9 Fundamental Aspects of Cognitive Representation

Stephen E. Palmer
University of California, Berkeley

This chapter was born of an ill-defined but definite feeling that we, as cognitive psychologists, do not really understand our concepts of representation. We propose them, talk about them, argue about them, and try to obtain evidence in support of them, but we do not understand them in any fundamental sense. Anyone who has attempted to read the literature related to cognitive representation quickly becomes confused – and with good reason. The field is obtuse, poorly defined, and embarrassingly disorganized. Among the most popular terms, one finds the following: visual codes, verbal codes, spatial codes, physical codes, name codes, image codes, analog representations, digital representations, propositional representations, first-order isomorphisms, second-order isomorphisms, multidimensional spaces, templates, features, structural descriptions, relational networks, multicomponent vectors, and even holograms. This abundance of language for talking about representation would be a good thing if all the distinctions were clear and if they fit together in a systematic way. The fact is that they are not clear and do not fit together. Different people use the same term in different ways and different terms in the same way. These are not characteristics of a scientific field with a deep understanding of its problem, much less its solution. This chapter is an extended inquiry into the nature of the problem of cognitive representation. The rationale is that a solution is more likely to be achieved if the problem is understood properly.

In order to make systematic progress on problems concerning cognitive representation, we must begin at the beginning: *What is representation?* This is a question few psychologists have ever asked and even fewer have made any serious attempt to answer. It is so basic a question that one might wonder whether its answer would be of any value to cognitive psychology. It is the main thesis of this chapter that the answer is enormously important and will change our under-

standing of both theories and experiments concerned with cognitive representation.

Let us assume that our goal is to specify as clearly as possible the nature of people's internal representations of the world. That is, we want to construct clear, concise theories of cognitive representation that can be evaluated using the methods of psychology as a behavioral science. The standard method of evaluation is to ask questions by experimental hypothesis testing. The answers to these questions delineate the nature of an adequate theory. If this is a reasonable characterization of the situation, is there any reason to believe that knowing (or having a theory of) what representation is will help us in either the theoretical or experimental enterprise?

Trying to determine the nature of cognitive representation without first knowing about representation as a general construct is much like trying to determine the nature of oak trees without first knowing about trees as a general class of objects. Suppose there are two botanists whose task is to describe the essential characteristics of oak trees by performing a series of tests on a given specimen. Botanist A has a thoroughgoing knowledge of trees: what their defining characteristics are, how types of trees differ from one another in ways that are relevant to classification, and how they differ in ways that are irrelevant. Botanist B knows none of these things. Botanist A works quickly and efficiently. She makes the measurements necessary to describe oaks versus other types of trees and only those measurements. She does not bother to note that oaks have bark and leaves because she knows that all trees have these attributes. Nor does she bother to count the number of branches on this particular tree, because she knows this to be irrelevant to her task. When she is done, she describes oaks as, say, trees with properties a, b, and c. Botanist B, however, necessarily performs many more measurements than A. If he is diligent enough, he eventually discovers the defining characteristics of oaks, but these are mixed together with properties that are relevant only to treeness and with other properties that are relevant only to this particular specimen. When he is done, his description might be that oaks are objects with properties a, b, c, d, ... Obviously, A has an advantage over B in knowing something about the general nature of trees. She is able to make fewer empirical tests, and her description is simpler and more specific to the relevant factors.

Although the analogy is rough, the major point is clear. If representations, like trees, have certain defining characteristics, certain relevant dimensions of variation, and certain irrelevant dimensions of variation, then knowledge of these things (or at least some working hypotheses about them) should be important to psychologists for very tangible reasons. It defines the kind of experiments that are deemed important and relevant. It specifies the general form of our theoretical descriptions and separates essential aspects from nonessential ones. In fact, without some "metatheoretical" framework of this sort, it is not clear that the theoretical and experimental enterprises are meaningful at all. In short, the answer to the representational question provides the larger framework for research that Kuhn (1962) has called a "paradigm."

If this is so, then we must currently have something that serves the function of a representational metatheory. As far as I can tell, the present framework for representations is a loose system of distinctions and classifications. Representational theories are defined by descriptive terms like those mentioned earlier: templates, features, structural descriptions, and so forth. Therefore, psychologists do experiments that purport to test these alternatives. Such experiments abound in the psychological literature. The issue is whether these are sensible questions to ask in our experiments. Ultimately, this boils down to asking whether our current framework for representation is sensible.

There are several observations suggesting that it is not. First, the distinctions we make do not follow from or lead to any coherent view of representation in general. This is why the basic question – what is representation – is so hard to answer. Second, virtually none of the current distinctions have ever been explicitly defined. We can point to some samples of each concept, but this ability is something less than a proper definition. Part of the reason they have not been defined is that good definitions are much easier to construct within a larger framework. Third, the distinctions do not relate to each other systematically. Is the template/feature distinction independent of, the same as, or otherwise related to the analog/propositional distinction? How do templates relate to prototypes? Our inability to provide good answers to such questions is symptomatic of our understanding of the concepts themselves. Fourth – and perhaps most apparent to experimental psychologists – the empirical tests of such distinctions rarely, if ever, lead to conclusive results. There always seem to be ready explanations from the allegedly discredited theoretical position. This fact can be partially ascribed to complications such as processing differences, because both representation and processing assumptions are required to predict performance. But often I suspect the problem is that many of the distinctions we purport to test are not mutually exclusive at a level that is meaningful for our goals and methods.

As Kuhn (1962) has noted, a scientific discipline does not abandon a reigning paradigm simply because it is seen to be defective. Rather, another must emerge to take its place. The new paradigm should be able to make sense of things that were formerly puzzling and bring a more elegant and coherent view to the domain of the field. In the rest of this chapter, I propose a new view of cognitive representation based explicitly on an answer to the general representational question. This view is developed informally with an emphasis on noncognitive representation. It is then applied to the concepts of the older framework. As far as possible, definitions are provided for constructs like templates, features, structural descriptions, prototypes, isomorphisms, propositions, and analogs. Relationships among them are clarified, and relevant aspects are separated from irrelevant ones. The results show that many of the mistakes we have made in understanding representation are alarmingly fundamental.

REPRESENTATIONAL METATHEORY

Let us turn now to the basic question: What is representation? The first problem is how to attack this question without considering cognitive representation itself. The answer, of course, is to examine noncognitive forms of representation, either real or artificially constructed for this particular purpose. Cognitive representation is exceedingly complex and difficult to study. Other sorts of representations are simple and easy to study. The plan is to move from simple representations to complex ones so that the basic issues are clear from the outset.

Some Examples of Representation

A representation is, first and foremost, something that stands for something else. In other words, it is some sort of model of the thing (or things) it represents. This description implies the existence of two related but functionally separate worlds: the *represented world* and the *representing world.* The job of the representing world is to reflect some aspects of the represented world in some fashion. Not all aspects of the represented world need to be modeled; not all aspects of the representing world need to model an aspect of the represented world. However, there must be some corresponding aspects if one world is to represent the other. In order to specify a representation completely, then, one must state: (1) what the represented world is; (2) what the representing world is; (3) what aspects of the represented world are being modeled; (4) what aspects of the representing world are doing the modeling; and (5) what are the correspondences between the two worlds. A representation is really a *representational system* that includes all five aspects.

Figure 9.1 shows some simple examples of representational systems that illustrate the previous points. In all cases, the represented world is the set of four rectangles shown in Fig. 9.1A. These drawings, simple as they are, contain many aspects that could be modeled in a representation. The representing Worlds B, C, and D show how different aspects of the same represented world can be modeled by the same representing world. World B reflects the relative height of the rectangles (a, b, c, d) by the relative length of the corresponding lines (a', b', c', d'). In other words, the fact that a is taller than b in World A is reflected by the fact that a' is longer than b' in World B. Similar statements can be made for any pair of rectangles in World A and the corresponding lines in World B. It is always true that if x is taller than y in World A, then x' is longer than y' in World B. One could describe this representational system by saying, "World B is a representation of World A in which the relative length of lines in B corresponds to the relative height of rectangles in A." The implication is that any question that could be answered about relative height in A could be equally well answered by considering relative length in B as long as the mapping of rectangles to lines were known.

World C reflects the relative width of rectangles in A by the relative length of lines. For example, the fact that d is wider than any other rectangle in World A

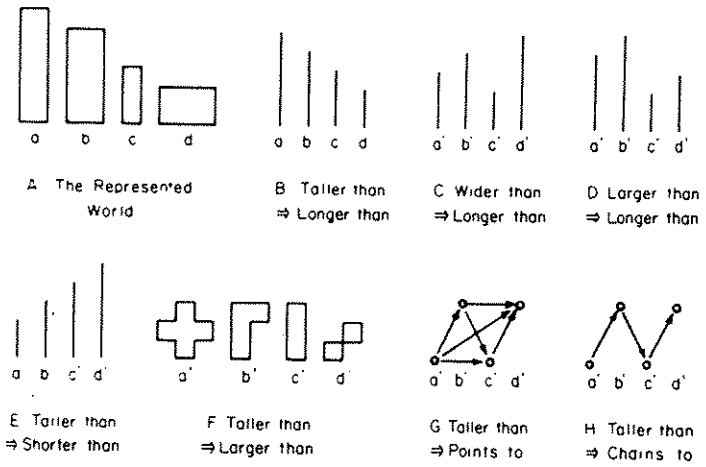

FIG. 9.1. Examples of representation. The represented world in each case consists of the objects shown in A. For each representing world, the correspondence of objects is indicated by the letters beneath them, and the correspondence of relations is indicated by the expression beneath each world.

is reflected in the fact that d' is longer than any other line in World C. World D performs the same kind of representational function for relative size of the rectangles in World A. These examples demonstrate that one cannot specify a representation simply by pointing to a representing world of objects. Without knowledge of the represented world, its modeled aspects, and the correspondence between the two worlds, representations B, C, and D are identical. Given this information, however, it is clear that they are quite different.

Worlds B, E, F, G, and H illustrate how the same aspect of a represented world can be modeled using different representing worlds. World B models height of rectangles in terms of line length; the taller the rectangle, the longer the line. World E also models rectangle height in terms of line length; but here, the taller the rectangle, the shorter the line. This representational system differs from that of B only in that a different relation ("shorter" rather than "longer") is used to model the represented relation ("taller").

World F reflects the height of rectangles by the size (area) of closed geometric forms. Note that object shape has no correlate in the represented world. In addition to using a different relation to represent "taller than," this example illustrates that there may be other aspects of the representing world that are irrelevant to its modeling function.

World G embodies a rather different way of preserving height relations among the rectangles in A. Here, the fact that a is taller than b in A is reflected in a' pointing to b' in G. What makes this kind of representation different from those discussed previously is that: (1) the representing objects corresponding to the

rectangles are identical; and (2) new elements (the arrows) have been added that correspond explicitly to the relation being modeled ("taller than"). It is important to notice that the arrows of World G are not "object elements" but "relational elements." That is, the presence of a given arrow in the representational world does not correspond to any single object in the represented world. It is tempting to characterize the difference between representations B and G by saying that B represents relations ("taller than") by relations ("longer than") while G represents relations ("taller than") by elements (the arrows). The real difference is more subtle, however. "Points to" (more accurately, "is arrow-connected to") in World G is a relation just as "longer than" is in World B. The difference is that although "longer than" is a relation that can hold between an ordered pair of objects, "is arrow-connected to" is a relation that can hold between an ordered pair of objects only by virtue of each being related to a third element (the relational arrow) in a particular way.

World H illustrates yet another type of representation. Like World G, it contains explicit relational elements, but the arrows are not sufficient to model the "taller than" relation in A. "Is arrow-connected to" models some other, more restricted relation that might be called "next-taller than." In order for "taller than" to be represented, the "points to" relation must be made transitive. The solution is to use the "chains to" relation (more accurately, "is arrow-path-connected to"), which is essentially a transitive version of "points to." It is easy to see that if all the "chains to" relations were made explicit in World H by arrows, the resulting representation would be identical to World G.

Intuitively, what all these representations have in common is that they contain information that reflects some information about the world they represent. The information contained by the representing worlds can be the same yet can reflect different information about the represented world. Worlds B, C, D, and E of Fig. 9.1 are examples of this situation. In contrast, the information contained by the representing worlds can be quite different, yet can reflect the same information about the represented world. Worlds B, E, F, G, and H are examples of this possibility. No two representational systems in Fig. 9.1 are exactly the same, but some are more similar than others. We examine these similarities and differences in more detail shortly.

Operational Relations

Thus far a representing world has been treated as a "thing" that stands for a represented world that is also a "thing." It does so by virtue of certain relationships between it and the world it represents. But the concept of representation also includes an operational component. The representing world can be used for certain purposes instead of the represented world. In order for this to happen, there must be processes to operate on the represented world. We now consider briefly the interdependence of representation and its processing environment.

It is axiomatic within an information-processing framework that one cannot discuss representation without considering processes. The role of processing operations in the present analysis is that they functionally determine the relations that hold among the object elements. Consider World G of Fig. 9.1. The arrows connecting the circles in the diagram are only meaningful and useful if there are operations for finding them and the circles they connect. The operations of finding an arrow and its associated circles *define* the "points to" relation. Similarly, in World B, some operations define the "longer than" relation between pairs of lines, and other operations define the "chains to" relation in World H.

In these cases, we relied on our intuitive notions of "longer than," "points to," and "chains to." This works because we all have more or less the same operational concepts for these relations. In constructing processes to use these representations, however, operational definitions must be specified in terms of what the processes *do* to determine whether or not a particular relation holds. It is possible to have a representation that seems nonsensical by intuitive notions but is appropriate and sensible given the processses that operate upon it. The height of rectangles, for example, might be represented by line length where there is no intuitively obvious relationship (such as our usual concepts of "longer than" or "shorter than") to model "taller than." But if there is a process that *interprets* the length of lines – whatever they may be – such that corresponding lengths are *functionally* ordered just as the rectangle heights are ordered, then there is an *operational relation* defined by this process that corresponds to the "taller than" relation in the represented world.

A more familiar example of the same general concept is the "next" element of a list in list-processing computer languages. There is no necessary relationship between one element of a list and the next element in terms of physical location in memory or their numerical addresses. The list-processing language operationally defines the "next" element as the one "pointed to" (i.e., addressed indirectly) by the current element. Thus the "next" element is *functionally* next to the current element in terms of access order by the interpretive process. In an array-processing language, the situation is different. The "next" element of an array is defined such that it is also the next numerical address in memory. This does not mean that "next" is any less operationally defined in the array-processing language but only that it is more intuitively obvious what the relationship is. "Next" is completely defined by the operations in both cases.

The importance of this argument is that the relations in a representation are operational relations rather than apparent relations. Operational relations are simply those defined by the processes that interpret the representation. [Pylyshyn's "semantic interpretation function" is very similar to the present concept of operational relations (Pylyshyn, 1975).] Thus, in talking about operational relations, we implicitly include certain aspects of processing operations in the representation itself. Without those processes, the representation is meaningless.

The dependence of representation on processing goes even deeper. There is an important sense in which the only information contained in a representation

is that for which operations are defined. In other words, it does not matter whether certain relationships could, in principle, be derived from a given representing world if there are no methods for doing so. A good example is Shepard's (1962a, 1962b) multidimensional scaling procedure. Suppose there is a matrix that tables the ordinal distances among, say, 30 cities in the United States. It is obvious to everyone that this representation contains weak information about relative distances between cities. What is not so obvious – indeed, what was not known until recently – is that this matrix contains a great deal more information about distance and some information about direction as well. In fact, the matrix contains enough information to produce a reasonable approximation to a map containing those cities. The proof that such information exists is that Shepard's scaling algorithm (and its descendants) is able to recover it from the original matrix. We now know that a great deal of locational information is implicitly contained in an ordinal matrix of this sort, but prior to Shepard's demonstration, this was not thought to be the case. More to the point, however, is that simply knowing it is theoretically there is not sufficient for it to be considered part of the represented information. No person or machine can derive it without actually performing the operations. In short, the implicit locational information is *not there at all* except for the computer programmed to extract it, and for that computer, it *is there* even though it does not seem to be.

This rather esoteric example illustrates a very simple fact. The only information contained in a representation is that for which operations are defined to obtain it. When stated in this way, the point seems almost trivial, but it is not. As we see later, the representational nature of several kinds of theories have been universally misunderstood precisely because this fact is not appreciated. In general, we must be very careful about deciding just what information is contained in a representing world. The notion of operational relations changes the way we view our constructs of representation.

The Nature of Representation

Let us stop now and consider what we have learned from the analysis thus far. First, a representation requires a certain kind of relationship between two functionally separate worlds. Each world consists of objects that are characterized by relations that hold among them. These relations are operationally defined. The function of a representing world is to preserve information about the represented world. We can tie all of this together by assuming that the information contained in the two worlds is the set of operational relations among objects. Preserving information, then, is equivalent to having corresponding relations in the two worlds.

The nature of representation is that there exists a correspondence (mapping) from objects in the represented world to objects in the representing world such that at least some relations in the represented world are structurally preserved

in the representing world.[1] In other words, if a represented relation, R, holds for ordered pairs of represented objects, $<x, y>$, then the representational mapping requires that a corresponding relation, R', holds for each corresponding pair of representing objects, $<x', y'>$. This is just a very abstract and general way of describing situations like those shown in Fig. 9.1. The "taller than" relation in World A is preserved by the "longer than" relation in World B, by the "shorter than" relation in World E, by the "bigger than" relation in World F, by the "points to" relation in World G, and by the "chains to" relation in World H.

The same sort of representational relationship can hold for properties of individual objects. All objects in the represented world that are, say, 2 feet tall must correspond to objects in the representing world that have whatever property corresponds to 2-feet-tallness. This fits our description of representation if we view properties of individual objects as relations that hold for single objects. In fact, this is just the way properties are modeled in set theory – as "unary" relations (defined for individual objects) that are no different in principle from "binary" relations (defined for pairs of objects) or "n-ary" relations (defined for sets of n objects). We use the terms "unary relations" and "properties" interchangeably.

We now have at least an informal answer to the basic representational question. A world, X, is a representation of another world, Y, if at least some of the relations for objects of X are preserved by relations for corresponding objects of Y. The second goal is to characterize the ways in which two representations can differ from one another. A "representation" obviously refers to a representing world in relation to its represented world. The question of how two representations can differ, then, is really a question about how two worlds can differ from one another in the way they relate to their respective represented worlds. Given our definition of representation, it is clear that two representations can differ in the objects and/or the relations they represent. Having noted that two representations can differ in the objects they represent, we focus our attention on how two representations can differ when the objects they represent are the same.

If a pair of representations model the same set of objects, then there are two major kinds of differences to consider. First, two representations can model different relations of the represented objects. This is the case in Worlds B, C, and D of Fig. 9.1. Second, two representations can model the same relations, but in different ways. This is the case in Worlds B, E, F, G, and H of Fig. 9.1. Because representation is concerned with preserving information, and information consists of relations, we call the latter situation *informationally equivalent representation*

[1] This definition has a straightforward formalization in terms of model theory (Tarski, 1954). The represented and representing worlds are relational systems, each consisting of a set of objects and sets of relations. A representational system is an ordered triple consisting of the two relational systems and a homomorphic function that maps the represented objects into the representing objects. The basic approach is similar to that used in measurement theory for numerical representation (Krantz et al., 1971; Scott & Suppes, 1958; Suppes & Zinnes, 1963). This formalization, however, is beyond the current level of discussion.

and the former *nonequivalent representation.* The only remaining cases are those in which the same relations are modeled in the same way. This situation is called *completely equivalent representation.*

Nonequivalent Representations

Two representations that reflect different relations of the same objects are not equivalent in the sense that they do not preserve the same information. In other words, given two such representations and their processing systems, one could not answer the same questions about the represented objects from both representations. There are many possible differences that could result in this situation.

Type of Information. The most obvious condition for nonequivalence is that two representations can model qualitatively different dimensions of variation in the represented world. For present purposes, a "dimension" is just a set of mutually exclusive relations, only one of which is true for each object or set of objects on which the relations are defined. Properties of individual objects like height, length, size, and so forth are unary dimensions, because each individual object has only one value for each. The "values" along a dimension are simply the relations that comprise the dimension. Thus, "being two feet tall," "being red," and "having a hand" are possible values for unary dimensions of height, color, and handedness. Binary dimensions are defined just as unary dimensions except that they can hold only for ordered pairs of objects. The distance between two objects and their relative sizes are examples of binary dimensions, because one-and-only-one of the component relations can hold between each ordered pair. Similarly, n-ary dimensions can be defined for larger sets of objects – e.g., the relative distance between an object and two others.

The intended notion of differences in type of information represented is that one representing world may preserve some (but not necessarily all) information about a given dimension, whereas the other representing world may preserve no information about that dimension. In other words, one world represents that dimension somehow, but the other does not. Worlds B and C are examples, because B represents relative height information whereas C does not, and C represents relative width information whereas B does not. Clearly, this is an important way in which two representations can differ, because they cannot be used to answer the same *kind* of questions about the represented objects, much less the same specific questions.

Resolution. If two representations model the same dimension, they can still differ in many ways. The dimensional representation in one world may contain just a few relations, whereas in another it may contain many. The limiting cases are two relations and an infinite number. Consider some possible representations

of the length dimension. In one representation, all lengths are categorized into just two values: "short" or "long." In the other, they are categorized into, say, 100 values. There will be many objects having the same representation of length in the first system that have different representations of length in the second. These two representations are nonequivalent, because there are questions for which they provide different answers. For example, two objects both classified as "long" in the two-valued system might fall in two different length categories within the 100-valued system.

In general, any dimension can be described as containing m relations. The number of relations comprising the dimension is one important aspect of its *resolution* or *grain.* The larger the number of relations in a dimension, the higher the resolution and the finer the grain. We are presently assuming that the assignment of relations (values along the dimension) is completely deterministic, but it need not be. One could define a probability distribution over the m dimensional values for each object to construct a probabilistic representational system. This possibility, however, is beyond our present level of inquiry.

The other aspect of resolution is concerned with the particular relations that are preserved. Two representations might each have the same number of levels without those levels containing objects that correspond to the same objects in the represented world. Two maps, for example, might have three levels of dot-size to represent city-size. If they used different criteria for assigning city-size to dot-size, the same city might be represented as a large dot in one map and a medium dot in the other. In order for the resolution of two representations to be identical for some dimension, then, they must not only have the same number of levels but must classify the represented dimension in the same way.

Uniqueness. If two representations represent the same dimension with the same resolution, they can still be nonequivalent in the sense defined earlier. Consider two maps again with three levels for their representation of city-size. In one map, the cities are represented as black dots of three different sizes such that the larger the city, the larger the dot. In the other map, the cities are represented as red, blue, and yellow dots of the same size. Using the first map, one can tell which of two cities is the larger if they are represented as different sized dots. That is, the ordering of dot sizes preserves the ordering of city sizes to some extent. In the second map, this is not obviously true. One can tell whether two cities are generally the same in size or different, but their order is not necessarily represented. If there is a "key" on this map that indicates the size-to-color mapping, one can figure out the relative sizes. The key provides what is needed – an operational ordering relation for the colors. Without the key, however, the two maps are not equivalent, because questions about relative size of cities could be answered from the first map but not from the second.

We call this kind of difference the *uniqueness* of a dimensional representation, because it is analogous to the concept of uniqueness in measurement theory

(Krantz et al., 1971; Suppes & Zinnes, 1963). In the case where the levels of a dimension are not functionally ordered, only same/different relations are defined for pairs of dimensional values. This kind of representation is called *nominal*, after the similar case of measurement scales. It might be thought that nominal representations are so weak that they are uninteresting, but this is not so. Nearly all current theories of language representation are exclusively nominal, and many theories of perceptual representation contain substantial nominal components. Any dimension in which only identity is preserved is nominal.

If the relations of a dimension functionally order the representing objects as the represented objects are ordered, then order information is preserved as well as identity information. Such representations are called *ordinal*, after the corresponding type of measurement scale. Ordinal representations preserve more information than nominal ones in the sense that additional higher-order relations are meaningful; "different" relations of nominal dimensions are divided into "more" and "less" relations in ordinal dimensions.

It is not clear how to describe other types of uniqueness properties for nonnumerical representation (e.g., interval, ratio, or absolute representations corresponding to those types of scales). Certainly, when the representation is numerical, these concepts are meaningful. Perhaps they are in other kinds of representation as well if the correlates of numerical transformations can be identified. For present purposes, we simply note that such an analysis seems plausible.

Informationally Equivalent Representations

Two representations that preserve the same relations about the same objects are called *informationally equivalent*, because they are indistinguishable in terms of the information they preserve *about the represented world*. This does not mean that the representations are identical, of course. They can preserve the same information in many different ways. The fact that their methods of representation differ should not obscure the fact that they provide essentially the same view of the world they represent.

There are countless ways in which informationally equivalent representations can differ. These differences may be subtle (e.g., Worlds B and E of Fig. 9.1) or obvious (e.g., Worlds B and G of Fig. 9.1). No attempt is made here to catalog all the possibilities. Rather, I focus on two distinctions that seem to be important.

Intrinsic Versus Extrinsic Representation. The first distinction is most clearly exemplified by the contrast between Worlds B and G as representations of rectangle height in Fig. 9.1A. Consider two facts about the nature of the represented relation "taller than." First, if an object x is taller than an object y, it cannot also be true that y is simultaneously taller than x. In the language of logic, this fact defines "taller than" as an *asymmetric relation*. Second, if x is taller than y, and y is taller than z, than it must be true that x is taller than z.

This fact defines "taller than" as a *transitive relation*. The asymmetry and transitivity of "taller than" seem to be inherent constraints in the physical world.

It follows from our present definition of representation that if "taller than" is to be represented by some other relation, it too must be functionally asymmetric and transitive. There are two quite different ways of achieving this result. In World B, for example, the "longer than" relation seems to have the same inherent constraints. It is asymmetric, because if line x is longer than line y, then y cannot be simultaneously longer than x. It is transitive, because if x is longer than y, and y is longer than z, then x must be longer than z. We call this method of preserving structure *intrinsic representation*. Representation is (purely) intrinsic whenever a representing relation has the same inherent constraints as its represented relation. That is, the logical structure required of the representing relation is intrinsic to the relation itself rather than imposed from outside. The representation of "taller than" would be intrinsic if it were modeled by "shorter than" (World E), "larger than" (World F), "brighter than," "more numerous than," or any other relation that is inherently asymmetric and transitive.

The situation is strikingly different in World G, however. Here, "is arrow-connected to" represents "taller than," but there seem to be no inherent constraints on this representing relation. If x is arrow-connected to y, then y might be arrow-connected to x, or it might not. If x is arrow-connected to y, and y is arrow-connected to z, then x might be arrow-connected to z, or it might not. Thus arrow-connectedness is not *necessarily* either asymmetric or transitive, although it is *possible* for it to be either or both. Its ability to represent "taller than" follows directly from this fact. Asymmetry and transitivity can be literally imposed on it by requiring that it preserve the structure of its represented relation. We call this method of preserving structure *extrinsic representation*. Representation is (purely) extrinsic whenever the inherent structure of a representing relation is totally arbitrary and that of its represented relation is not. Whatever structure the representing relation has, then, is imposed on it by the relation it represents.

There are two ways in which intrinsic and extrinsic representation can be mixed. Obviously, one relation in a given representation can be modeled intrinsically and the other extrinsically. Not so obviously, both can be used in modeling the same relation. World H is an example of this. The "is arrow-path-connected to" relation is inherently transitive, because if there exists a path of arrows from x to y and another from y to z, then there must be a path from x to z. It is not inherently asymmetric, however. If there is a path of arrows from x to y, then there is no reason why there could not be one from y to x as well. Thus the transitivity of the "taller than" relation is represented intrinsically, whereas its asymmetry is represented extrinsically.

A word of caution is necessary about the distinction between intrinsic and extrinsic representation. The caution is that it rests on the concept of "inherent structure," a notion fraught with deep philosophical problems. After a moment's reflection, it is seen that "inherent structure" is closely related to the philosophical

concepts of "a priori knowledge" and "analytic and synthetic statements." In fact, intrinsic representation could just as well be called "analytic" and extrinsic representation called "synthetic." These are ideas about which philosophers have been arguing for centuries (e.g., Grice & Strawson, 1956; Quine, 1951). Despite such problems, I think the intrinsic–extrinsic distinction and the underlying notion of inherent structure are intuitively clear enough to be useful. As we see later, the distinction lies at the heart of a current psychological controversy.

Direct Versus Derived Representation. Another way in which two informationally equivalent representations can differ is in terms of how basic the information is within the representations. Intuitively, the distinction is between representing a relation so that it is a representational "primitive" and representing it so that it must be computed from other, more primitive relations. In World G, for example, the representation of "taller than" by "is arrow-connected to" seems more basic than in World H, where it is represented by "is arrow-path-connected to." The reason is that the latter relation relies on the former relation for its definition. In other words, one must make use of the "is arrow-connected-to" relation in order to evaluate the "is arrow-path-connected to" relation.

We call a representation of a relation *direct* if its operational definition relies on no other relations. Otherwise, the relation is *derived.* Any derived relation could be based on relations that are themselves either direct or derived. The dependencies that exist among relations determine the *derivational structure* of the system. Each relation can be specified in terms of how it is computed from other more basic relations.

There are some sticky problems involved with claiming that derivational structure is a representational issue. Strictly speaking, it is a question about how the representation is processed, because the definitions of relations are claimed to be operational. Still, there are cases in which it is obvious that one relation is derived from another – e.g., in World H of Figure 9.1. Direct representation is especially clear when representation is extrinsic, for reasons that become obvious later. With intrinsic representation, derivational structure is often obscure.

Completely Equivalent Representation

There is not a great deal to say about completely equivalent representations. They are simply informationally equivalent representations in which the same relations are modeled in precisely the same way.

It is worth mentioning, however, that no form of representational equivalence guarantees that performance characteristics will be the same for two representations embedded in process models. Even two completely equivalent representations may not have the same temporal characteristics, because a set of operations performed sequentially in one model may be performed simultaneously in the other. Error characteristics are similarly opaque without considering the processing environments for the representations in detail. The simple fact is that there

are a multitude of nonrepresentational factors that contribute to performance characteristics, and these can differ no matter how similar the representations might be.

Complex Representations

Thus far we have been discussing "simple" representations that model a single dimension of their referent worlds. The situation is far more complex in cognitive representation and in most real-world representations. Many different dimensions of the represented world can be modeled in the same system. This allows for the possibility that different aspects of the represented world may have qualitatively different representations. Consider a typical road map. Dots representing cities and lines representing roads are laid out in a spatial arrangement that simultaneously preserves a number of different dimensions of the real world. The location of cities is represented by the location of dots. The population of cities is represented by the size of dots. The condition of roads (paved, unpaved, highway, etc.) is often represented by the color of lines.

From the view of representation developed here, such a map is not representationally homogeneous. Representing city-location by dot-location provides a very high resolution, whereas representing city-size by dot-size generally has very low resolution. It seems, then, that the best way to characterize complex representations is in terms of the simpler dimensional representations that we have been considering. There is no single, acceptable description of the map as a whole, but we can say sensible things about it when broken down into dimensional components.

Interdimensional Structure. The separate dimensional pieces of a complex representation do not tell the whole story, however. When a representation contains more than one dimension, there is the possibility that pairs of them will not be independent. To the extent that this is true, there is interdimensional structure that must be preserved in the representing world. To use our familiar example, the height, width, and area of rectangles are not independent dimensions. If a rectangle is both tall and wide, it cannot be small. Height and width determine area in a fundamental way that prohibits such a combination.

Whatever interdimensional structure is present in the represented world must be preserved in the representing world for the modeled dimensions. This is not much of a problem if only a basic set of completely independent dimensions are represented directly and all others are derived from them. In such cases, the derivation generally takes into account the interdimensional structure. Otherwise, there is potentially a problem in preserving this information.

Once again, we can distinguish between intrinsic and extrinsic methods of preserving structure. If the height of rectangles were modeled, say, by the volume of spheres, and if the width of rectangles were modeled by their density, then the area of rectangles would be intrinsically represented by their mass.

This is so because the mass of spheres bears the same relationship to their volume and density as the area of rectangles does to their height and width. In other words, the inherent structure among dimensions in the representing world is the same as the inherent structure of dimensions in the represented worlds. The problem with this solution is that it very quickly becomes difficult to find analogous physical systems with all the required dimensional structure. For sufficiently complex representations, the constraints become so fierce that only a scaled model of the represented world will suffice. Although this is a satisfactory solution in some applications, it is not tractable for mental representation.

The other solution is to represent interdimensional structure extrinsically. That is, one could choose representing dimensions that are inherently independent and make them dependent by virtue of building in that structure. For example, the height, width, and area of rectangles could be represented by the length, brightness, and orientation of lines. Because length and brightness do not in any sense determine orientation, it would simply have to be the case that in the representing world, long, bright lines are oriented more vertically than are short, dark lines. Note that even though the individual dimensions involved are largely intrinsic representations of their referent dimensions, they are extrinsic at the higher level of interdimensional structure. Naturally, a representation can be extrinsic for both unidimensional and multidimensional structures.

Our hope that complex representations would be analyzable into simple representations turns out to be only partly realized. As more and more dimensions are added, higher-order structure increases drastically. Still, this general approach seems preferable to an unsystematic one. More importantly, by having considered simple representations first, we have come a long way toward our goal of a general framework – a metatheory, if you will – for representation. At least we have a coherent set of assumptions about what representation is and how representations can differ from one another at different levels.

COGNITIVE REPRESENTATION

We now turn our attention to the form of representation in which we were interested all along – cognitive representation. The plan is to use the framework developed for representation in the previous section to analyze the problem of cognitive representation. It must be clear from the outset that the goal is not to present a new and better theory of cognitive representation. Rather, it is to understand in a new and more fundamental way how cognitive psychology should approach mental representation and how we have been doing it for the past decade or so.

The discussion focuses on perceptual representation for two reasons. First, the concepts of perceptual representation currently in use are more confused and confusing than for any other cognitive domain. Second, the range of different proposals about perceptual representation seems greater than any other.

In fact, some are very similar to forms of representation currently in use in other domains such as language and various kinds of memory. In short, it is a microcosm of the state of cognitive representation as a whole.

Representation and the Cognitive Approach

The first thing about which we must be clear is exactly what we are doing when we construct a model or theory of mental representation from a cognitive point of view. Following Weizenbaum (1976), we make a distinction between theories and models. A theory of something is essentially a description of it at some level of analysis. It expresses the structural laws that hold in the object of study at a level of abstraction appropriate for the goals and methods of the scientific enterprise for which it is constructed. A theory, then, does not include aspects that are more concrete than can be verified by empirical observations of the sort indigenous to the science. A model is a concrete embodiment of a theory. Its relationship to its theory is that it satisfies the assumptions of the theory. Because there are many ways in which a given theory may be satisfied, there are many models that are consistent with it. All of these are described equally well by the theory. Thus the theory is simultaneously a description of its object of study and its many models.

For the current discussion, the object of study is mental representation of the world, perceptual representation in particular. The scientific field is cognitive psychology complete with its goals and methods. The question at hand is how cognitive theories and models relate to mental representation and its referent, the real world. Further, we want to know the scope of cognitive psychology in characterizing the nature of mental representation.

The proposed view of the situation is diagrammed in Fig. 9.2. To begin with, the "mental world" in which we are interested is some kind of representation of

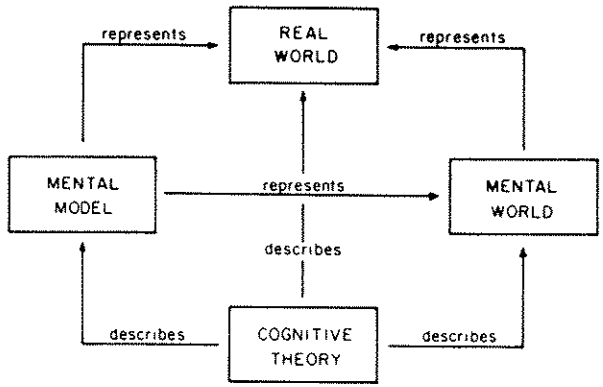

FIG. 9.2. A view of cognitive representation. Relationships among constructs are indicated by labeled arrows. See text for discussion.

the "real world." This is indicated by the "represents" arrow pointing from the mental world to the real world. A cognitive model of this mental world (the "mental model") is, in turn, a representation of that mental world. Thus the mental model is a representation of a representation of the real world. Almost by accident, the mental model is a representation of the real world in its own right.[2] This situation should begin to sound familiar. Both the mental model and the mental world are representations of the same represented world – the real world. This is just the case we considered earlier when discussing equivalence of representations. A relevant question, then, is what sort of equivalence can be achieved between the mental world and our hypothetical cognitive model of that world. The answer is that they should be as equivalent as is meaningful for the goals and methods of cognitive psychology. Not accidentally, this kind of equivalence is also the level of abstraction appropriate for the cognitive theory of mental representation. The theory should simultaneously be the proper description of both the mental world and the mental model. Indirectly, the cognitive theory is also a description of the real world, although it will differ substantially from, say, a physicist's. The fact that a cognitive theory also provides a description of the real world is significant. It explains why some theorists – notably Garner (1974) and J. J. Gibson (1966) – have been able to make important contributions to cognitive psychology by analyzing aspects of the world (the stimulus) rather than the representation of that world.[3] Note that they do not talk about the world in physical terms (like frequency and amplitude of mechanical deformations) but in psychological correlates of physical terms (like pitch and loudness of sound). Thus they are essentially describing the world in a way that is equally applicable to a mental model or mental world in an abstract sense.

[2]This idea can be demonstrated as follows. Suppose that the objects of the real world are a, b, c, and d and that those of the mental world are a', b', c', and d'. Because the mental world is a representation of the real world, there exists a mapping function (correspondence) from the real world objects to the mental world objects that could be expressed as $x' = f(x)$. Because the model of the mental world is also a representation of the mental world, its objects, a'', b'', c'', d'', can also be considered part of a similar mapping function from the mental world, $x'' = g(x')$. Now we see that the objects of the mental model can also be expressed by a mapping function from the objects of the real world, $x'' = g[f(x)]$ or $x'' = h(x)$, where h is just the composite of functions g and f. Thus, h is the representational mapping from the real world to the mental model. Although it is not universally true that this will result in the same relations being preserved in the mental model as in the mental world, it will be true if the mental model is an isomorphic (rather than homomorphic) representation of the mental world.

[3]The fact that Gibson does not acknowledge the existence of mental representation is irrelevant. His claims about the information available in the real world are still important for psychologists who postulate mental representations.

Following Neisser's (1967) classical statement, let us assume that the goal of cognitive psychology is to describe the "software" rather than the "hardware" of the mind. This assumption is justified by the kinds of experiments we perform. By and large, they are behavioral, not physiological, even when their object of study is a physiological distinction like hemispheric function. Even more to the point, scalpels are not included in our apparatus, and surgery is not part of our training. This means that neither cognitive theories nor experiments are properly concerned with the concrete way in which mental representation is accomplished within the brain and nervous system. Our theories and experiments *are* concerned with the nature of the information represented about the external world. Moreover, we want our models to be as equivalent to the mental world as possible in terms of the information contained about the world. In the language defined earlier, cognitive psychology can aspire only to informational equivalence between its models and actual mental representation inside the head.

This fact determines the proper level of discourse for cognitive theory as the level of abstraction defined by informationally equivalent systems. The representational issues of concern for cognitive theory are things like the types of information represented, the resolution of the dimensions represented, their uniqueness properties, and the higher-order structure that exists among different dimensions. These are our tools of analysis for dealing with mental representation from a cognitive approach. Questions about the concrete physical aspects of the mental world are inappropriate and irrelevant. The distinction between intrinsic and extrinsic representation is also beyond our reach. Derivational structure is somewhat unclear, because it is irrelevant as a representational construct alone but probably is relevant as a processing construct. Within this framework for cognitive representation, let us try to understand our current concepts of perceptual representation.

Notation and Illustration

When a theorist proposes a theory of representation, he or she usually draws one or more diagrams to illustrate the nature of the theory. These diagrams are essentially small pieces of a model of the theory being proposed. One problem with the current view of representation is the pervasive belief that these diagrams can be taken uncritically as the theory being put forth. A more subtle form of the same mistake is to assume that even though the figures are not the theory, they are intuitively transparent to the theory.

To demonstrate the flaw in this way of thinking, Fig. 9.3 shows seven standard, easily recognizable types of representation for a diagonal line. Within the usual classificational system, there is a template (A), a neural network (B), a digital matrix (C), a multicomponent vector (D), a binary feature set (E), a list of propositions (F), and a relational network (G). Most of these are currently thought

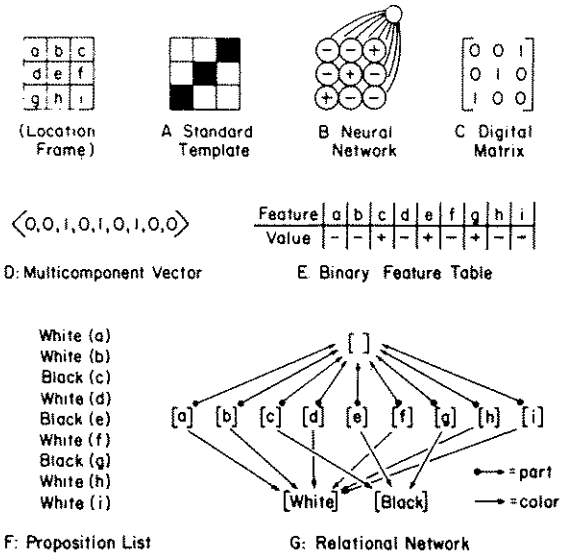

FIG. 9.3. Seven standard cognitive representations of a diagonal line. The letters *a–i* in Representations E, F, and G refer to the locations indicated in the Location Frame in the upper left corner. In Representation D, the vector components are in alphabetical order of the location-frame letters (*a, b, c, ...*).

to be different theories of representation. What are the differences, and are they meaningful for cognitive theory? They look different, and we talk about them differently. But these things may or may not reflect substantive differences at the level of cognitive theories of representation.

Let us examine them from the new view of representation. Each one consists of nine components, where each component represents a point (or, equivalently, a point's location) in the pattern. What dimensions are represented? Each representation contains information about the location and color of the individual points. In A, B, and C, location information is preserved by spatial location, in D by position in the vector, in E by identity of the feature, and in F and G by identity of the arguments of the relations. In A, color is represented by color, in B and E by "+" and "–", in C and D by "1" and "0" entries, and in F and G by the labels "white" and "black." In each case, then, color is a two-valued dimension, and location is a nine-valued dimension. The representation of both color and location is nominal in most cases, although A, B, and C seem to contain more information about location. As argued earlier, one must consider the processes that operate on the representations to determine what information is actually included in the representation. As we see later, the processing assumptions for standard templates, neural networks, and digital matrices reveal that location is represented nominally.

If this is a reasonable analysis, all seven representations in Fig. 9.3 are informationally equivalent but not completely equivalent. That is, they differ only in the concrete way information is preserved, not in what information is preserved about their referents. This means that they are all models of the same cognitive theory. In fact, they all turn out to be models of standard template theory.

I have obviously taken some liberties in constructing these drawings. No one has ever (to my knowledge) proposed a feature theory quite like the one shown. But why not? It is against the rules of feature theories? If so, why? These are questions that must be answered in order to understand our theories properly. The pictures we draw must not be confused with the representational assumptions contained in the theory itself. We must see *through* the surface form of those pictures to the information they contain about the represented world. That is what cognitive theories of representation are all about.

Templates

Until recently, templates have been the perennial "straw men" of perceptual representation. Discussions by Neisser (1967) and Lindsay and Norman (1972) have succeeded in convincing a whole generation of cognitive psychologists that templates – whatever they are – are useless representations for pattern-recognition systems. Two developments have brought about a resurgence of interest in templates. One is the construct of "prototypes" in representing categories (Rosch, 1973; Chapter 2 of this volume). The other is recent work on image rotation (Cooper, 1975; Cooper & Shepard, 1973; Shepard & Metzler, 1971) and image scanning (Kosslyn, 1973; 1975a; Chapter 8 of this volume). Rightly or wrongly, these phenomena have been seen as possible evidence for the existence of template-like representations. In addition, templates seem to be good candidates for the kind of low-level visual information storage studied by Sperling (1960).

The construct of templates has been with us for a long time, but no one has ever really defined it properly. Perhaps the most common "definition" is to point to a figure that displays a digitized pattern overlapping to a certain extent with an input pattern (Neisser, 1967, p. 51; Lindsay & Norman, 1972, pp. 2–6). The figure and accompanying text provide the reader with an intuitive feel for how template matching systems work but no real definition of the representations on which they operate. Worse still, processing assumptions that are totally independent of representational assumptions (e.g., parallel matching) are often confused with the form of representation.

The basic problem in understanding the fundamental nature of templates is that they are displayed as *pictures* of the patterns they are intended to represent. These pictures have all the information in their referents, at least implicitly. The

template-picture has lines and angles plus properties like closedness and symmetry. It is not at all transparent to the nature of the underlying theory. Only when the operations performed on templates are considered do the assumptions about representation become clear.

Standard Templates. The simplest case is a standard template match without any "preprocessing operations" (Neisser, 1967). Consider what happens when a template match is performed. The template in memory is compared to the input pattern in a point-to-point fashion, where location determines the correspondence of points. For each pair of corresponding points that have the same color (both white or both black), a match is registered. If there are n locations, then the number of matches can vary from no points to all n points. This defines a similarity dimension with $n + 1$ values for pairs of templates. It is used to classify patterns according to some decision strategy, usually of the best-fit variety.

There are two important things to notice about this process. First, no components are considered except individual points, and no dimensions are considered except location and color. Thus angles, lines, closedness, and all the rest of the information in the template-picture are irrelevant. They are not represented information, because there are no operations that define them or use them. Second, the matching process requires only information about the identity of locations and colors. It does not matter in the slightest whether one location is above, below, or close to another location, as long as it is matched to the corresponding point in the input pattern. In sum, both location and color are nominally represented dimensions.

Standard templates, then, are defined as follows. Templates are representations in which each pattern is composed of n points, and each point is defined by just two dimensional values: one from an n-valued, nominal dimension representing location and the other from a two-valued, nominal dimension representing color. Thus, all of the illustrations shown in Fig. 9.3 are surface variations (i.e., informationally equivalent models) of standard template theory when the usual processing operations are employed. Notice that many of the bizarre forms are far more revealing of the essence of templates than is the usual form.

Preprocessed Templates. The problems with standard templates are well known. Because they are position-, orientation-, and size-specific, trivial changes in these parameters of the input pattern have catastrophic consequences for classification performance. To correct these difficulties, Neisser (1967) suggested "preprocessing operations" to normalize (translate, rotate, dilate, and "clean up") the input pattern prior to matching. The extent to which such operations actually solve such problems is not at issue here. The important point is that these preprocessing operations require a significant change in representational assumptions.

It is still true that points, locations, and colors are the basic kinds of information represented. But the representation of location cannot be nominal. In order to "fill in" a light point surrounded by black points, relations like "between" must be operationally defined. In a nominal representation of location, "filling in" operations could not be performed. Similarly, shifting a pattern x units in a given direction cannot be done with nominal representation of location. In fact, the usual preprocessing operations seem to require that location be represented at least intervally, because units of locational dimensions must be constant.

Similar changes could be made in the representation of color. The resolution of the color dimension could be greatly increased, and its values could be made ordinal or interval. Such information could be used to construct more powerful matching procedures that include partial matches for intermediate levels of grayness. In any case, the representational assumptions of preprocessed templates differ significantly from those of standard templates.

Hierarchical Templates. A slightly more deviant type of template theory is what might be called "hierarchical templates" or "minitemplates." The basic notion is that the components of a complex pattern might be a set of simpler templates rather than just a set of points. The simpler templates would then be defined by either even-simpler templates or by a set of individual points. Thus this representation is a hierarchy with individual points at the terminals. Such representations border on structural descriptions in that they have an articulated structure of higher-order parts.

The general nature of templates, then, rests largely on the kind of information represented: points, locations, and colors. There seems to be some agreement on color being represented as a nominal, two-valued dimension, but that could be relaxed. Note that the assumptions are actually of the sort that should be important for a cognitive theory of representation. Once the surface form is disregarded, templates are a fairly well-defined theory. Perhaps this is why template theory is so easily and frequently shown to be false. As we see later, most other purported theories are so vauge that they cannot really be tested.

Features

Feature representations were invented as an alternative to templates. As initially proposed, features were things like horizontal lines, angles, curves, and so forth. Since then, many additional features have been postulated for special purposes: closedness, complexity, wiggliness, and height of forehead, to name just a few. They are probably the most widely used form of representation. A large part of their popularity stems from their flexibility; anything can be a feature. This is

simultaneously their greatest strength and greatest weakness. It makes them convenient to use to explain data, but it makes them inherently ill-defined as a theory.

There are a number of feature theories in current use. The three most popular seem to be binary features, multidimensional spaces, and hierarchical features. We consider the representational assumptions of each in turn and clarify some of the relationships among them.

Binary Features. Perhaps the most common type of feature theory is sets of binary features. Prominent examples include E.J. Gibson's (1969) distinctive feature theory of letters and Jacobson, Fant, and Halle's (1961) distinctive feature theory of phonemes.

Binary feature theories operate more or less as follows. A set of n operational feature tests are applied to an input pattern – either serially or simultaneously. Each test has two possible outcomes, one for presence of the feature and the other for absence of the feature. (Precisely how this happens is seldom discussed.) In Gibson's theory, for example, letters are defined by presence or absence of properties like having horizontal lines, being closed, and being symmetrical. The results of these tests are compared to a set of stored representations of pattern types, each one being defined by the outcomes of the same feature tests. For each feature, a match is registered if the input pattern has the same value as the stored representation – i.e., both have the feature, or both fail to have it. Some measure of similarity is computed according to a function that integrates the matching results for each feature. The simplest computation is the number of matches, although various weighting parameters can be introduced to reflect the saliency of different features. The resulting similarity dimension is then used to classify the pattern according to some decision strategy. The most frequently employed rule is a 100% threshold, requiring that the input pattern have all the same features as the memory representation of the pattern type. Other more complex decision strategies are sometimes used (e.g., Smith, Shoben, & Rips, 1974).

The representational assumptions of standard binary feature theories are quite clear. Patterns are represented along n different unary (property) dimensions. Each dimension has just two values defined by the results of their operational definitions. The dimensions are nominal in that only same/different relations are defined for the matching procedure. Thus each pattern is a set of values along n different, two-valued, nominal, unary dimensions.

Although these assumptions are clear, they are not very specific. Because the informational nature of the dimensions is unspecified, the range of possible theories within this class is enormous. Note that if the features are exclusively location-colors (see Figure 9.3E), the result is a standard template theory. In other words, if the features are strictly of the form "is dark at location x" (where x is identified as a different location for each feature), the foregoing

procedure would operate exactly as the template matching system described earlier. Thus standard templates are a special case of binary features.

An obvious extension of standard binary feature representations would be nominal m-valued feature systems. All that changes is the resolution of the dimensions. An example would be a set of color names (red, blue, yellow, green), in which each object has one color only and the only information in the color values is whether they are the same or different.

Multidimensional Spaces. Another common form of feature representation is a multidimensional space. The basic metaphor is that mental objects can be modeled as points in a metric space of n dimensions. Relations among groups of objects are preserved by spatial relationships among sets of points. For example, overall psychological similarity is usually assumed to be reflected in distance relationships. The illustrations used to depict such representations are usually low-dimensional spaces with points labeled by object names and with coordinate axes labeled by dimensional feature names. Most multidimensional space representations are derived from the nonmetric scaling techniques originally developed by Shepard (1962a, 1962b). Prominent examples include spatial representations of animals (e.g., Rips, Shoben, & Smith, 1973), states (Shepard & Chipman, 1970) and numbers (Shepard, Kilpatrick, & Cunningham, 1975).

Multidimensional space representations are used to classify patterns in the following way. A set of n operational feature tests are applied to the input pattern. Each test has m possible outcomes, where m is a relatively large number usually assumed to approach infinity. The outcomes generally represent the *degree* to which the instance has that feature and are interpreted as interval representations. The results of these tests specify the point within the n-dimensional space occupied by the input pattern. The point is then compared to a set of stored representations of pattern types defined for outcomes of the same feature tests. At this stage, two different classification methods diverge: the point method and the region method. In the point method, the stored representations of pattern types are single (or sometimes multiple) points in the metric space. The input pattern point is compared to the category point(s) using some form of distance metric, usually Euclidean or city-block. This metric is taken as a representation of psychological similarity, and the pattern is classified with the category to which it is closest (most similar). In the region method, the representation of pattern types is in terms of a region with in the space. The input pattern is then classified as an instance of the category within whose region it falls. [See Reed (1973) for a more comprehensive treatment of spatial models of categorization.]

Notice that the basis for multidimensional space representations is essentially an analogy. It specified that a possible *model* for cognitive entities is a spatial one. But the spatial aspects of the model cannot be taken as part of the underlying cognitive theory, because the medium of representation is not within the

cognitive level of analysis. That is, the "space" is metaphorical. The essence of multidimensional spaces as a *theory* of representation must lie elsewhere.

Without loss of generality, each point in an n-dimensional space can be described by a vector of n ordered values. Each component of the vector specifies the projection of the point on one of the axes of the space. Spatial relationships like distance are mathematical relations on vectors. Because the vector form of multidimensional feature theory contains the same information as the spatial form but without the spatial assumptions, it is more transparent to the underlying theory of representation. This theory assumes that there exists some number of highly resolved, interval, unary dimensions in a mental representation. In practice, the number of dimensions represented in multidimensional spaces is small. This is more a constraint of illustrating the representation than of the underlying theory itself, however.

Multidimensional spaces and binary feature theories seem very different from each other. Binary feature models are presented as tables, whereas multidimensional spaces are presented as spaces. These surface differences are not very revealing of the underlying similarities and differences. The commonality that seems to make them both versions of feature theory is that they both represent unary dimensions (properties of individual objects). The differences are that binary feature dimensions are two-valued and nominal, whereas multidimensional spaces are multivalued and at least interval.

Having noticed these relationships between binary features and multidimensional spaces, a number of other things become clear. For example, the language we use to talk about these models becomes as arbitrary as their surface form in illustrations. Consider the following spatial characterization of templates: "A template is a point in a discrete dimensional space. Each dimension represents the color of a particular location of the pattern, and if there are n locations, there are n dimensions in the space. For each dimension, there are just two values: 0 if the pattern is light at the location and 1 if it is dark. Thus, the input pattern and each memory representation occupies a point in this discrete space. The input pattern is then classified as an instance of the pattern type to which it is closest according to a city-block distance metric." On the surface, this seems very different from the usual description of templates, yet the theory described is the same. Because templates are a special case of binary features, the same sort of translation could be done for any standard binary feature theory.

The point of this discussion is not to argue that templates, binary features, and multidimensional spaces are all the same. As these concepts are *used,* they are certainly not. But they do have similarities as well as differences. These relationships are not obvious from the pictures used to illustrate them or the language used to talk about them. They are found in the basic representational assumptions.

Hierarchical Features. Another version of feature theory is that in which feature dimensions are structured according to their interrelationships. The basic idea is that complex features can be defined in terms of more primitive ones. A familiar example is found in Lindsay and Norman's (1972) pandemonium model. At a concrete level, each pattern is defined by a number of orientation-specific features such as the number of horizontal lines, the number of vertical lines, and the number of oblique lines. At an abstract level, patterns are defined by orientation-free features such as the number of lines of any orientation. The higher-order dimension of number of lines is the sum of the values for the lower-order dimensions that comprise it. Thus, heirarchical feature theories are one way to specify logical dependencies that exist among different dimensions.

There are other relationships among dimensions that are more psychological than logical. Many of these concern what Garner (1974, 1976; Chapter 5 of this volume) calls "integrality" and "separability" of dimensions. As a paradigm case of integrality, hue, saturation, and brightness seem to be closely related psychological dimensions. We even have a name for this set of dimensions – "color." There seems to be a level of analysis at which these three dimensions combine into a unitary aspect of the stimulus. People can make separate judgments about the component dimensions, but it is difficult. This is a facet of dimensional representation that cannot be modeled in either standard binary feature systems or multidimensional spaces. Hierarchical features provide a mechanism for doing so by allowing structural relationships among different dimensions to be represented. It is not clear whether this violates the assumptions of feature theory in general. Just as hierarchical templates are deviant versions of template theory, so are hierarchical feature theories deviant versions of feature theory.

Structural Descriptions

Some theorists came to reject both features and templates as representations for pattern recognition because of their poor modeling of structural interrelationships among patterns and their parts. For example, although a feature theory can represent a given pattern having three lines and three angles, there is no representation of which lines are part of which angles, nor of how the lines are connected to form the angles. A more powerful representation is needed for such information, because a pattern is no longer defined solely by unary dimensions (properties). One must represent facts such as that Pattern P contains X and Y as parts and that the top of X is joined to the middle of Y – i.e., that X "is top-middle-connected to" Y.

The initial attempts at structural description theories followed the formalisms of generative grammars (Narasimhan, 1969; Narasimhan & Reddy, 1967). The representation of a pattern was the set of production rules required to generate

it from primitive patterns. The production rules specified the manner in which parts were to be combined into wholes. Since then, grammatical formalisms have largely disappeared, but the legacy of representing *n*-ary relations has remained. It is essentially the power of representing relations on more than single objects that differentiates structural descriptions from features.

Simple Structural Descriptions. Simple structural descriptions are closely related to hierarchical templates. The basic idea is that a pattern is defined by relationships among subpatterns. Each subpattern can be thought of as a mini-template. The specified relationship among the minitemplates is satisfied by moving and turning the minitemplates until the proper relations hold. Each of the subpatterns, of course, could be either a primitive (usually, points or lines) or relationships among further subpatterns. Simple structural descriptions differ from hierarchical templates in that the relations among subpatterns are subject to variation. It is this variation in relationships that is represented by the structural relations in the representation.

In a structural description, each pattern is represented by a set of values for *m* *n*-ary relational dimensions. Each value of an *n*-ary dimension is a relation among *n* subpatterns. Because this is a rather complex concept, let us consider an example. The letter "T" might be defined in the following kind of structural description: "T" is the connection of the middle of a horizontal line to the top of a vertical line. This description actually contains three components: "T" contains a horizontal line, "T" contains a vertical line, and the horizontal line is middle-top-connected to the vertical line. The first two are binary relations between T and its subpatterns, and the third is a binary relation between the two subpatterns themselves. It is the third component that is important and distinguishes structural descriptions from feature theories.

Consider how to translate the three components into unary relations for a feature theory. The "contains" relations between T and its parts are no problem. One simply defines two different unary relations by binding the two parts to the second argument of the "contains" relation. The features of T then become "contains-a-vertical-line" and "contains-a-horizontal-line," both of which are familiar features in existing feature theories. But the last relation is a problem for feature theory. It is not a problem for structural description, because both the horizontal and vertical lines are independent object-elements in their descriptions. Therefore, representing the relation between them is no different in principle from representing the relations between T and the subpatterns. In a feature theory, however, the horizontal and vertical lines do not have independent status as representational objects. They are inextricably bound within the two constructed unary relations. The usual solution is to add more unary relations that indirectly constrain how the horizontal and vertical lines can be arranged. For example, many possible arrangements of the two lines are ruled out when the following features are added to the definition of T: "has an intersection,"

"has two right angles," "is open," and "is symmetrical." Despite these complex, higher-order features, the indirect constraints are seldom tight enough to rule out all illegal patterns. The only sure-fire feature would be something like "has-a-horizontal-line-whose-middle-is-connected-to-the-top-of-a-vertical-line." At this level of complexity, one might just as well have a feature called "looks-like-a-T."

The sure-fire solution for feature theory has the unfortunate side effect of proliferating represented relations. Suppose that there are just 10 basic subpatterns for a given class of patterns and 10 possible relations that might hold between any pair. The structural description theory would have 20 primitive elements, one for each subpattern and one for each relation. They would be combined in appropriate ways to define the patterns. The corresponding feature theory would require, in principle, 1000 features to have the same power. Each complex feature would be pairing of each of the 100 possible ordered pairs of basic subpatterns with each of the 10 possible relations. Thus, it may well be the case that there is always a feature theory that is informationally equivalent to a given structural description theory, but the latter is preferable on grounds of simplicity.

Augmented Structural Descriptions. One potential drawback to simple structural descriptions is that, like templates, their representations ultimately rely on just locations and colors of points. Although these primitives are combined in powerful ways into higher-order parts and patterns, every bit of information must be derived from them. The problem is that when higher-order parts are formed by relationships among component parts, the larger patterns frequently have "emergent properties" not defined for the components. For example, a line can be defined as a particular relationship among a set of points, but the line has properties like length and orientation that are not properties of the component points. Similarly, when lines are combined into a square, the square has properties like area and closedness that are not properties of the component lines.

An obvious solution to this problem is to augment simple structural descriptions with features for the higher-order patterns. The result is a hybrid of feature theory and structural descriptions that we call *augmented structural descriptions.* The general assumptions are that any pattern is represented *both* as a set of unary dimensional values and as a set of relationships among component parts. This provides the power necessary to represent emergent properties like those mentioned previously. Examples include the models of Palmer (1975a) and Winston (1975).

By this point, we have reached a type of "theory" that is so powerful that it no longer is any theory at all. Notice that there are virtually no constraints on what information can be represented. There can be any number of dimensions that can hold among any number of objects. The dimensions can represent any kind of information with any resolution and any uniqueness properties. In their

most general form, then, augmented structural descriptions are an untestable theory of perceptual representation as a general class. Particular examples of the class, however, can be tested. But one must do things like specify what dimensions are represented and how they are represented. Only then does an augmented structural description theory become more than an abstract framework in which to construct particular theories.

Prototypes

A great deal of interest has recently been generated about the possible role of prototypes in cognitive representation and processing. This development is due in large part to the seminal work of Eleanor Rosch on categorical prototypes (Rosch, 1973, 1977; Chapter 2 of this volume). She has demonstrated the existence of prototypes for natural categories like colors and animals as well as for artificial categories like dot patterns and schematic drawings. The evidence that prototypes of some sort play a critical role in human categorization is compelling. The question that concerns us here is the nature of the representational assumptions required by prototype theories.

The most common belief about prototypes is that they must be templates of some sort. This is partly because prototypes are frequently discussed as "images" and because they are associated with particular examples of the category. The association between prototypes and templates is further strengthened by the fact that templates are universally described as "prototypical" examples of their class. Because everyone who has read an introductory text in cognition has been informed that templates are *wrong*, however, prototypes are usually thought to be only "template-like." For example, in a recent book on pattern recognition (Reed, 1973), we find the following discussion of results demonstrating the importance of prototypes:

> Insofar as a prototype may be thought of as a type of template, these results also support a template theory. But a prototype is not an unanalyzed template in which the amount of overlap is used to judge its similarity to other patterns. Instead, a prototype consists of features and when it represents the central tendency, is determined by the mean value of each feature when the mean is calculated from all patterns in the category. [p. 32].

This passage illustrates the confusion about prototypes; they are template-like and yet they are not templates but features. In what sense are they like templates and in what sense like features? Are they necessarily related to these concepts at all?

Let us begin by considering how a prototype theory of perceptual classification operates. There are categorical representations stored in memory. These are assumed to be highly specific in the sense that they approximate the "most typical" or the "ideal" instance of the category. However these representations are constructed, the input pattern is represented along the same dimensions as the prototypes. A measure of similarity is computed between the input pattern and each categorical prototype. The similarity is assumed to be highly resolved or even continuous. A decision strategy is used to assign the pattern to a category on the basis of degree of similarity. The most common classification rule is of the best-fit variety such that each pattern is classified into one-and-only-one category.

The general form of this process is the same as before. The input pattern is represented and compared to stored representations for similarity. The similarity measures are then used to categorize the pattern. The major elaborations for prototype theories are: (1) the stored representations are highly specific; (2) the similarity dimension is highly resolved; and (3) a best-fit criterion is used for classification. Note that both (2) and (3) are pragmatically determined by (1). That is, if the categorical representation is highly specific, then using a similarity dimension with low resolution and/or using strict 100% threshold decision criteria would result in unacceptable categorization performance. Either too few instances would be classified at all, or some instances would be classified into many categories. Thus the basic assumptions of prototype process-models of classification all follow from the single assumption that categorical representations are highly specific.

Let us consider this assumption more carefully. In what sense is the categorical representation highly specific? The prototype approach is properly considered in opposition to the "invariant attribute" approach (e.g., E.J. Gibson, 1969). The essence of these theories is a very general representation of categories such that each instance is completely and equally consistent with it. For this to be true, invariant-attribute representations of categories cannot represent dimensions that vary within the category. They represent only dimensions that vary across categories. Now it becomes clear that the specificity of prototypes is with respect to within-category variation. A prototype representation is one that has relatively high resolution for dimensions of information that vary within the category.

The standard view seems to be that the prototype and invariant-attribute approaches are dichotomous. The present analysis leads to a different view – namely, that underlying the dichotomy is a broad range of possible theories, which differ in their representation of within-category variation. At one extreme is the usual prototype approach – the prototypical prototype approach, if you will – in which every aspect of within-category variation is represented. At the other extreme is the standard invariant-attributes approach, in which no aspect

of within-category variation is represented. In between are numerous sensible compromises. A categorical representation need not be uniformly specific for all dimensions of variation, for example. Some might be quite specific, others less so, and still others not at all.

Nothing has yet been said about exactly what the dimensions are nor just how they are represented. This is because prototype theories as a general class do not require such assumptions except when they are particularized for a given category. Certainly, standard templates satisfy the specificity constraint, but they are not the only kind of theory that does so. Binary feature theories can also represent within-category variations, provided there are sufficiently many such features (e.g., Smith, Shoben, & Rips, 1974). In fact, any reasonable theory of perceptual representation in general will have to be consistent with the notion of prototypes, because it will have to have the capability of representing highly specific instances. If not, the theory would never be able to account for how people can distinguish their own house from other houses or their own dog from other dogs. In short, prototypes are a construct of categorical representations, not of representations in general. As a class, they are equally compatible with virtually any theory that can represent specific instances.

Shepard's Principles of Isomorphism

Roger Shepard has discussed the nature of possible forms of isomorphism that might hold between the real world and people's internal representations of that world (Shepard & Chipman, 1970; Shepard, 1975). In particular, he has distinguished between what he calls "first-order" and "second-order" isomorphism. These concepts have gained wide currency within cognitive psychology and are believed to have implications for the field. I discuss the nature and importance of Shepard's proposals because they seem to be frequently misunderstood. First, I formulate them within the new framework of representation to clarify them. Then I argue that second-order isomorphism, in its most general sense, is not a theory of representation but a definition. Finally, I argue that the distinction between first- and second-order isomorphism is irrelevant for cognitive psychology.

First-Order Isomorphism. Simply stated, first-order isomorphism is a concept of mental representation in which the properties of real-world objects are retained in the internal representation of those objects. To use Shepard's original examples (Shepard & Chipman, 1970), in a first-order isomorphism, the representations of green things must be themselves green, and those of square things must be themselves square. Shepard rejects this notion of the correspondence between the real world and the mental world by arguing that it is physiologically absurd to suppose that the internal representations of green things are themselves green and that it is unnecessary and implausible to suppose that those of square

things are themselves square. In a later paper, Shepard (1975) refers to this as concrete first-order isomorphism.

Translating this concept into the present framework is straightforward. A representation is concretely first-order isomorphic to its referent if the unary relations (properties) of the represented objects are preserved by the physically equivalent unary relations (properties) of the representing objects. Thus, greenness is reflected by greenness and squareness by squareness. The most sensible interpretation for the name of this concept is that "first-order" signifies that relations on individual objects (properties) constitute the information of interest and that "concrete" signifies that the corresponding relations must be physically the same.

Shepard distinguishes concrete from abstract first-order isomorphism in a later paper (Shepard, 1975). His example is that an abstract first-order isomorphism would hold if the internal representation of a square contained four parts, each of which correspondended to a corner of the square. Unfortunately, the example is ambiguous, and Shepard does not discuss it fully enough for the reader to know exactly what he intended.

The "abstract" version of our previous definition should be that a representation is abstractly first-order isomorphic if the unary relations of the represented objects are preserved by functionally (or operationally) equivalent unary relations of the representing objects. This example is then interpreted as follows. One of the properties of squares is that they have corners as parts, four of them in fact. This is reflected in the representing world by the square-representation having corner-representations as parts, four of them in fact. The isomorphism is not concrete, because "having-four-corners-as-parts" is not physically equivalent to "having-four-corner-representations-as-parts." In the special case where the corner-representations are themselves corners, concrete first-order isomorphism would hold.

In the other interpretation, we consider binary relations within each world rather than unary ones. The square has a relationship to each of its corners: Namely, the corners are "part of" the square. Similarly, the square-representation has a relationship to each of its corner-representations: Namely, the corner-representations are "part of" the square-representation. Now it seems that the represented binary relation is physically the same as the representing binary relation. In other words, the example might also be considered an instance of concrete second-order isomorphism by simple extension of our previous definition. I am not sure that Shepard actually meant to convey either of these concepts, but within the present framework, they are both possible.

Second-Order Isomorphism. Shepard's alternative to first-order isomorphism is second-order isomorphism. His example is that the internal representation of a square need not be itself square, but – whatever it is – it must be functionally more similar to a rectangle than to a green flash or the taste of persimmon

(Shepard & Chipman, 1970). It is important to realize that more than one thing has changed from the initial example of concrete first-order isomorphism. As the name implies, one change is from talking about the correspondence between properties of single objects to talking about that between relationships among more than one object. One would think this is the important change, but it is not.

Consider the binary relation "greener than" in contrast to the unary relation "green." First-order isomorphism implies that if an external object is green, its internal representation is also green. If second-orderness is the essence of second-order isomorphism, then replacing "green" with "greener than" relationships should yield an example of second-order isomorphism. Thus, if object A is greener than object B, then the representation of A should be greener than the representation of B. The only constraint that has been lifted is that the properties of individual objects in the external and internal worlds need not be identical. That is, the representation of A need not be itself green (it might be blue-green or even blue), but it still must be greener than the representation of B.

Now we see that the other change is from physical sameness to functional sameness. This is the important one. Because of it, external greenness need not be like internal greenness in any physical sense but only in a functional sense. That is, when the greenness of an external object changes, there is *some* corresponding change in the internal object that may be nothing like changes in greenness. This functional correspondences is what effectively decouples the internal and external world in terms of resemblances. Shepard's terminology is unfortunate, because it emphasizes the wrong change. It would be better to call first-order isomorphism "physical isomorphism" and second-order isomorphism "functional isomorphism," where either can hold for properties or any higher-order relations.

Second-Order Isomorphism: Theory or Definition? Regardless of what one chooses to call it, the basic concept referred to as second-order isomorphism can be defined as follows. A representation is second-order isomorphic to its referent world if the similarity of represented objects is functionally reflected by the similarity of the corresponding representing objects. What makes this difficult to pin down is the construct of similarity in both the external and internal worlds. Whether second-order isomorphism is a theory (or class of theories) of cognitive representation or just a definition depends on how broadly the concept of similarity is interpreted.

In its most general sense, similarity is a binary dimension containing at least two values. That is, the crudest sort of similarity is a one-bit classification into "same" or "different" relations. Inserting this into the previous definition we have: A representation is second-order isomorphic to its referent world if a binary dimension containing at least two values is functionally reflected by a binary dimension containing at least two values for the corresponding representing objects. This must be true even for purely nominal representations, because they preserve same/different binary relations. Thus, verbal labels for objects

satisfy this broad interpretation of second-order isomorphism. (It is perhaps worth noting at this point that verbal descriptions constructed by selecting one label from each of n sets of possible descriptions are no different in principle from nominal feature sets of the kind discussed earlier. Both are purely nominal, unary dimensions in which only same/different relationships are meaningful.)

I suspect Shepard's intention was to convey a more restricted concept of similarity, however. The intuitive notion of similarity connotes a highly resolved or even continuous dimension to most psychologists. With this notion of similariry, second-order isomorphism becomes more specific. It rules out, for example, representations containing only a few nominal dimensions. If the concept of second-order isomorphism is further required to hold within a single dimension, then it rules out all representations with low resolution of their dimensions (e.g., standard binary features). In any case, the status of second-order isomorphism as a theory or a definition of representation is unclear as long as the construct of similarity is left undefined.

Distinguishing First- from Second-Order Isomorphism. As Shepard has pointed out, all first-order isomorphisms are necessarily second-order isomorphisms, but the reverse is not true. The present question is whether cognitive psychology can hope to distinguish between them given the methods of experimental, behavioral psychology.

Suppose that there are two models of cognitive representation, one first-order and one second-order isomorph of the external world. The first-order model specifies that external greenness is modeled by internal greenness and external squareness by internal squareness. The second-order model specifies that external greenness is modeled by internal squareness and external squareness by internal greenness. There are two points to be made. First, a second-order isomorphism cannot be distinguished from a first-order one without "looking inside the head" at the actual representing world. Moreover, just looking inside the head will not be sufficient to determine whether the representation is first- or second-order isomorphic to the world, because both models are characterized by just internal-greenness and internal-squareness. In order to test the two models, one must have access to the mapping function from the outside world. Does internal greenness correspond to external greenness or external squareness? Both looking inside the head and determining the mapping function are tasks for physiological psychology, not cognitive psychology.

The second point is that a second-order isomorphism is not necessarily any more plausible than a first-order isomorphism in physiological terms. The present example illustrates this obvious fact, assuming that no part of the nervous system is actually green or that, if it is, this fact is irrelevant to how it represents information. The only thing that matters is whether the postulated physical properties of the internal representation are consistent with the known physical properties of the nervous system. One can try to figure out what dimensions

could possibly be represented as first-order isomorphisms, but this is armchair speculation for cognitive psychologists. More than anything else, the notion of second-order isomorphism (in its general interpretation) is simply a philosophical comfort to cognitive psychologists, because it provides a justification for not worrying about precisely those issues involved in first-order isomorphism.

The Propositional/Analog Controversy

One of the most hotly debated issues of cognition these days is whether representations are "propositional" or "analog" in nature (e.g., Pylyshyn, 1973, 1975; Kosslyn, 1975b; Kosslyn & Pomerantz, 1977; Palmer, 1975b). The arguments began with the appearance of Pylyshyn's (1973) influential paper attacking the "picture metaphor" of visual imagery. The arguments seem to have spread to perceptually related representations in general. The terms "propositional" and "analog" have become emotionally charged buzz words capable of provoking arguments almost instantaneously.

Naturally, the entire controversy rests on the presupposition that propositional and analog representations are fundamentally opposed in some way that is relevant to cognitive theories and/or experiments. Whether this is true or not depends on how the terms are defined and what one takes to be the domain of cognitive psychology. In the following subsection, I define propositional representations and discuss their implications. Then I discuss a few of the concepts of analog representation that seem to be in current use. I also suggest a new way of looking at the notions of propositional and analog representation based on the distinction between intrinsic and extrinsic representation. To anticipate the conclusion, it turns out that propositions and analogs are fundamentally opposed but in ways that are not relevant for cognitive psychology. Further, I suggest that the reason for the controversy lies in differences between the two camps in terms of theoretical goals and styles.

Propositional Representation. The fundamental nature of propositional representation is quite simple. Recall the definition of representation: A world, X, is a representation of another world, Y, if at least some of the relations among objects of X are preserved by relations among objects of Y. This definition requires that any representation must have "object elements" to correspond to the represented objects. It might also have "relational elements" that model the relations (e.g., the arrows of World G in Fig. 9.1) or it might not (e.g., Worlds B–F in Fig. 9.1). Propositional representations are simply those in which there exist relational elements that model relations by virtue of themselves being related to object elements. The result is that relationships among n object elements cannot be determined simply by examining those n elements. One must determine them by examining their relationships to the additional relational elements.

Language is the paradigm case of propositional representation. Words referring to objects are related in syntactically ordered strings through relational words (verbs, prepositions, and the like). The sentence, "The ball is under the table," specifies a relationship between the ball and the table that can only be understood by virtue of their syntactic relationships to the relational construction "is under." Other kinds of propositional representations have the same basic properties. Lists of sentences in predicate calculus notation contain explicit relational elements and devices for specifying ordered-connectedness to them. Relational networks do also. Somewhat surprisingly, binary features are usually expressed propositionally. In a feature table, for example, the element representing the object is usually a row (or column) and the element representing the unary relation is usually a column (or row). The relational elements are "predicated" of the object elements by virtue of their being row–column connected by "+" or "1" rather than by a "–" or "0". In fact, there are many quite different-looking propositional representations in current use, including some with bizzarre notational devices (e.g., Schank, 1972; Leeuwenberg, 1971).

Analog Representations. The problem in attempting to characterize analog representations is that they seem to be different things to different people. I briefly characterize three notions that seem to be used most frequently.

The clearest and most obvious interpretation is that "analog" representations are those in which dimensions are continuous rather than discrete. The intended contrast is between analog and digital computers. Analog computers represent information in a physical dimension that, for all functional purposes, varies continuously (voltage). Digital computers do so in discrete, quantized units (bits). Although the continuous/discrete distinction may be of theoretical interest, it is not a question that seems to be answerable given state-of-the-art behavioral techniques. The cruder question of high versus low resolution within a given dimension is answerable, but having high resolution is not the same thing as being continuous.

The other two meanings for "analog" are both related to the visual imagery controversy. The claim is that visual images are, in some sense, "spatial" (Kosslyn, 1975a; Chapter 8 of this volume). The weaker version of this claim is that visual images preserve spatial information about that which they represent. For example, if object A is above object B in the represented world, then the representing world – whatever it is – will have objects A and B in some relationship that functionally corresponds to aboveness in the external world. Thus the weak spatial claim is equivalent to proposing that the image represents spatial dimensions in some fashion. This is a very mild and sensible position, one that is not opposed to propositional representation in any way.

The stronger spatial claim is, I suspect, the more usual one. In this interpretation, spatial information is not only preserved, but it is preserved (1) in a spatial medium, and (2) in such a way that the image *resembles* that which it represents.

For example, if A is above B in the external world, the strong claim would be that the representation of A is physically above (or perhaps below, if the image is inverted) the representation of B in the image. Although this is not a physiologically absurd type of first-order isomorphism, it is a first-order isomorphism nevertheless. Like all proposals of first-order isomorphism, it is functionally indistinguishable from informationally equivalent representations that are not first-order isomorphic.

Intrinsic Versus Extrinsic Representation. Although I have never heard anyone define "analog" in quite the way I now suggest, I suspect it is close to what most "analog" theorists have in mind. It is a weaker claim than physical isomorphism but a stronger one that functional isomorphism. Moreover, it does put analog and propositional representations in opposition to one another.

Recall that in discussing methods of preserving relational structure, we noted two different approaches. The intrinsic method was to model a represented relation or dimension by using a representing relation or dimension that has the same inherent structure as that which it represents. In such cases, the preservation of logical structure is a "natural" consequence of the representing relation or dimension chosen. The extrinsic method was to model the represented information using a relation or dimension that has no inherent structure, but to build the necessary structure into the system explicitly to conform to the represented world. The propositional/analog controversy makes sense if we associate analog representation with intrinsic methods and propositional representations with extrinsic methods. Let us see how this proposal ties in with what has been discussed so far.

The previous discussion of propositional representation was mainly concerned with surface manifestations. At a deeper level, the significance of using relational elements in representing relations is that propositions are extrinsic representations. The reason is that any object can, in principle, be connected to any relational element in any fashion. Hooking up object elements by relations to relational elements places no constraints whatsoever on the nature of the relations represented. Thus, *whatever structure there is in a propositional representation exists solely by virtue of the extrinsic constraints placed on it by the truth-preserving informational correspondence with the represented world.*

This fact is closely related to characterizations of propositional representations as "descriptive" and "interpreted" (Pylyshyn, 1973). The essence of a description is that it can be either true or false. There is nothing about descriptions that precludes contradiction with fact. One can *say*, for example, "A is above B," and "B is above A," although not both descriptions could be true of the relationship between B and A in the world. But if one is constrained to make only *true* statements, not both descriptions could be used. Thus, if descriptions are to serve as a representation, the constraints are external to the descriptive world. By definition, then, descriptions are extrinsic representations.

Analog representations are now seen to be those that contain no relational elements – i.e., nonpropositional representations. In such cases, the properties of individual represented objects are modeled by properties of individual representing objects, and relationships among sets of represented objects are modeled by relationships among sets of the corresponding representing objects. These representations are necessarily intrinsic, because the structure of the representing relations is inherent and therefore determines completely the kind of represented relations they can model. Thus, *whatever structure is present in an analog representation exists by virtue of the inherent constraints within the representing world itself, without reference to the represented world.*

Another way to view the nature of analog representation is in terms of a form of isomorphism somewhere between physical (first-order) and functional (second-order) isomorphisms. Recall that physical isomorphisms preserve information by virtue of the representing relations being themselves the same as their represented relations in a physical sense. Any physical isomorphism is analog (intrinsic), because the same relations must have the same inherent structure in both worlds, provided their operational definitions are constant across worlds. Now, suppose we relax the strict interpretation of physical sameness to allow representing relations to be physically the same in a more abstract sense. In the present framework, the sense in which representing and represented relations are the same is precisely that they have the same inherent structural constraints. We might call this concept "natural isomorphism" to emphasize that structure is preserved by the nature of corresponding relations themselves. Whatever one calls it, this concept is a stronger claim than functional (second-order) isomorphism, because the latter requires only a correspondence that preserves structure, regardless of whether it is done by intrinsic or extrinsic means. Thus propositional representations are functional isomorphisms but not natural isomorphisms. In general, any representation that is physically isomorphic is necessarily both naturally and functionally isomorphic to its represented world. Any representation that is naturally isomorphic is necessarily functionally isomorphic, but not physically isomorphic, to its represented world. Finally, any representation that is functionally isomorphic is not necessarily either naturally or physically isomorphic to its represented world. Thus there is a strict hierarchy of isomorphisms in which physical isomorphism is the most concrete and functional isomorphism the most abstract.

Relevance to Cognitive Psychology. If the distinction between intrinsic and extrinsic representation is actually the fundamental issue underlying the propositional/analog controversy, is it relevant to and resolvable by cognitive psychology? One can attempt to construct experiments to distinguish between the propositions and analogs on the basis of intuitive notions about the properties of "descriptive" versus "nondescriptive" or "interpreted" versus "noninterpreted" representations. It has been my experience that such experiments are never

convincing. Usually they are based on a simplistic notion of how the alternative type of theory might operate (see Kosslyn & Pomerantz, 1977, for some examples). Although such experiments may succeed in showing that analog-theory-X predicts better than propositional-theory-Y, the results do not seem to generalize beyond the particular examples.

The most compelling reason to believe that analog and propositional representations are not distinguishably different requires no experiments at all. Assuming that the intrinsic/extrinsic distinction is the fundamental issue, the answer to the controversy rests on the inherent nature of the representing relations and their relationship to the inherent nature of the corresponding represented relations. I see no way this can be determined without "looking inside the head." The concept of the inherent nature of representing relations concerns the physical medium that caries information. It is an *abstract* question about their physical nature, but one that concerns the physical medium nevertheless. Therefore, resolving the controversy over whether mental representation is analog or propositional is a task beyond the scope of cognitive psychology. It should be relegated to physiological psychologists, whose job it is to figure out the physical nature of inside-the-head and to determine its correspondence to outside-the-head.[4]

Approaches to Cognitive Models. There remains the interesting question of why the controversy arose in the first place. The answer provides some insights into different approaches to modeling in cognitive psychology.

There are two camps involved in the controversy. By and large, they divide cleanly in that proponents of the propositional view construct models by writing computer simulations, whereas proponents of the analog view do so by formulating analogies to known physical or formal systems. [Kosslyn is an exception, because he first worked by analogy (Kosslyn, 1973, 1975a; Kosslyn & Pomerantz, 1977) but has since simulated this analogy (Chapter 8 of this volume).] According to the analysis given here, this division is no accident.

[4]It should be emphasized that the argument is that analog representations are indistinguishable from propositional ones because they can be informationally equivalent as representations. It is possible that they differ in some nonrepresentational way that makes one preferable to the other. Perhaps testable differences exist in processing operations, for example. Whether this is true or not depends on the outcome of a rigorous analysis of the fundamental properties of processing operations. I suspect that such an analysis will yield a result parallel to the present one. That is, given that some operation has certain performance characteristics, it might be that these characteristics are a *necessary* consequence of the physical operation (i.e., intrinsic to the operation) or merely a *possible* consequence of the physical operation (i.e., extrinsic to the operation). If so, intrinsic and extrinsic methods of modeling performance characteristics are just as indistinguishable as intrinsic and extrinsic methods of preserving information. There are also other grounds for preferring one type of theory over another, such as parsimony, simplicity, and other pragmatic or esthetic considerations.

As Weizenbaum (1976) has pointed out, computers are systems in which the programmer is freed from the constraints of the physical world. He or she works in the domain of "possible worlds" that can be constructed within the computer. Therefore, when a program is intended to model something else, the programmer needs to know everything about those aspects of the modeled world that are of interest in order to simulate them. In the most general symbolic languages, virtually nothing can be hidden in implicit structure, because in the domain of possible computer-worlds, there is no structure in which it can be hidden. Thus, if a computer-modeled relation is transitive or connected, it is because either the programmer or something in the program makes it have these properties. Any unspecified aspect is "magic" until it can be spelled out precisely. Once the behavior of the simulated world is known to the required level of analysis, it can be simulated. Being able to predict new things not specifically built into the program is largely an accidental byproduct of constructing very complex programs. There is little or no intent to predict.

Modeling by analogy is quite different. The analogizer works in the domain of actual worlds that nearly always contain a great deal of structure. In some sense, the more structure that is hidden in the inherent constraints of the analogous system the better. This structure provides a rich base of possible future predictions from aspects of the system over which the modeler has no direct control. A good example of modeling by analogy is the notion of multidimensional spaces. If psychological concepts are conceived as points in a multidimensional space and similarity as distances between points, a great deal of structure is built into the model "accidentally." Of course, it is not really an accident at all but the mark of a clever theorist who has recognized at least some of the structural correspondences. In the spatial analogy, all of the dimensional features automatically have mutually exclusive values, because this is inherent in the axes of a space. In addition, similarity automatically has properties of symmetry, minimality, and the triangle inequality (see Tversky, 1977; Chapter 4 of this volume). That many of these structural properties are hidden is evidenced by the fact that the assumptions underlying the similarity-as-distance analogy have been tested only recently, more than a decade after the analogy gained popularity.

In short, computer modelers and systematic analogizers differ in the amount of structure they choose to have in their modeling medium. Computer modelers use an exceptionally flexible, unstructured medium. As a result, most structure must be built in explicitly to correspond to that of the modeled system. Analogizers work by choosing from among many rigid, highly structured media, each of which has constraints independent of its potential as a model of something else. It is not too surprising, then, that propositional models were developed by those who work with computers and that analog models were developed by those who work with more structured media. The amount of structure provided by the representing medium is precisely the difference between propositional (extrinsic) and analog (intrinsic) approaches to representation.

CONCLUSION

Representation is a complex and elusive concept, much more so than is generally supposed. Within psychology at least, it has been associated with an information-containing "thing" that is operated upon by processes. I used to believe that if I was looking at that "thing" in a diagram or illustration, then I knew the nature of that representation. The view presented in this chapter indicates this to be untrue. Our understanding of the current concepts in representation is based largely on superficial trappings that have little to do with their fundamental nature. If we are to make significant progress on the nature of cognitive representation, we need a deeper understanding of our theories.

One aspect of this deeper understanding is the realization that the representational nature of this "thing" cannot be dissociated from the operations that define the information it contains. Considering those operations may reveal that much of the information that *seems* to be there is not really there at all. Conversely, it may turn out that much information that does not seem to be there actually is. To determine what the representational nature of the "thing" is, one must first consider its functional information content as defined by those processes that use it. This is not to say that representation is indistinguishable from processing. There are many aspects of processing that are entirely independent of representational assumptions and other aspects that are only partly dependent on them. In general, however, operations are much more intricately woven into the fabric of representation than is usually acknowledged.

Once the information content has been discovered, it must be related back to the world it represents. In order for a "thing" to be a representation of any sort, it must preserve at least some information about its referent world. There is an important sense in which the nature of a representation is simply the view it presents of the represented world. Representations that provide the same view of the same world are at least informationally equivalent. Representations that provide views of the same world are not representationally equivalent, no matter how similar they may seem on the surface. The importance of the correspondence between represented and representing worlds should not be underestimated. More than anything else, it is what representation is all about.

For the purposes of cognitive psychology, there is yet another step in understanding our models of representation. Once the information has been discovered and related to the represented world, all other aspects of the representing world must be disregarded. In other words, the *only* thing that matters about the model of representation is what information it preserves about the representing world. Issues that pertain to any physical aspects of the representing world are simply beyond the scope of cognitive theories. The proper level of discourse for cognitive theory concerns information, not the medium used to carry it. It is this last step that allows us to see *through* the "thing" we find in illustrations to the representational assumptions that define the theory it embodies.

Perhaps the most general lesson to be learned from our discussions is that we cannot properly understand our theories and models of cognitive representation without some larger, metatheoretical framework in which to view them. The concepts currently used to talk about representation are seriously confused and inadequate. As a result, we lack the insight that allows us to separate relevant issues from irrelevant ones and to see the relationships among our models and theories in a clear and systematic way. I have attempted to provide the sort of framework I believe is necessary. It has helped me to notice things that were previously obscure and to clarify things that were previously confused. Once we understand the problems involved in cognitive representation properly, perhaps we can solve them more quickly.

ACKNOWLEDGMENTS

I would like to thank Amos Tversky, Donald Norman, Zenon Pylyshyn, and Elizabeth Bates for their helpful comments on earlier versions of this chapter. Most of all, I thank Eleanor Rosch for her constant encouragement; without it I would have given up on this project long ago.

REFERENCES

Cooper, L. A. Mental transformation of random two-dimensional shapes. *Cognitive Psychology*, 1975, *7*, 20–43.

Cooper, L. A., & Shepard, R. N. Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.

Garner, W. R. *The processing of information and structure*. Potomac, Md.: Lawrence Erlbaum Associates, 1974.

Garner, W. R. Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, 1976, *8*, 98–123.

Gibson, E. J. *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts, 1969.

Gibson, J. J. *The senses considered as perceptual systems*. Boston: Houghton-Mifflin, 1966.

Grice, H. P., & Strawson, P. F. In defense of dogma. *Philosophical Review*, 1956, *65*, 141–158.

Jacobson, R., Fant, G. G. M., & Halle, M. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, Mass.: M.I.T. Press, 1961.

Kosslyn, S. M. Scanning visual images: Some structural implications. *Perception and Psychophysics*, 1973, *14*, 90–94.

Kosslyn, S. M. Information representation in visual images. *Cognitive Psychology*, 1975, *7*, 341–370. (a)

Kosslyn, S. M. On retrieving information from visual images. In R. C. Schank & B. L. Schank & B. L. Nash-Webber (Eds.), *Theoretical issues in natural language processing*. Arlington, Va.: Tinlap Press, 1975. (b)

Kosslyn, S. M., & Pomerantz, J. R. Imagery, propositions, and the form of internal representations. *Cognitive Psychology*, 1977, *9*, 52–76.

Krantz, D. H., Luce, R. D., Suppes, P. H., & Tversky, A. *Foundations of measurement* (Vol. 1). New York: Academic Press, 1971.

Kuhn, T. S. *The structure of scientific revolutions.* Chicago: University of Chicago Press, 1962.

Leeuwenberg, E. L. J. A perceptual coding language for visual and auditory patterns. *American Journal of Psychology*, 1971, *84*, 307–350.

Lindsay, P. H., & Norman, D. A. *Human information processing.* New York: Academic Press, 1972.

Narasimhan, R. On the description, generation, and recognition of classes of pictures. In A. Grasselli (Ed.), *Automatic interpretation and classification of images.* New York: Academic Press, 1969.

Narasimhan, R., & Reddy, V. S. N. A generative model for handprinted English letters and its computer implementation. *ICC Bulletin*, 1967, *6*, 275–287.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Palmer, S. E. Visual perception and world knowledge: Notes on a model of sensory–cognitive interaction. In D. A. Norman, D. E. Rumelhart, & LNR Research Group, *Explorations in cognition.* San Francisco: Freeman, 1975. (a)

Palmer, S. E. The nature of perceptual representation: An examination of the analog/propositional debate. In R. C. Schank & B. L. Nash-Webber (Eds.), *Theoretical issues in natural language processing.* Arlington, Va.: Tinlap Press, 1975. (b)

Pylyshyn, Z. W. What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 1973, *80*, 1–24.

Pylyshyn, Z. W. Do we need images and analogues? In R. C. Schank & B. L. Nash-Webber (Eds.), *Theoretical issues in natural language processing.* Arlington, Va.: Tinlap Press, 1975.

Quine, W. V. Two dogmas of empiricism. *Philosophical Review*, 1951, *60*, 20–43.

Reed, S. K. *Psychological processes in pattern recognition.* New York: Academic Press, 1973.

Rips, L. J., Shoben, E. J., & Smith, E. E. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 1–20.

Rosch, E. H. On the internal structure of perceptual and semantic categories. In T. M. Moore (Ed.), *Cognitive development and the acquisition of language.* New York: Academic Press, 1973.

Rosch, E. H. Human categorization. In N. Warren (Ed.), *Advances in cross-cultural psychology* (Vol. 1). London: Academic Press, 1977.

Schank, R. C. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, 1972, *3*, 552–631.

Scott, D., & Suppes, P. H. Foundational aspects of theories of measurement. *Journal of Symbolic Logic*, 1958, *23*, 113–128.

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 1962, *27*, 125–140. (a)

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, 1962, *27*, 219–246. (b)

Shepard, R. N. Form, formation, and transformation of internal representations. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1975.

Shepard, R. N., & Chipman, S. Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1970, *1*, 1–17.

Shepard, R. N., & Metzler, J. Mental rotation of three-dimensional objects. *Science*, 1971, *171*, 801–703.

Shepard, R. N., Kilpatrick, D. W., & Cunningham, J. P. The internal representation of numbers. *Cognitive Psychology*, 1975, *7*, 82–138.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and processing in semantic memory: A feature model for semantic decision. *Psychological Review*, 1974, *81*, 214–241.

Sperling, G. A. The information available in brief visual presentation. *Psychological Monographs*, 1960, *74*(Whole No. 498).

Suppes, P. H., & Zinnes, J. L. Basic measurement theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, (Vol. 1). New York: Wiley, 1963.

Tarski, A. Contributions to the theory of models, I, II. *Indigationes Mathematicae*, 1954, *16*, 572–288.

Tversky, A. Features of similarity. *Psychological Review*, 1977, *84*, 327–352.

Weizenbaum, J. *Computer power and human reason.* San Francisco: Freeman, 1976.

Winston, P. H. Learning structural description from examples. In P. H. Winston (Ed.O, *The psychology of computer vision.* New York: McGraw-Hill, 1975.