowak, Szamrej, & Latané, 1990; Vallacher & Nowak, 1994; Vallacher, Nowak, & Kaufman, 1994). Connectionism holds great promise for social psychology, allowing us to integrate classic concerns with contemporary questions, and weaving a rich tapestry from the often disparate threads of our discipline.

References

Abelson, R. P. (1968). Psychological implication. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg,

& P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook* (pp. 112–139). Chicago: Rand McNally.

Abelson, R. P. (1983). Whatever became of consistency theory? *Personality and Social Psychology Bulletin, 9,* 37–54.

Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J., & Tannenbaum, P. H. (Eds.). (1968). *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally.

Abelson, R. P., & Rosenberg, M. J. (1958). Symbolic psycho-logic: A model of attitudinal cognition. *Behavioral Science, 3,* 1–8.

Allport, F. H. (1924). *Social psychology.* Boston: Houghton-Mifflin.

Amit, D. J. (1989). *Modeling brain function: The world of attractor neural networks.* Cambridge, England: Cambridge University Press.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39,* 1037–1049.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84,* 413–451.

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review, 85,* 249–277.

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic.

Aronson, E. (1968). Dissonance theory: Progress and problems. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook* (pp. 5–27). Chicago: Rand McNally.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41,* 258–290.

Asch, S. E. (1952). *Social psychology.* Englewood Cliffs, NJ: Prentice-Hall.

Asch, S. E., & Zukier, H. (1984). Thinking about persons. *Journal of Personality and Social Psychology, 46,* 1230–1240.

Barsalou, L. W. (1992). *Cognitive psychology: An overview for cognitive scientists.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Cantor, N., Mischel, W., & Schwartz, J. (1982). A prototype analysis of psychological situations. *Cognitive Psychology, 14,* 45–77.

Cartwright, D., & Harary, F. (1956). Structural balance: A generalization of Heider's theory. *Psychological Review, 63,* 277–293.

Deutsch, M. (1968). Field theory in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (pp. 412–487). Reading, MA: Addison-Wesley.

Eiser, J. R. (1994). *Attitudes, chaos, and the connectionist mind.* Oxford, England: Blackwell.

Festinger, L. (1950). Informal social communication. *Psychological Review, 57,* 271–282.

Festinger, L. (1957). *A theory of cognitive dissonance.* Evanston, IL: Row, Peterson.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology, 58,* 203–210.

Gabrys, G. (1989). Heider: A simulation of attitude consistency and attitude change. In S. Ohlsson (Ed.), *Aspects of cognitive conflict and cognitive change* (Tech. Rep. No. KUL–89–04). Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality, 54,* 733–740.

Graesser, A. C., & Clark, L. F. (1985). *Structures and procedures of implicit knowledge.* Norwood, NJ: Ablex.

Hamilton, D. L., & Zanna, M. P. (1972). Differential weighting of favorable and unfavorable attributions in impressions of personality. *Journal of Experimental Research in Personality, 6,* 204–212.

Hastie, R., Schroeder, C., & Weber, R. (1990). Creating complex social conjunction categories from simple categories. *Bulletin of the Psychonomic Society, 28,* 242–247.

Heider, F. (1944). Social perception and phenomenal causality. *Psychological Review, 51,* 358–374.

Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology, 21,* 107–112.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

Henle, M. (1986). *1879 and all that: Essays in the theory and history of psychology.* New York: Columbia University Press.

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

Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 282–317). Cambridge, MA: MIT Press/Bradford.

Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science, 13,* 295–355.

Holyoak, K. J., & Spellman, B. A. (1993). Thinking. *Annual Review of Psychology, 44,* 265–315.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA, 79,* 2554–2558.

Hopfield, J. J. (1984). Neurons with graded responses have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences, USA, 81,* 3088–3092.

John, O. P. (1986). How shall a trait be called: A feature analysis of altruism. In A. Angleitner, A. Furnham, & G. Van Heck (Eds.), *Personality psychology in Europe: Current trends and controversies* (pp. 117–140). Berwyn, IL: Swets North America.

Jordan, N. (1968). Cognitive balance as an aspect of Heider's cognitive psychology. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook* (pp. 169–178). Chicago: Rand McNally.

Kashima, Y., & Kerekes, A. R. Z. (1994). A distributed memory model of averaging phenomena in person impression formation. *Journal of Experimental Social Psychology, 30,* 407–455.

Kemler-Nelson, D. G. (1984). The effect of intention on what concepts are acquired. *Journal of Verbal Learning and Verbal Behavior, 23,* 734–759.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review, 95,* 163–182.

Koffka, K. (1935). *Principles of Gestalt psychology.* New York: Harcourt, Brace.

Köhler, W. (1920). *Die physischen Gestalten in Ruhe and im stationären Zustand.* Braunschweig, Germany: F. Vieweg.

Köhler, W. (1929). *Gestalt psychology.* New York: Liveright.

Krech, D., & Crutchfield, R. S. (1948). *Theory and problems of social psychology.* New York: McGraw-Hill.

Krull, D. S. (1993). Does the grist change the mill? The effects of perceiver's inferential goal on the process of social inference. *Personality and Social Psychology, 19,* 340–348.

Kunda, Z., Miller, D. T., & Claire, T. (1990). Combining social concepts: The role of causal reasoning. *Cognitive Science, 14,* 551–577.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel constraint satisfaction theory. *Psychological Review, 103,* 284–308.

Lewin, K. (1935). *A dynamic theory of personality.* New York: McGraw-Hill.

Lewin, K. (1947a). Frontiers in group dynamics: I. *Human Relations, 1,* 2–38.

Lewin, K. (1947b). Frontiers in group dynamics: II. *Human Relations, 1,* 143–153.

Lutz, C. (1988). *Unnatural emotions: Everyday sentiments on a Micronesian atoll and their challenge to Western theory.* Chicago: University of Chicago Press.

Maes, P. (1990). Situated agents can have goals. *Journal of Robotics and Autonomous Systems, 6,* 49–70.

Maes, P. (1991). Adaptive action selection. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 108–113). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Mannes, S. M., & Kintsch, W. (1991). Routine computing tasks: Planning as understanding. *Cognitive Science, 15,* 305–342.

Markus, H. R., & Kitayama, S. (1994). The cultural construction of self and emotion: Implications for social behavior. In S. Kitayama & H. R. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 133–196). Washington, DC: American Psychological Association.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.

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/Bradford.

Medin, D. L., & Schwanenflugel, P. L. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 355–368.

Medin, D. L., & Wattenmaker, W. D. (1987). Category cohesiveness, theories, and cognitive archeology. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 25–62). Cambridge, England: Cambridge University Press.

Miikkulainen, R. (1993). *Subsymbolic natural language processing: An integrated model of scripts, lexicon, and memory.* Cambridge, MA: MIT Press/Bradford.

Miller, G. A., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehart & Winston.

Miller, L. C., Bettencourt, B. A., DeBro, S. C., & Hoffman, V. (1993). Negotiating safer sex: A dynamic interpersonal process. In J. Pryor & G. Reeder (Eds.), *The social psychology of HIV infection* (pp. 85–123). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Miller, L. C., & Burns, D. (1992, May). *Why didn't you use a condom? On the coherence of gay men's explanations for unsafe sexual practices.* Paper presented at the meeting of the International Communication Association, Miami, FL.

Miller, L. C., Cody, M. J., & McLaughlin, M. L. (1994). Situations and goals as fundamental constructs in interpersonal communication research. In M. L. Knapp & G. R. Miller (Eds.), *Handbook of interpersonal communication* (pp. 162–198). Thousand Oaks, CA: Sage.

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: Plenum.

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

Minsky, M., & Papert, S. (1969). *Perceptrons.* Cambridge, MA: MIT Press.

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

Murre, J. M. J. (1992). *Learning and categorization in modular neural networks.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nowak, A., Szamrej, J., & Latané, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review, 97,* 362–376.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology, 51,* 242–258.

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 521–533.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62,* 189–206.

Read, S. J. (1987). Constructing causal scenarios: A knowledge structure approach to causal reasoning. *Journal of Personality and Social Psychology, 52,* 288–302.

Read, S. J., & Lincer, H. (1994). *Explanatory coherence and dispositional inference: The role of parallel constraint satisfaction processes.* Unpublished manuscript, University of Southern California, Los Angeles.

Read, S. J., & Marcus-Newhall, A. R. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology, 65,* 429–447.

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: Lawrence Erlbaum Associates, Inc.

Read, S. J., & Miller, L. C. (1993). Rapist or "regular guy": Explanatory coherence in the construction of mental models of others. *Personality and Social Psychology Bulletin, 19,* 526–540.

Read, S. J., & Miller, L. C. (1994). Dissonance and balance in belief systems: The promise of parallel constraint satisfaction processes and connectionist modeling approaches. In R. C. Schank & E. J. Langer (Eds.), *Beliefs, reasoning, and decision making: Psycho-logic in honor of Bob Abelson* (pp. 209–235). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Roseman, I. J., Spindel, J. S., & Jose, P E. (1990). Appraisals of emotion-eliciting events: Testing a theory of discrete emotions. *Journal of Personality and Social Psychology, 59,* 899–915.

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attribution processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32,* 880–892.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60–94.

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

Rumelhart, D. E., & McClelland, J. L. (1986b). 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. Foundations* (pp. 110–146). Cambridge, MA: MIT Press/Bradford.

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

Seiter, J. (1993). *A coherence analysis of deception detection.* Unpublished doctoral dissertation, University of Southern California, Los Angeles.

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology, 52,* 1061–1086.

Shultz, T. R., & Lepper, M. R. (1992). A constraint satisfaction model of cognitive dissonance phenomena. In *The Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 462–467). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Shultz, T. R., & Lepper, M. R. (1996). Cognitive dissonance reduction as constraint satisfaction. *Psychological Review, 103,* 219–240.

Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 194–281). Cambridge, MA: MIT Press/Bradford Books.

Spellman, B. A., & Holyoak, K. J. (1992). If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles. *Journal of Personality and Social Psychology, 62,* 913–933.

Spellman, B. A., Ullman, J. B., & Holyoak, K. J. (1993). A coherence model of cognitive consistency. *Journal of Social Issues, 4,* 147–165.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12,* 435–467.

Thagard, P. (1992). *Conceptual revolutions.* Princeton, NJ: Princeton University Press.

Thagard, P., & Millgram, E. (1995). Inference to the best plan: A coherence theory of decision. In A. Ram & D. B. Leake (Eds.), *Goal-driven learning (pp. 439–454).* Cambridge, MA: MIT Press.

Vallacher, R. R., & Nowak, A. J. (1994). *Dynamical systems in social psychology.* San Diego, CA: Academic.

Vallacher, R. R., Nowak, A., & Kaufman, J. (1994). Intrinsic dynamics of social judgment. *Journal of Personality and Social Psychology, 67,* 20–34.

Wattenmaker, W. D., Dewey, G. I., Murphy, T. D., & Medin, D. L. (1986). Linear separability and context learning: Context, relational properties, and concept naturalness. *Cognitive Psychology, 18,* 158–194.

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie, 61,* 161–265.

Wierzbicka, A. (1994). Emotion, language, and cultural scripts. In S. Kitayama & H. R. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 133–196). Washington, DC: American Psychological Association.

Wilensky, R. (1983). *Planning and understanding: A computational approach to human reasoning.* Reading, MA: Addison-Wesley.

Wyer, R. S., Jr. (1974). Changes in meaning and halo effects in personality impression formation. *Journal of Personality and Social Psychology, 29,* 829–835.

Zanna, M. P., & Hamilton, D. L. (1977). Further evidence for meaning change in impression formation. *Journal of Experimental Social Psychology, 13,* 224–238.