Gestalt Principles and Parallel Constraint Satisfaction Processes: The Parallels

Eric J. Vanman
Department of Psychology
University of Southern California
Los Angeles, CA 90089-1061
vanman@usc.edu

Stephen J. Read
Department of Psychology
University of Southern California
Los Angeles, CA 90089-1061
read@almaak.usc.edu

Lynn C. Miller
CAAS
University of Southern California
Los Angeles, CA 90089-1694
lmiller@uscvm

Abstract

This paper examines the tremendous similarities between the Parallel Constraint Satisfaction Processes that are a central part of many connectionist models and the Gestalt principles that played a central role in the history of Psychology. Gestalt Psychology played a major role in a number of areas in psychology, such as perception, reasoning and problem solving, causal reasoning, and many key aspects of social psychology, such as social perception, group interaction, and belief consistency. Many of the key assumptions of Gestalt Psychology have resurfaced in recent connectionist models. We propose that Parallel Constraint Satisfaction Processes provide a computational implementation of many of the central principles of Gestalt Psychology. In this paper we discuss the clear parallels between each of five key assumptions of Gestalt Psychology and aspects of Parallel Constraint Satisfaction Processes. The five assumptions we examine are: (1) psychological processing can be treated as interactions in fields of forces, (2) psychological processing is holistic, (3) the whole is greater than the sum its parts, (4) the importance of the structure of cognitive elements; how things are connected and related, and (5) the emphasis on cognitive dynamics, and such concepts as change, equilibrium, and tension.

Introduction

The historical contributions of gestalt psychology to the foundations of social psychology, cognitive psychology, and especially, perception, are well known. Koffka, Köhler, and Wertheimer influenced psychology not only directly through their seminal theory and research, but also through the training of others who made significant contributions to different areas of psychology. For example, the names of Asch, Lewin, Heider, and Festinger are well known to the social psychologist when discussing such concepts as dissonance, balance, and person perception. Gestalt principles were central to each. Yet, despite their historical and theoretical importance, Gestalt principles are largely absent from most current psychological theorizing. Why? Gestalt processes stressed holistic 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 interaction implied by basic Gestalt principles. Individuals 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 a narrow time. But, capturing such dynamics may have seemed beyond the reach of the empirical and theoretical tools of the day. However, recent work in connectionism, specifically parallel constraint satisfaction models, suggests that seemingly metaphorical Gestalt processes can be given a concrete implementation (Spellman & Holyoak, 1992).

In this paper, we first briefly discuss the general concept of parallel constraint satisfaction processes. We follow this with an analysis of the commonalties between these processes and the various Gestalt concepts. The main point of this paper is to demonstrate that many of the issues and insights addressed by connectionist models have a long history in psychology, going back at least to the early 1900s in the work of the Gestalt psychologists.

Parallel Constraint Satisfaction Processes

In many connectionist models processing can be viewed as a parallel constraint satisfaction *process*, where activation is passed around the nodes in the network until all the node activations asymptote or "relax" into a state that satisfies the constraints among the nodes (e.g., Hertz, Krogh, & Palmer, 1991; Murre, 1992; Rumelhart & McClelland, 1986). A parallel constraint satisfaction network consists of a set of nodes and the links among them, where the nodes represent hypotheses about the presence or absence of various features, and positive and negative links represent the extent to which the hypotheses are consistent or inconsistent with one another. The weights on the links indicate the strength of the consistency or inconsistency between the nodes or hypotheses.

Thus, the links represent a set of constraints among the hypotheses. Activation is spread among all the nodes until the activation of each node asymptotes and the network "settles". Because the activation of a node is a result of all of its positive and negative links to other nodes, its final activation is a solution to the constraints represented by the links. Moreover, because activation is spread in parallel, this process results in a global solution to the constraints among the entire set of nodes.

Hopfield (1982, 1984; also see Hertz et al., 1991) has shown that such a system can be treated as if it has an energy, where the energy of the system is a function of the activations of the nodes and the weights among them. Moreover, evaluating a parallel constraint satisfaction network with symmetric links can be viewed as an attempt to minimize the system's energy. Solving for the constraints is a gradient descent process, moving toward a minimum (or valley) in an energy surface that represents all the possible states of the system (Hertz et al., 1991). 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 for the *entire system.* Such networks are dynamic systems, as their state evolves over time. Essentially, the energy of the system corresponds to its degree of organization. High energy 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.

The Relation of Gestalt Principles of Parallel Constraint Satisfaction Processes

We turn now to an examination of some of the basic assumptions of Gestalt psychology and their close parallels with the characteristics of parallel constraint satisfaction systems. 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 holistic 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, and (5) that the psychological field is a dynamic system.

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 behavior's causal relation to these fields, and to identify and understand the forces that caused behavior to occur.

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, Hertz et al. (1991), following work by others, point 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 between individual atoms in this magnetic material correspond to the patterns of influence among neurons in a Hopfield network. Further, researchers have applied a host of ideas from statistical mechanics and thermodynamics to the analysis of neural network models (e.g., Hertz et al., 1991; Hinton & Sejnowski, 1986). Therefore, 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. Thus, the intuitions of the Gestalt psychologists may have been on the right track and their attempts to analyze psychological phenomena in terms of interacting psychological and social fields of forces, may not have been misdirected.

Psychological Processing is Holistic

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 writes "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.

Holistic 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. Asch proposed that we perceive other individuals as whole units. Like one of the visual illusions studied by Gestalt psychologists, the perception of personality traits is holistic. "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 holistic, 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.

Holistic processing of information is precisely what happens in neural network models. Items simultaneously send and receive activation to and 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. 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) demonstrated 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.

One problem Gestalt Psychology always had was that as much as this idea seemed to fit 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 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.

One reasons why earlier work on neural networks largely stopped is because Minsky and Papert (1969), in their critique of one kind of neural network, the perceptron, demonstrated that these early networks could only handle linear problems. Yet, researchers recognized that psychological processes often 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, 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 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 a 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 is in work on human categorization. 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. Further, linearly separable categories are no easier to learn than are non linearly separable categories (Medin & Schwanenflugel, 1981). Medin and Wattenmaker (1987) argue 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."

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 not only applied to spatial relations, but 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 influenced the creation of causal attributions. Later (Heider, 1946), he extended this analysis in balance theory. 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 had to know how elements were organized, what was related to what, and how they were related. One could not just sum up all the elements, one had to know how they were organized. 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. The activation (and thus the interpretation) of the elements in the network critically depends on the nature of the connections among the elements. 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, 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).

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 of a system under tension that tends to evolve toward a state of minimal tension is remarkably similar to a parallel constraint satisfaction 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 as an attempt to minimize the degree of tension or conflict in the system. 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.

In addition, neural network models can be explicitly characterized as dynamic systems where the state of the system changes over time. For example, one can look at how, following initial input, the system evolves over time to an increasing degree of organization, and one can 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 examine how the system evolves to a new state. Further, certain kinds of 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, where the minimums in the energy surface are attractors toward which the state of the system tends or is "pulled" (Hertz et al., 1991).

Summary

We have outlined how some of the key ideas and insights of recent work in connectionism have a long history in psychology that can be traced back to the founders of Gestalt psychology. In fact, Gestalt psychologists introduced these ideas in a number of areas in psychology. And many of these ideas have periodically resurfaced through the years, with parallel constraint satisfaction models providing the latest instance. However, one major advantage of current parallel constraint satisfaction models is that they allow us to develop explicit process models that provide computational implementations of many of these recurring insights, as well as allowing us to push our investigations far beyond the bounds of this earlier work.

References

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

Deutsch, M. (1968). Field theory in social psychology. In G. Lindzey & E. Aronson (Eds.), *The Handbook of Social Psychology* (2nd Ed.), (Vol. 1, pp. 412-487). Reading, MA: Addison-Wesley.

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.

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

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

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.

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

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

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

Medin, D. L., & Schwanenflugel, P. L. (1981). Linear separability in classification learning. *Journal of Experiemental 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 University Press: Cambridge, England.

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

Murre, J. M. J. (1992*). Learning and Categorization in Modular Neural Networks*. Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., & McClelland, J. L. (1986). PDP models and general issues in Cognitive Science. In Rumelhart, D. E., & McClelland, J. L. (Eds). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1. Foundations*. (pp. 110-146). 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.

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie*, 61, 161-265. (Translation in T. Shipley (Ed.*), Classics in psychology*. New York: Philosophical Libray, 1961.)