Cognition 129 (2013) 138–162

Contents lists available at SciVerse ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

The GIST of concepts

Ronaldo Vigo *

Ohio University, Athens, OH 45701, United States

ARTICLE INFO

Article history:
Received 2 June 2012
Revised 14 May 2013
Accepted 15 May 2013

Keywords:
Categorization
Invariance
Complexity
Ideotype
Pattern detection
Concept learning

ABSTRACT

A unified general theory of human concept learning based on the idea that humans detect invariance patterns in categorical stimuli as a necessary precursor to concept formation is proposed and tested. In GIST (generalized invariance structure theory) *invariants* are detected via a perturbation mechanism of dimension suppression referred to as *dimensional binding*. Structural information acquired by this process is stored as a compound memory trace termed an *ideotype*. Ideotypes inform the subsystems that are responsible for learnability judgments, rule formation, and other types of concept representations. We show that GIST is more general (e.g., it works on continuous, semi-continuous, and binary stimuli) and makes much more accurate predictions than the leading models of concept learning difficulty, such as those based on a complexity reduction principle (e.g., number of mental models, structural invariance, algebraic complexity, and minimal description length) and those based on selective attention and similarity (GCM, ALCOVE, and SUSTAIN). GIST unifies these two key aspects of concept learning and categorization. Empirical evidence from three experiments corroborates the predictions made by the theory and its core model which we propose as a candidate law of human conceptual behavior.

1. Introduction

The ability to form concepts (i.e., sparse mental representations of multiplicity of entities in the environment and mind) lies at the very core of human cognition. Without it, humans would not be able to efficiently classify, organize, identify, nor store complex information – in short, humans would not be able to make sense of the world in which they live. Indeed, organisms in general would not be able to survive because a classification task as simple as distinguishing a poisonous food source from a non-poisonous one depends on having acquired an appropriately general representation (i.e., a concept) of non-poisonous food sources. In view of the basic role that concepts play in our everyday physical and mental lives, one of the ultimate goals of cognitive science has been to discover the laws that govern concept learning and categorization behavior and to characterize them with the same level of systematicity and rigor found in the physical sciences. To achieve this goal, several mathematical and computational models that aim to accurately predict the degree of learning difficulty of concepts of all types have been developed with different constructs and principles at their core.

For example, two of the most influential models of concept learning and categorization, the generalized context model (GCM; Nosofsky, 1984, 1986) and ALCOVE (Kruschke, 1992), are based on the basic principle that the mind forms concepts by determining similarities between different examples of the concept. More specifically, the GCM computes the probability that an object from a category of objects will be classified correctly by determining its similarity to remembered exemplars (i.e., memory traces of objects) whose categories have already been mentally encoded. Similarity between the exemplars is regulated by the assignment of attention weights to their dimensions. The resulting probability scores can be transformed into an overall percentage of classification errors for the particular category which then may be used to

* Tel.: +1 7405931707.
E-mail address: vigo@ohio.edu

0010-0277/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.05.008

operationalize how difficult it is to learn a concept from it (for details, see the appendix of Nosofsky, 1984). The same similarity and selective attention framework of the GCM was implemented in ALCOVE (Kruschke, 1992) as a three-layered, feed-forward connectionist network model of concept learning.

Similarity and selective attention also play a fundamental role in a more recent connectionist model named SUSTAIN (Love, Medin, & Gureckis, 2004). SUSTAIN functions by creating clusters in multi-dimensional space that represent a concept structure. Concept learning difficulty in SUSTAIN is a function of the number of clusters in the concept representation, where an increase in the number of clusters indicates an increase in learning difficulty. Although the GCM, ALCOVE, and SUSTAIN have been successful in predicting the well-known concept learning difficulty ordering of categorical stimuli consisting of four object-stimuli defined over three dimensions (Shepard, Hovland, & Jenkins, 1961), they have not been able to account for human classification performance with respect to large classes of categorical stimuli defined by logical rules on binary dimensions (Feldman, 2006; Goodwin & Johnson-Laird, 2011). One reason may be that they do not capture the specific kinds of patterns that humans are able to detect in categorical stimuli and that are necessary for concept formation.

This emphasis on stimulus structure was proposed by Gibson (1966) and explored by Garner in subsequent years (Garner, 1963; Garner and Felfoldy, 1970; Garner, 1974). Ever since, an abundance of laboratory experiments have supported the core idea suggested by these researchers that, in order to learn concepts, subjects extract rules from perceived patterns or regularities in categorical stimuli (Bourne, 1966; Estes, 1994). Inspired by this research and by the early probabilistic and connectionist models of categorization, alternative deterministic formal models of degree of concept learning difficulty have emerged which place the construct of complexity reduction or simplification at their core. These accounts have focused on concepts defined by functions from first order sentential logic and are known as *Boolean concepts*. Although several of these deterministic accounts have emerged recently, we shall focus on the three leading ones.

The first account, due to Feldman (2000), posits that since humans report forming rules when performing laboratory categorization tasks, one can then measure the degree of concept learning difficulty associated with a categorical stimulus by the length of the shortest logical rule that defines it. So, for example, if the rule $xy + xy'$ consisting of four literals (i.e., unprimed and primed variables) describes the category consisting of a black round object and a black square object (where x is black and y is round): such rule may then be reduced to a single literal rule x by the Boolean algebraic laws of distribution and complementarity. That is, we can factor out the variable x and cancel out y and its negation. This yields the single literal x which represents the minimal one dimensional rule "black", and this is the rule that we derive to make categorization decisions.

This proposal, referred to as "minimization complexity" (MinC), does not answer key questions about concept learning as a rule-oriented process: notably, what is the nature of the relational pattern detection process that must precede (and that is necessary for) the formation of efficient rules and simplification heuristics in the first place, and what are the limits of our capacity to detect such relational patterns? In other words, rule simplification procedures based on the symbolic calculus of Boolean logic should, but do not, give a deep rationale for why it is easier to form rules about certain sets of stimuli but not about others. Such a rationale is necessary to better understand why categorization performance is often inconsistent with rule-based accounts of concept learning (Vigo, 2006; Lafond et al., 2007). Indeed, MinC does not predict the canonical learning difficulty ordering of categorical stimuli consisting of four objects defined over three dimensions (Shepard et al., 1961). For a discussion of this and other challenges facing MinC the reader is referred to Vigo (2006).

A second and somewhat more elaborate structural approach referred to as the "Algebraic Complexity Model" (ACM) was introduced by Feldman (2006). The ACM describes how a Boolean concept may be decomposed algebraically into a "spectrum" of component patterns or regularities, each of which is a simpler or more "atomic" regularity. Regularities of higher degree represent more idiosyncratic patterns while regularities of lower degree represent simpler patterns in the original concept. The full spectral breakdown of the component patterns of a concept in terms of minimal component regularities is known as the power series of the pattern. These are expressed in terms of what are called "implication polynomials". The power spectrum of a Boolean concept is the number of minimal implication polynomials associated with the concept. The algebraic complexity of a concept is then defined by the weighted sum of its power spectrum where the weights increase linearly with respect to the degree of decomposition and the sum of their absolute value equals one. In other words, the weighted average of the complexity of these "atomic" rules (as measured by the number of variables they instantiate) per level of decomposition is a measure of the algebraic complexity of the concept.

Another model of concept learning difficulty that is based on complexity reduction has been proposed recently by Goodwin and Johnson-Laird (2011). The "number of mental models" model (NOMM) is based on the notion of a "mental model" which has found its greatest exponent in Johnson-Laird (1983). There has been a long debate concerning the nature, definition, and validity of the notion (e.g., Bonatti, 1994; O'Brien et al., 1994) so we shall not attempt to define it in depth here. Instead, we give the specific definition used by Goodwin and Johnson-Laird in their 2011 paper. They define the mental models representation of the extension of a Boolean concept as a disjunction of its possible instances, where each instance is represented as a conjunction of properties, though some of them may be negative. In other words, mental models have a conjunctive symbolic structure that is represented as a list of conjunctions of variables, where each variable stands for a particular property and each conjunction for a concept instance.

Take, for example, the concept representing a category containing two objects defined by two dimensions whose

structure is spelled out by the Boolean rule $xy + xy'$. The two mental models corresponding to this rule are represented symbolically in NOMM by two conjunctions of the two properties represented by the variables x and y, and the negation of the second property as follows: xy and xy'. The authors then give a simple heuristic for reducing the number of mental models. The basic idea is that eliminating "*irrelevant*" dimensions is at the heart of concept learning. The minimization heuristic employs two rules of Boolean algebra: distribution and complementarity. For example, y and y' cancel each other, making y irrelevant and leaving x as the only mental model left after the reduction. According to NOMM, the number of mental models we are left with after the application of this simple reduction heuristic is a good predictor of the degree of learning difficulty of a Boolean concept. In other words, the minimal number of mental models that a Boolean concept can be reduced to determines how difficult it is to learn it.

Note that NOMM's minimization heuristic is simpler than MinC's since it is consistent primarily with the use of only two simplification laws from Boolean algebra: distribution and complementarity. In contrast, MinC's heuristic achieves a greater variety of possible simplifications, as well as shorter minimal representations due to the use of more Boolean algebraic rules (for the rules involved in MinC, see Feldman, 2000). Note also that these two complexity reduction approaches are, in spirit, similar (not surprisingly, both models provide equally accurate fits to data from our classification experiment in Section 6). That is, both rely on a symbolic representation that is reduced by a set of simple Boolean algebraic axioms and/or meta-rules. Nonetheless, Goodwin and Johnson-Laird (2011) believe that the very nature of a mental model makes their model more "cognitive" than alternative proposals and suggest that the symbolic representation of a mental model should not be construed as the mental model: "...though actual mental models aren't strings of words but representations of situations in the world" (p. 42).

Notwithstanding, the three complexity reduction approaches discussed boil down to the use of Boolean algebraic rules on some symbolic representation of the Boolean concept. We propose that this approach does not reveal the implicit and structurally rich sub-symbolic laws, computations, and processes that make such symbolic calculi (and the generation of explicit symbolic rules) feasible in the first place. Researchers have attempted to reconcile these two aspects of a concept learning system. For example, Ashby and associates proposed a theory that describes how implicit subsymbolic processing and procedural explicit rule processing combined to facilitate concept learning (Ashby et al., 1998). Similarly, Pothos and Chater (2002) proposed "the simplicity model" which attempts to explain the process of simplifying sets of object-stimuli as a perceptual organization principle using cost functions on various aspects of cluster analysis. The model aimed to account for unsupervised categorization performance with respect to the way that humans will spontaneously divide into groups a set of objects.

Although these theories and their core models have contributed significantly to our understanding of the nature of concept learning beyond Boolean concepts, the discovery of a mathematically precise and elegant relational principle of concept learning that is sufficiently general to predict highly accurately the degree of learning difficulty of a wide range of category structures (e.g., defined with either discrete or continuous dimensions) remains an open problem. In this paper, we attempt to solve this problem with an alternative theory of concept learning named Generalized Invariance Structure Theory (GIST) that is a natural and direct descendant of categorical invariance theory (CIT; Vigo, 2009). The theory posits that the human conceptual system, at its core, functions as a particular kind of subsymbolic qualitative pattern detection system: namely, one that is sensitive to invariance. More specifically, the theory propounds that the process of concept formation necessitates the detection of qualitative patterns referred to as "invariants". These are revealed by a process named "dimensional binding" where dimensional values are temporarily suppressed. This mechanism, along with other key constructs of the theory, are explained in the next few sections.

1.1. Invariance and pattern detection

Many seminal ideas in the physical and mathematical sciences may be traced to invariance principles. Loosely speaking, invariance is the property of entities to stay the same in some respect after undergoing some transformation or change. Invariance is all around us in obvious and not so obvious ways: for example, when a piece of paper is crumpled, neither its color nor weight changes. Similarly, when a melody is transformed from one key to another, or played by different instruments, we recognize it as the same melody because several of its characteristics are invariant in respect to such changes. In short, invariance is permanency or constancy in change. Perhaps because it facilitates the predictability of events and attributes, and because it facilitates the identification of key features of the objects in the complex world around us, the ability to detect invariance is of paramount importance to the human cognitive system.

Some cognitive scientists have suggested that invariance plays a role in higher level cognition (Garner, 1963, 1970; Garner & Felfoldy, 1970; Leyton, 1992). However, these suggestions have been limited in three respects. First, they have lacked generality because they have been based on ideas of coherent geometric or spatial structure (see, for example, Shepard, 1984), such as symmetry, that limit their scope to spatial domains. Secondly, they were not designed to account for key results in the concept learning and categorization literature. Thirdly, the majority have lacked mathematical and/or computational precision. These limitations were addressed in CIT (Vigo, 2009) and have been inherited by GIST, the theory introduced in the next few sections.

GIST and its core model overcomes several stumbling blocks for theories of concept learning: namely, (1) it predicts the learning difficulty ordering of the six key category structures corresponding to the class of categorical stimuli with four objects defined over three dimensions (Shepard et al., 1961); (2) it is able to accurately account for the

learnability of complementary stimulus structures (i.e., in "down parity"); (3) it accurately fits the data and, more generally, accounts for the learnability of an unprecedented number of 84 categorical stimulus structures tested in Experiment 1 ($R^2 = .91$, $p < 0.0001$); (4) using a scaling parameter, it has the potential to account for key individual differences in classification performance; (5) it introduces an original mathematical and deterministic framework for the study of concept learning behavior and cognition in general; (6) it predicts the learning difficulty ordering of categories defined over *multivalued*, *dichotomous*, *and continuous* dimensions; (7) it unifies in precise quantitative terms key and ubiquitous constructs in universal science – such as pattern detection, symmetry, invariance, similarity, and complexity – from the perspective of concept learning research. With respect to these seven points, GIST outperforms the most successful models to date.

1.2. Informal terminology

We begin by informally defining some terms. By a categorical stimulus we shall mean a set of dimensionally definable objects that, by virtue of sharing dimensions, are related in some way. Concepts, on the other hand, we shall define roughly as sparse mental representations of categorical stimuli. Accordingly, categorical stimuli are the raw material from which concepts are learned or formed.

Dimensionally-definable stimulus objects are objects that can be characterized in terms of a fixed number of shared attributes or properties (i.e., dimensions), each ranging over a continuum or over discrete values. For example, the properties of brightness, shape, and size, as well as the more subjective attributes of satisfaction and personal worth, are all possible dimensions of the objects of some categorical stimulus. In addition, we shall assume that all of the dimensions associated with a specific categorical stimulus range over a specific and fixed number of values that combined specify a gradient (standardized in the [0,1] interval) for the particular dimensions. For example, the brightness dimension may have five fixed values representing five levels of brightness on a continuum standardized from 0 to 1 (from least bright to most bright). This continuum may be established empirically ahead of the concept learning experiment by eliciting judgments about the degree of the attributes of the objects comprising the categories. In fact, we employed this procedure in part 1 of Experiment 2. On the other hand, we shall assume that whenever the dimensions range, in principle, over an infinite number of possible values, it simply means that they are continuous.

Since one of the goals of this paper is to introduce a more general theory of concept learning than theories that exclusively pertain to Boolean concepts, we introduce notation to designate the various kinds of non-Boolean concepts that should be accounted for by a general theory. We shall say that a concept associated with a categorical stimulus consisting of p objects defined over D dimensions, and for which the dimensions have n standardized values in the real number interval [0,1], is a $D_n[p]$ type concept. Furthermore, we shall say that such categorical stimulus is a member of the $D_n[p]$ class of categorical stimuli. Please note that p refers to the specific number of objects in a given categorical stimulus and not to the total number of objects in both the given categorical stimulus and a negative or complimentary categorical stimulus (a complimentary categorical stimulus consists of those objects definable by the same given dimensions but that are not in the categorical stimulus). Also, as mentioned, in the extreme case that the dimensions of the stimulus have an infinite number of possible values, we let $n = \infty$. For example, a categorical stimulus with four objects whose four dimensional values lay on a continuum is said to belong to the $4_\infty[4]$ class of categorical stimuli. In general, the greater the number of dimensions and dimensional values used to define a categorical stimulus, the more ill-defined it will appear to be. Indeed, our ultimate aim is to develop a theory of classification performance on dimensionally-defined category structures ranging from the Boolean variety, to the semi-continuous, and the continuous variety.

Six examples of categorical stimuli of the $3_2[4]$ class of categorical stimuli and consisting of objects defined over the three separable binary dimensions of color, shape, and size are given in Fig. 1b. Note that each of the six categorical stimuli contains four objects and has a certain structure, which is to say that each displays a specific relationship between its dimensional values. There are exactly six possible structures associated with the $3_2[4]$ class of categorical stimuli (see Higonnet & Grea, for a proof). These six unique structures are represented by the six cubes of Fig. 1a. We shall call the set of structures corresponding to any $D_n[p]$ class of categorical stimuli "the $D_n[p]$ structure family". On the other hand, the categorical stimuli conforming to each structure are called its structure instances.

One of the ultimate goals of theories of concept learning difficulty is to accurately account for the different degrees of learnability associated with the different types of structures corresponding to different classes of categorical stimuli. In addition, due to their specific binary dimensional nature, they may be represented by Boolean algebraic expressions or, simply stated, logical rules (i.e., expressions consisting of disjunctions, conjunctions, and negations of variables that stand for binary dimensions). These algebraic representations of a categorical stimulus are referred to as *concept functions*. Concept functions are useful in spelling out the logical structure of a stimulus set. For example, suppose that x stands for blue, x' stands for red, y stands for round, and y' stands for square, then the two-variable concept function $(x' \cdot y) + (x \cdot y')$(where "+" denotes "or", "$\cdot$" denotes "and", and "x'" denotes "not-x") defines the category which contains two objects: a red and round object and a blue and square object. Clearly, the choice of labels in the expression is arbitrary. Hence, there are many Boolean expressions that define the same category structure (for a detailed explanation see Vigo, 2006).

The six category structures depicted in Fig. 1a and b were studied empirically by Shepard, Hovland, and Jenkins (1961) who observed the following increasing concept learning difficulty ordering between the six structures based on classification performance: I < II < [III, IV, V] < VI

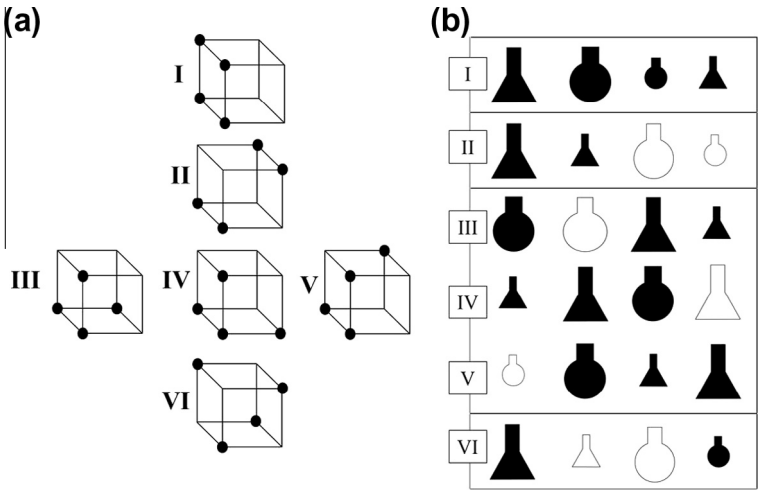

Fig. 1. Examples of the $3_2[4]$ category structures and structure instances used in Experiment 1. (a) Shows the stimuli of each type denoted by the corners of a cube where the sides of the cube represent dimensions. Corners with circles represent positive examples whereas empty corners are negative examples of the category. (b) Shows examples of structure instances used in Experiment 1 for each structure type. These consist of four flasks defined over three dimensions (in this example: color, shape, and size).

(with types III, IV, and V of approximately the same degree of learning difficulty). This ordering, which henceforth we shall refer to as the "SHJ ordering", has been empirically replicated numerous times by several researchers (Kruschke, 1992; Shepard et al., 1961; Nosofsky et al., 1994a; Love & Medin, 1998), but has been challenging to predict quantitatively. Throughout this article, we shall revisit this empirical result which has become an essential benchmark for theories of classification performance.

In addition to the $3_2[4]$ family ordering, another benchmark for classification performance may be found in a more recent and broader study by Feldman (2000). He observed an approximate empirical difficulty ordering for 76 category structures from the $3_2[2]$, $3_2[3]$, $3_2[4]$, $4_2[2]$, $4_2[3]$, and $4_2[4]$ families along with their complementary "down parity" counterparts. A category is in down parity whenever it has more objects than its complementary category (Feldman, 2000); otherwise, it's in "up parity". Note that the complement of a category is the set of objects that are also definable by D dimensions but that are not in the category.

Finally, we contrast between two experimental paradigms in the concept learning literature. Under the first and more popular paradigm, subjects are asked to classify, in succession, members of a category and a contrasting category without prior knowledge of the category. Henceforth, we shall call such empirical situations *serioinformative* (*serio* for serial) to distinguish them from their *parainformative* counterparts (*para* for parallel) where a category and its contrasting category are shown simultaneously for a certain amount of time before their members are displayed sequentially during the classification phase. While the majority of the well-known concept learning experimental paradigms are serioinformative in nature (e.g., Kruschke, 1992; Shepard et al., 1961; Nosofsky, et al., 1994a; Love & Medin, 1998), and thus, of a primarily inductive character, there have been a significant number of experiments of a parainformative nature in the literature (for examples see Haygood and Bourne, 1965; Garner, 1974; Feldman, 2000). For the remainder of this paper we shall focus exclusively on parainformative tasks as the target of our models, predictions, and tests.

2. Categorical invariance theory

In this section we lay the foundation for a general theory of concept learning named "Generalized Invariance Structure Theory" (GIST). Its core model, the *generalized invariance structure theory model* (GISTM), and its variant, the GISTM-SE will be derived using the construct of categorical invariance introduced in CIT (Vigo, 2009). Appropriately, we begin with an introduction to CIT. CIT was developed primarily as a *stimulus-oriented* theory of conceptual behavior where the degree of concept learning difficulty of a categorical stimulus is modeled quantitatively as the ratio between its size and its degree of *categorical invariance* (to be explained). This idea is now known as the *structural invariance* or *categorical invariance* model (CIM, Vigo, 2009). In contrast, GIST is an *observer-centered* theory that focuses on the ability of observers to detect invariance patterns in categorical stimuli.

GIST uses the mathematical notion of categorical invariance introduced first in CIT (Vigo, 2009, 2011a, 2011b) to describe the kinds of patterns that humans are sensitive to in categorical stimuli. To understand how categorical invariance and its related invariance measure work, consider a simple example. The categorical stimulus consisting of a triangle that is black and small and a circle that is black and small and a circle that is white and large is described by the concept function $xyz + x'yz + x'y'z'$ (note that, for readability, we have eliminated the symbol "·" representing "and"). Let's encode the features of the objects in this categorical stimulus using the digits "1" and "0" so that each object may be represented by a vector of zeros and ones. For

example, the vector (1, 1, 1) stands for the first object when $x = 1 =$ *triangular*, $y = 1 =$ *small*, and $z = 1 =$ *black*. Thus, the entire categorical stimulus can be represented by C = {(1, 1, 1), (0, 1, 1), (0, 0, 0)}. If we perturb this categorical stimulus with respect to the shape dimension (dimension 1) by assigning the opposite shape value to each of the object-stimuli in the set, we will get the perturbed categorical stimulus $T_1(C) = \{(0, 1, 1), (1, 1, 1), (1, 0, 0)\}$ which indicates a transformation along the first dimension. More generally, if C is a categorical stimulus defined over D binary dimensions (where $D \geqslant 1$), then, for any dimension i ($1 \leqslant i \leqslant D$), the transformation T_i on C is defined as follows: $T_i(C) = \{(x_1, \ldots, x_i', \ldots, x_D) | (x_1, \ldots, x_i, \ldots, x_D) \in C\}$ where $x_i' = 1$ if $x_i = 0$ and $x_i' = 0$ if $x_i = 1$.

Now, if we compare the original categorical stimulus to the perturbed set, we see that they have two object-stimuli in common. The object-stimuli that "survive" the transformation (from the standpoint of membership in the categorical stimulus) are called "invariants". Thus, two out of three objects are invariants or remain the same. Intuitively, the ratio of invariants to number of objects in the categorical stimulus may be construed as a measure of the partial homogeneity of the categorical stimulus with respect to the dimension of shape and can be written more formally as $|C \cap T_i(C)|/|C|$. In this expression, |C| stands for the number of object-stimuli in the categorical stimulus C and $|C \cap T_i(C)|$ for the number of object-stimuli that C and its perturbed counterpart $T_i(C)$ share (Vigo, 2009).

The first pane of Fig. 2 illustrates this qualitative transformative process. Doing this for each of the dimensions generates the partial invariance scores of the categorical stimulus C which are then arranged as a vector referred to as the *logical manifold* Λ of the Boolean categorical stimulus C as shown by Eq. (1).

$$\Lambda(C) = \left(\frac{|C \cap T_1(C)|}{|C|}, \frac{|C \cap T_2(C)|}{|C|}, \ldots, \frac{|C \cap T_D(C)|}{|C|} \right) \tag{1}$$

Each component of this vector represents a degree of partial invariance for the categorical stimulus. But we also wish to characterize the global or total degree of gestalt invariance of the categorical stimulus. We do this by taking the Euclidean distance of each logical manifold from the zero logical manifold whose components are all zeros (i.e., $\mathbf{0} = (0, \ldots, 0)$). Thus, the overall degree of categorical invariance Φ of a categorical stimulus C defined over D binary dimensions is given by Eq. (2) (where $|C| > 0$:

$$\Phi(C) = \left[\sum_{i=1}^{D} \left[0 - \frac{|C \cap T_i(C)|}{|C|} \right]^2 \right]^{1/2} = \left[\sum_{i=1}^{D} \left[\frac{|C \cap T_i(C)|}{|C|} \right]^2 \right]^{1/2} \tag{2}$$

Intuitively, degree of categorical invariance may be construed as a measure of the overall gestalt homogeneity of the categorical stimulus or its coherence. Furthermore, the larger the degree of categorical invariance of a categorical stimulus, the easier it should be to learn a concept from it. Using our example from the first pane of Fig. 2, we showed that the original categorical stimulus and the perturbed categorical stimulus have two elements in common (out of the three transformed elements) with respect to the shape dimension; thus, its degree of partial invariance is expressed by the ratio 2/3. Conducting a similar analysis with respect to the dimensions of color and size, its logical manifold computes to $\left(\frac{2}{3}, \frac{0}{3}, \frac{0}{3}\right)$ and its overall degree of categorical invariance is given by Eq. (3).

$$\Phi(\{(1,1,1),(0,1,1),(0,0,0)\}) = \sqrt{\left(\frac{2}{3}\right)^2 + \left(\frac{0}{3}\right)^2 + \left(\frac{0}{3}\right)^2} = .67 \tag{3}$$

There are two glaring limitations of the above formal framework of invariance extraction as a psychological theory. The first is that it applies only to categorical stimuli defined over binary dimensions. The second is that the idea of perturbing categorical stimuli may seem psychologically implausible as a theory of how humans detect categorical invariants. GIST overcomes both limitations in the following section by reconceptualizing the invariance pattern extraction process described here in terms of the capacities of discrimination, similarity assessment, attention, and short-term memory. However, this more psychologically plausible and observer-oriented description should be construed as a high-level cognitive description of how people accomplish the discussed computations and not as a

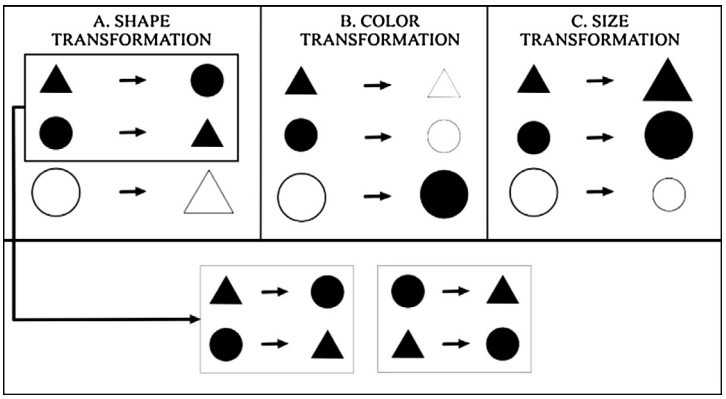

Fig. 2. Logical manifold transformations across the dimensions of shape, color, and size for a $3_2[3]$ type structure instance.

detailed process account describing precisely the concept learning mechanisms and representations underlying the human conceptual system.

3. Generalized invariance structure theory

The central working hypothesis of GIST is that humans detect optimally (i.e., in a way that maximizes classification performance within the limits of their cognitive capacities), categorical invariants with respect to each dimension of a categorical stimulus. Invariants were precisely described in the previous section. An example of a pair of invariants (note their pairwise symmetry) is shown in the bottom pane of Fig. 2 for a very simple stimulus consisting of a small white triangle and a small white circle. In GIST, the proportion of invariants (with respect to a particular stimulus dimension) to the number of objects in the categorical stimulus is referred to as a *structural kernel* or SK. Each structural kernel is simply each individual component of the sum in Eq. (2). Furthermore, each structural kernel is interpreted as a measure of the degree of partial (with respect to a particular dimension) homogeneity of a categorical stimulus. We shall explain the connection between invariance and homogeneity later in this section.

There are as many SKs corresponding to a categorical stimulus and its corresponding concept as the number of dimensions that defines it. As we shall see, SKs carry critical information about the classification potential of each dimension of the categorical stimulus. The SKs of a categorical stimulus are stored as a compound memory trace referred to as an *ideotype*. Ideotypes contain the essential structural information of categorical stimuli and are represented by points whose coordinates are the values of their SKs in a psychological space (see Fig. 3 to visualize). Because an ideotype carries only structural information, a single ideotype may correspond to different categorical stimuli. Indeed, in this sense, an ideotype may be regarded as a type of meta-concept. In addition, the distance in psychological space between the ideotype of a categorical stimulus and the zero ideotype (i.e., the ideotype containing only zero SKs and indicating a total lack of invariants) determines the degree of overall gestalt homogeneity or coherence of the categorical stimulus that it encodes.

In the previous section, we defined a mathematical operator (the logical manifold operator Λ) that characterized in precise mathematical terms the detection of invariants and the formation of logical manifolds. However, the process required the perturbation of categorical stimuli. In GIST, we reconceptualize the operator Λ as a mental operator that is capable of generating SKs and ideotypes. We do this by reducing the process of the detection of invariants to a process grounded on the ubiquitous cognitive capacities of similarity and attention. Specifically, invariants are detected by first determining how the suppression of each dimension defining the categorical stimulus influences the similarity between its members. This is achieved by a simple goal-oriented attention process that we shall refer to as *dimensional binding*. In effect, dimensional binding now becomes the perturbation mechanism of invariance detection defined in CIT (Vigo, 2009) and described in Section 2.

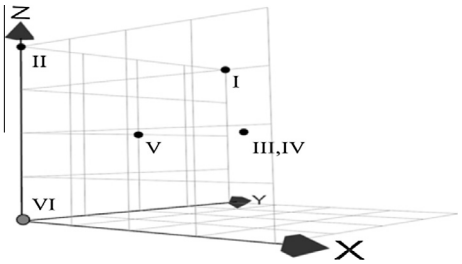

Fig. 3. Cartesian coordinates representation of ideotypes corresponding to the $3_2[4]$ class of categorical stimuli graphed in ideotype (psychological) space. Note that type VI, the most difficult to learn and the one perceived to be the most difficult to learn, coincides or is closest to the zero ideotype.

In order to better understand the process underlying dimensional binding, we turn to the debate concerning the nature of selective visual attention. Several researchers have argued and presented empirical evidence (Allport, 1987; Neumann, 1987; Prinz, 1983; Schneider, 1993) in support of the idea that attentional selection should not be viewed solely as a limited capacity of the human visual system, but more as a facility for constraining possible cognitive actions. The basic idea underlying this view is that our sensory system is able to detect many different stimuli simultaneously, but our higher-level cognitive system is normally limited to carrying out actions serially. Under this view of goal-directed control of attention, attentional processes are needed to constrain the selection of the appropriate action on the basis of the incoming information. For example, Neumann (1987) writes that this type of attention prevents "the behavioral chaos that would result from an attempt to simultaneously perform all possible actions for which sufficient causes exist" (p. 347).

With this notion of goal-directed attention in mind, we interpret the detection of invariants and SKs in a categorical stimulus as a "partial similarity" assessment process involving three levels of attention on the stimulus set. First, at the stimulus-object level, the observer rapidly shifts her attention from one pair of objects to another to select the objects that will be compared next. Second, after a pair of objects have been selected, the observer will temporally *bind* (i.e., disregard or suppress) one of the dimensions. Third, while in binding mode, the observer will pay attention to the "free" (non-bound) dimensions and assess the similarity between the two objects based on these free dimensions. This three-stage attention-regulated process of assessing similarities is significantly different from that found in other concept learning models based on similarity assessment (e.g., Nosofsky's GCM, 1984 and Kruschke's ALCOVE, 1992) – the key difference being the way that dimensional binding changes the character of similarity assessment.

Dimensional binding constrains similarity assessment in a way that is equivalent to the process of extracting invariance patterns in categorical stimuli described in Section 2 (see Fig. 2). More specifically, perturbing a dimension of a pair of object-stimuli in a set, and determining whether the generated pair of objects remained in the set, is equivalent to determining, after a dimension has

been bound, whether the two distinct objects in the set are identical with respect to their free dimensions. If so, then they are categorically-invariant: that is, no matter what the value of the bound dimension is, the objects remain the same in the sense that they both continue to be members of the (unperturbed) categorical stimulus. On the other hand, if they are not identical, then they will not be members of the categorical stimulus. We refer to this relationship as the *invariance-similarity equivalence principle*. Fig. 4 illustrates this equivalence between the categorical invariance detection process and the partial similarity comparison between two objects with respect to the bound dimension of shape. The mathematical theory describing the equivalence between invariance structure detection and the detection of partial similarities via dimensional binding has been sketched in the technical Appendix B and discussed in greater detail in the supporting documents website.

Note that now we can clarify what was meant at the beginning of this section by the statement that the structural kernels (representing proportion of invariants) are measures of degree of partial homogeneity. To illustrate, consider that the simple categorical stimulus of Fig. 2 consisting of three objects is partially homogeneous (objects are partially alike) when their shape dimension is ignored. Specifically, only two out of the three objects are identical when we bind the shape dimension. Likewise, if we bind the color dimension, no objects look alike in the categorical stimulus. This alternative theory of the invariance detection process, based on dimensional binding and similarity assessment, does not have the limitations of the theory presented in Section 2. Indeed, the theory is now psychologically plausible and it applies to a wider range of categorical stimuli beyond the Boolean variety. The "mental" Λ operator (henceforth referred to as the structural manifold operator) will now stand for a more general invariance detection operator; one that extracts SKs and ideotypes from categorical stimuli defined over binary, multivalued, and continuous dimensions via the cognitive operations of similarity assessment and dimensional binding.

In GIST, the processes of SK detection and ideotype formation inform two distinct systems respectively: the first is responsible for the formation of symbolic rules, the second for the formation of magnitude judgments (or more accurately, meta-judgments) about concept learnability. Note that in this article we will not discuss the mechanisms underlying these secondary systems (however, we refer the reader to RULEX (Nosofsky, Palmeri, & McKinley, 1994b) for an example of the kind of cognitive mechanism that may be at play after the degree of diagnosticity of each dimension is determined by the core model in GIST). Instead, we explain how and why ideotypes and SKs act as precursors to other types of concept representations. First, the perceived degree of learning difficulty of a categorical stimulus is a function of the distance between its ideotype and the zero ideotype in psychological space. The shorter this distance is, the less homogenous the categorical stimulus is perceived to be and, consequently, the more difficult it is judged to be from the standpoint of concept formation.

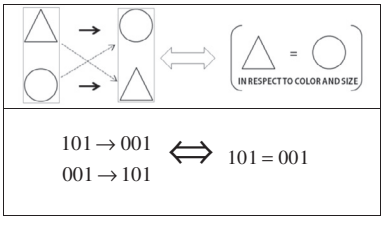

Fig. 4. Equivalence of invariance to partial similarity across two dimensions. The two objects (a large white triangle and a large white circle) are identical when an observer binds the shape dimension. This renders the shape dimension as redundant because it cannot determine set membership (i.e., both objects belong in the set). Note the qualitative symmetry between the two invariant objects (i.e., invariants) indicated by the pair of broken arrows. These are referred to as "invariant-symmetries".

Secondly, ideotypes are precursors to rules because the SKs of the ideotype contain necessary information for rule formation about the degree of "diagnosticity" (the quality of completely determining category membership) and "redundancy" (the quality of playing no role in determining category membership) of the dimensions on which the categorical stimulus is defined. But how do SKs determine the degree of diagnosticity of the dimensions? Recall that the SKs are generated by the process of invariant detection. If the act of binding a dimension results in absence of invariants, this signals that the categorical stimulus and its perturbed counterpart do not have any object-stimuli in common. This means that the particular dimension (e.g., color) can perfectly determine membership in the categorical stimulus for any of it objects and any of the objects of its perturbed counterpart. Accordingly, as the number of invariants increase, the less diagnostic the dimension becomes and the more it becomes redundant or non-essential. Thus, the degree of diagnosticity of the dimension can be precisely characterized by the proportion of invariants associated with it.

For example, take the two object category of Fig. 4. Note that when the categorical stimulus is perturbed or, equivalently, bound on the shape dimension, two invariants emerge (i.e., namely the entire content of the categorical stimulus). Thus, the shape dimension cannot be used as a basis for distinguishing between the original and the perturbed categorical stimulus. Although dimensions with correspondingly high SKs do not carry diagnostic information about the categorical stimulus that they define, they signal the presence of degrees of redundant information that may be eventually eliminated. For example, take the first categorical stimulus shown on Fig. 1b consisting of three dimensions: color, size, and shape. This is also known as Type I of the $3_2[4]$ family of structures. Clearly, this is a simple stimulus with its most relevant or diagnostic dimension being color (and in particular, the color black). Now, applying the structural manifold operator on this categorical stimulus yields the (0,1,1) ideotype which has three SK values (one per dimension in the following order: color, shape, and size).

Note that the first dimension is the most diagnostic because it has the lowest proportion of invariants associated with it: namely, zero. The other two dimensions have ker-

nel values of one, indicating that they are fully redundant or dispensable because they have the highest proportion of invariants possible (i.e., transforming the categorical stimulus objects along these two dimensions yields the same original categorical stimulus). One of the core assumptions in GIST is that our conceptual system strongly favors and is greatly biased toward the extreme SK values of 0 and 1 because of their greater utility in forming clear cut rules. This partly stems from the fact that it is harder to ascertain the exact degrees of partial invariance or homogeneity that lie between these two extreme values.

The key points in this section may be summarized in terms of the following two principles that link categorical invariance to concept learning difficulty:

I. *Concept–invariance principle*: The more sensitive an organism is to the invariants of a stimulus set, the easier it is for the organism to learn a concept from it and to determine its learnability.
II. *Invariance–parsimony principle*: For any given ideotype, the extreme SK values of 0 and 1 have a much greater relative impact on the perceived learnability of a concept than the SK values between 0 and 1.

In addition to these two principles, we will adopt the following principle, referred to as the *invariance-learnability principle*, about the nature of the learnability of categories: the percentage change in the perceived initial raw complexity of a categorical stimulus relative to said initial raw complexity is negatively proportional to the degree of perceived invariance of the stimulus or, in other words, its gestalt homogeneity (where the perceived initial raw complexity of a categorical stimulus is measured by the number of objects it contains). For a more detailed explanation of this principle the reader is referred to Appendix A. From this principle and from the mathematical description of the similarity-based mechanism underlying invariance extraction, we formally derive a simple candidate mathematical law of conceptual behavior in Appendix A: namely, that the degree of subjective learning difficulty ψ of a categorical stimulus X is directly proportional to its cardinality or size p and inversely proportional to the exponent of its perceived gestalt homogeneity as measured by categorical invariance (see Eq. (4)). In Eq. (4), the hatted Greek capital phi $\widehat{\Phi}$ stands for the more general and process-oriented measure of degree of categorical invariance (i.e., one which applies to categorical stimuli defined over dichotomous, semi-continuous, and continuous dimensions) based on the "mental" structural manifold operator Λ introduced formally in Appendix B and discussed informally at the beginning of this section.

The scaling parameter k ($0 \leqslant k < \infty$) is a distance scaling parameter reflecting overall discriminability (or the capacity to discriminate overall) in the ideotype psychological space (e.g., the ability to discriminate ideotypes of different dimensionality in ideotype space). It accounts for differences in categorization performance on categories with different number of dimensions when their degree of invariance is equal and non-zero. Under this interpretation, only one parameter value is required for all ideotypes of the same dimensionality.

$$\psi(\mathrm{X}) = p e^{-k\widehat{\Phi}(\mathrm{X})} \tag{4}$$

In addition, Eq. (4) is a complete generalization (via the generalization of degree of invariance $\widehat{\Phi}$ of a similar equation introduced in Vigo (2009, 2011a). We refer to it as the ECIM (*exponential categorical invariance model*). It should be construed as a steady-state behavior model (Bush, Luce, & Rose, 1964) or a phenomenological description (Luce, 1995) of the role that invariance pattern information plays in the concept learning process. Support for an exponential function of invariance comes from converging evidence. First, as mentioned above, it is consistent with the invariance-learnability principle and other basic assumptions in GIST (for a proof see Appendix A). Second, the exponential relation provides the overall best fits (when compared to other mathematical forms) for the data from two large scale classification experiments: the Feldman (2000) study of 76 category structures and our own experiment labeled Experiment 1 under the methods section involving 84 category structures. Indeed, in other areas of cognitive research, candidate laws have been proposed solely on the basis of accurate data fits and criticized for this very reason. In particular, Steven's power law (Stevens, 1955) comes to mind. Third, the mathematical framework in Appendix B suggests that there may be a strong exponential link to invariance via a lower-level type process of generalization. This suggests a possible connection to Shepard's law of universal generalization.

An intuitive way of interpreting the exponential GISTM concerns the tradeoff between the initial "perceived raw complexity" of a categorical stimulus and its degree of perceived invariance. In particular, the invariance-learnability principle articulates this tradeoff precisely in terms of a "percentage change" in order to establish an initial, scale-independent, upper bound level of perceived raw complexity that is then reducible by the degree of homogeneity of the categorical stimulus. Although we measure the perceived raw complexity of a categorical stimulus before it is processed as the number of items it contains, more psychological measures, such as the total number of pairwise comparisons between its objects (i.e., p^2), are also plausible. Regardless of the raw complexity measure chosen, this quantity is subject to exponential decay (as degree of homogeneity increases) if it is decreased at a rate proportional to its value (see Appendix A).

Although, Eq. (4) yields accurate fits to data from key historical experiments and our own experiments (described under the methods section), the invariance-parsimony principle, which states that the SKs with 0 and 1 values should play a much greater role in determining degree of difficulty, is not satisfied by the simple exponential model of Eq. (4). We remedy this and other limitations by adjusting the basic functional form of the model in Section 4.

4. The generalized invariance structure theory model

The first additional extension is based on the previously discussed invariance-parsimony principle involving SKs: namely, humans highly favor SK-values of 0 and 1 and in-between values play a disproportionate lesser role in

lowering perceived difficulty. In other words, there is an inherent strong bias toward the dimensions whose kernels are zeros or ones. We capture this disproportional contribution of the kernel values by squaring the degree of invariance. Squaring works because the contributions of the kernels with values of 1 or 0 to the overall degree of invariance are amplified when compared to the contributions of all the SKs with in-between values (for a formal justification and discussion of this assumption and notion see Vigo, 2009); the power of two is the smallest positive integer power to accomplish this.

Ultimately, squaring the degree of categorical invariance transforms the exponential form of the function to a Gaussian form as shown in Eq. (5) where X is a continuous or dichotomous category X, ψ is degree of subjective learning difficulty (aka, subjective complexity), p is the number of objects in the categorical stimulus, and k is a scaling parameter interpreted in the same way as in our discussion of Eq.(4).

$$\psi(\mathrm{X}) = pe^{-k\widehat{\Phi}^2(\mathrm{X})} \tag{5}$$

Note that letting the scaling parameter k above be the discrimination index defined by $k=D_0/D$ (where D is the number of dimensions of the categorical stimulus and $D_0=2$ a lower bound on the number of dimensions), gives a non-parametric version of Eq. (5) (named GISTM-NP) with virtually the same fits to our data (see Figure 6C)[1]. For a discussion of the non-parametric variants of the GISTM, the reader is referred to the supporting documents website. Next, we discuss another variant of Eq. 5 that introduces a potentially powerful new construct for the analysis of concepts and concept learning.

4.1. Structural equilibrium as a moderator

In GIST, degree of categorical invariance is characterized as a *global* distance metric on ideotype space. Yet, each ideotype also contains *local* information (i.e., independent of other ideotypes) that may contribute to overall concept learning difficulty. For example, associated with each ideotype is a certain degree of "structural equilibrium". A categorical stimulus X is in structural equilibrium (SE) whenever the SKs of its ideotype are all zero or, in other words, whenever each of its dimensions is diagnostic or essential. SE indicates perfect structural stability because when a categorical stimulus is in SE, each of its dimensions plays the exact same structural role (for a more detailed explanation of this construct, see the supporting documents to this article). Psychologically, degree of SE, as measured by the percentage of zero SKs in the ideotype, is indicative of the degree of perceived independence between the dimensions of a stimulus or, equivalently, of how easy it is to discriminate between the structural roles that the dimensions play in a categorical stimulus. We propose that high SE exerts a positive moderating effect in concept learnability by facilitating the identification of the dimensions that are subsequently processed by a rule formation system. In other words, perfectly diagnostic dimensions make categorization easy.

For example, a categorical stimulus X that is a structure instance of structure IV in Fig. 1, has an ideotype represented by the (.5,.5,.5) structural manifold. There is a strong interaction between dimensions here because there are no zero SKs present. It is this perceptual confound in the diagnostic role that the dimensions play that makes it just as hard to learn this type of stimulus as it would be to learn one whose overall degree of homogeneity is slightly lower but whose ideotype has a value of (0,.5,.5) (i.e., a structure instance of structure V in Fig. 1). For this latter ideotype, our conceptual system readily and unequivocally recognizes that the first dimension is the most diagnostic. In this sense, high degree of SE may play a moderating secondary and minor influence in lowering the perception of concept learning difficulty by clarifying the diagnostic role of the dimensions of the categorical stimulus. Note that ideotypes corresponding to the down parity structure types tested in our experiment do not contain any zero kernels (see Table 1) so the learning difficulty of their associated categorical stimuli is not alleviated by this moderating factor. The degree of structural equilibrium of a categorical stimulus X is the percentage of zero SKs in its ideotype (+1 to avoid zero percentages and division by zero). Based on this quantity, we defined the structural equilibrium coefficient $\eta(\mathrm{X})$ (see supporting documents website). Incorporating $\eta(\mathrm{X})$ (abbreviated as η) into Eq. (5) results in the following variant of the GISTM (Eq. (6)) which we shall refer to as the GISTM-SE; in the next section, we test both models.

$$\psi(\mathrm{X}) = \frac{pe^{-k\widehat{\Phi}^2(\mathrm{X})}}{\eta} = \eta^{-1}pe^{-k\widehat{\Phi}^2(\mathrm{X})} \tag{6}$$

5. Methods and empirical evidence

Three experiments were conducted to test the qualitative and quantitative predictions made by the GISTM and GISTM-SE and to compare these to those made by other well-known categorization models. Of the three experiments, only Experiments 1 and 2 are described in detail. The reader is referred to the supporting documents website for a discussion of Experiment 3. Experiment 1 aimed to test, extend, and replicate the results obtained by Feldman (2000) in his study involving 76 category structures. Like Feldman, in our experiment we used a parainformative task. However, we increased the number of structures tested to 84: namely, the 76 structures studied by Feldman, the four structures involving two dimensions, and another four structures (of three and four dimensions) in up and down parity involving a single object. The structures tested are listed in Table 1 in terms of the Boolean formulae that define them).

In contrast, Experiment 2 aimed to test and compare the aforementioned models on 24 category structures defined over three quaternary dimensions whose values lied on a [0,1] gradient. Only four of the eight models tested

[1] By transforming the Euclidean metric that defines degree of categorical invariance Φ in appendix B to the following metric instead $\widehat{\Phi}'(\mathrm{X}) = \sum_{d=1}^{D}[H_{|d|}(\mathrm{X})]^2$, then Eq. (5) may retain the original external non-Gaussian form of Eq. (4): namely, $\psi(\mathrm{X}) = pe^{-k\Phi'(\mathrm{X})}$) and $\psi(\mathrm{X}) = pe^{-(\frac{D_0}{D})\Phi'(\mathrm{X})}$ for the non-parametric variant.

Table 1
List of the 84 category structures tested. The first column displays the structure labels, the second column the Boolean functional description of the structure, the third column the structural manifold of the structure, the fourth column the degree of categorical invariance of the structure. The **U** stands for up parity and the **D** stands for down parity. Asterisks on structure labels in the first column identify the structures that were not among the 76 structures studied by Feldman (2000).

TYPE	BOOLEAN FUNCTIONAL DESCRIPTION OF STRUCTURE	STRUCTURAL MANIFOLD	$\bar{\Phi}$
U-2_2[1]-1*	*(x′y′)*	(0,0)	0.00
U-2_2[2]-1*	*((x′y′)+(x′y))*	(0,1)	1.00
U-2_2[2]-2*	*((x′y′)+(xy))*	(0,0)	0.00
U-3_2[1]-1*	*(x′y′z′)*	(0,0,0)	0.00
U-3_2[2]-1	*((x′y′z′)+(x′y′z))*	(0,0,1)	1.00
U-3_2[2]-2	*((x′y′z′)+(x′yz))*	(0,0,0)	0.00
U-3_2[2]-3	*((x′y′z′)+(xyz))*	(0,0,0)	0.00
U-3_2[3]-1	*((x′y′z′)+(x′y′z)+(x′yz′))*	(0,.67,.67)	0.94
U-3_2[3]-2	*((x′y′z′)+(x′y′z)+(xyz′))*	(0,0,.67)	0.67
U-3_2[3]-3	*((x′y′z′)+(x′yz)+(xy′z))*	(0,0,0)	0.00
3_2[4]-1	*((x′y′z′)+(x′y′z)+(x′yz′)+(x′yz))*	(0,1,1)	1.41
3_2[4]-2	*((x′y′z′)+(x′y′z)+(xyz′)+(xyz))*	(0,0,1)	1.00
3_2[4]-3	*((x′y′z′)+(x′y′z)+(x′yz′)+(xy′z))*	(.50,.50,.50)	0.87
3_2[4]-4	*((x′y′z′)+(x′y′z)+(x′yz′)+(xy′z′))*	(.50,.50,.50)	0.87
3_2[4]-5	*((x′y′z′)+(x′y′z)+(x′yz′)+(xyz))*	(0,.50,.50)	0.71
3_2[4]-6	*((x′y′z′)+(x′yz)+(xy′z)+(xyz′))*	(0,0,0)	0.00
U-4_2[1]-1 *	*(x′y′z′w′)*	(0,0,0,0)	0.00
U-4_2[2]-1	*((x′y′z′w′)+(x′y′z′w))*	(0,0,0,1)	1.00
U-4_2[2]-2	*((x′y′z′w′)+(x′y′zw))*	(0,0,0,0)	0.00
U-4_2[2]-3	*((x′y′z′w′)+(x′yzw))*	(0,0,0,0)	0.00
U-4_2[2]-4	*((x′y′z′w′)+(xyzw))*	(0,0,0,0)	0.00
U-4_2[3]-1	*((x′y′z′w′)+(x′y′z′w)+(x′y′zw′))*	(0,0,.67,.67)	0.94
U-4_2[3]-2	*((x′y′z′w′)+(x′y′z′w)+(x′yzw′))*	(0,0,0,.67)	0.67
U-4_2[3]-3	*((x′y′z′w′)+(x′y′z′w)+(xyzw′))*	(0,0,0,.67)	0.67
U-4_2[3]-4	*((x′y′z′w′)+(x′y′zw)+(x′yz′w))*	(0,0,0,0)	0.00
U-4_2[3]-5	*((x′y′z′w′)+(x′y′zw)+(xyz′w′))*	(0,0,0,0)	0.00
U-4_2[3]-6	*((x′y′z′w′)+(x′y′zw)+(xyz′w))*	(0,0,0,0)	0.00
U-4_2[4]-1	*((x′y′z′w′)+(x′y′z′w)+(x′y′zw′)+(x′y′zw))*	(0,0,1,1)	1.41
U-4_2[4]-2	*((x′y′z′w′)+(x′y′z′w)+(x′y′zw′)+(x′yz′w′))*	(0,.50,.50,.50)	0.87
U-4_2[4]-3	*((x′y′z′w′)+(x′y′z′w)+(x′y′zw′)+(x′yz′w))*	(0,.50,.50,.50)	0.87
U-4_2[4]-4	*((x′y′z′w′)+(x′y′z′w)+(x′y′zw′)+(x′yzw))*	(0,0,.50,.50)	0.71
U-4_2[4]-5	*((x′y′z′w′)+(x′y′z′w)+(x′y′zw′)+(xyz′w′))*	(0,0,.50,.50)	0.71
U-4_2[4]-6	*((x′y′z′w′)+(x′y′z′w)+(x′y′zw′)+(xyz′w))*	(0,0,.50,.50)	0.71
U-4_2[4]-7	*((x′y′z′w′)+(x′y′z′w)+(x′y′zw′)+(xyzw))*	(0,0,.50,.50)	0.71
U-4_2[4]-8	*((x′y′z′w′)+(x′y′z′w)+(x′yzw′)+(x′yzw))*	(0,0,0,1)	1.00
U-4_2[4]-9	*((x′y′z′w′)+(x′y′z′w)+(x′yzw′)+(xy′zw′))*	(0,0,0,.50)	0.50
U-4_2[4]-10	*((x′y′z′w′)+(x′y′z′w)+(x′yzw′)+(xy′zw))*	(0,0,0,.50)	0.50
U-4_2[4]-11	*((x′y′z′w′)+(x′y′z′w)+(x′yzw′)+(xyzw′))*	(.50,0,0,.50)	0.71
U-4_2[4]-12	*((x′y′z′w′)+(x′y′z′w)+(x′yzw′)+(xyzw))*	(0,0,0,.50)	0.50
U-4_2[4]-13	*((x′y′z′w′)+(x′y′z′w)+(xyzw′)+(xyzw))*	(0,0,0,1)	1.00
U-4_2[4]-14	*((x′y′z′w′)+(x′y′zw)+(x′yz′w)+(x′yzw′))*	(0,0,0,0)	0.00
U-4_2[4]-15	*((x′y′z′w′)+(x′y′zw)+(x′yz′w)+(xy′z′w))*	(0,0,0,0)	0.00
U-4_2[4]-16	*((x′y′z′w′)+(x′y′zw)+(x′yz′w)+(xy′zw′))*	(0,0,0,0)	0.00
U-4_2[4]-17	*((x′y′z′w′)+(x′y′zw)+(x′yz′w)+(xyzw′))*	(0,0,0,0)	0.00
U-4_2[4]-18	*((x′y′z′w′)+(x′y′zw)+(xyz′w′)+(xyzw))*	(0,0,0,0)	0.00
U-4_2[4]-19	*((x′y′z′w′)+(x′y′zw)+(xyz′w)+(xyzw′))*	(0,0,0,0)	0.00
D-3_2[1]-1 *	*((x′y′z)+(x′yz′)+(x′yz)+(xy′z′)+(xy′z)+(xyz′)+(xyz))*	(.86,.86,.86)	1.48
D-3_2[2]-1	*((x'yz')+(x'yz)+(xy'z')+(xy'z)+(xyz')+(xyz))*	(.67,.67,1)	1.37
D-3_2[2]-2	*((x'y'z)+(x'yz')+(xy'z')+(xy'z)+(xyz')+(xyz))*	(.67,.67,.67)	1.15
D-3_2[2]-3	*((x'y'z)+(x'yz')+(x'yz)+(xy'z')+(xy'z)+(xyz'))*	(.67,.67,.67)	1.15
D-2_2[1]-1 *	*((x′y)+(xy′)+(xy))*	(.67,.67)	0.94
D-3_2[3]-1	*((x'yz)+(xy'z')+(xy'z)+(xyz')+(xyz))*	(.40,.80,.80)	1.20

D-3$_2$[3]-2	*((x'yz')+(x'yz)+(xy'z')+(xy'z)+(xyz))*	(.40,.40,.80)	0.98
D-3$_2$[3]-3	*((x'y'z)+(x'yz')+(xy'z')+(xyz')+(xyz))*	(.40,.40,.40)	0.69
D-4$_2$[1]-1*	*((x'y'z'w)+(x'y'zw')+(x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw'))*	(.93,.93,.93,.93)	1.87
D-4$_2$[2]-1	*((x'y'zw')+(x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.86,.86,.86,1)	1.79
D-4$_2$[2]-2	*((x'y'z'w)+(x'y'zw')+(x'yz'w')+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.86,.86,.86,.86)	1.71
D-4$_2$[2]-3	*((x'y'z'w)+(x'y'zw')+(x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw')+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.86,.86,.86,.86)	1.71
D-4$_2$[2]-4	*((x'y'z'w)+(x'y'zw')+(x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw'))*	(.86,.86,.86,.86)	1.71
D-4$_2$[3]-1	*((x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.77,.77,.92,.92)	1.70
D-4$_2$[3]-2	*((x'y'zw')+(x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw')+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.77,.77,.77,.92)	1.62
D-4$_2$[3]-3	*((x'y'zw')+(x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.77,.77,.77,.92)	1.62
D-4$_2$[3]-4	*((x'y'z'w)+(x'y'zw')+(x'yz'w')+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.77,.77,.77,.77)	1.54
D-4$_2$[3]-5	*((x'y'z'w)+(x'y'zw')+(x'yz'w')+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w)+(xyz'w')+(xyzw'))*	(.77,.77,.77,.77)	1.54
D-4$_2$[3]-6	*((x'y'z'w)+(x'y'zw')+(x'yz'w')+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w)+(xyz'w')+(xyzw'))*	(.77,.77,.77,.77)	1.54
D-4$_2$[4]-1	*((x'yz'w')+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.67,.67,1,1)	1.70
D-4$_2$[4]-2	*((x'y'zw)+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.67,.83,.83,.83)	1.59
D-4$_2$[4]-3	*((x'y'zw)+(x'yz'w')+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.67,.83,.83,.83)	1.44
D-4$_2$[4]-4	*((x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw')+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.67,.67,.83,.83)	1.51
D-4$_2$[4]-5	*((x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw')+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.67,.67,.83,.83)	1.51
D-4$_2$[4]-6	*((x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.67,.67,.83,.83)	1.51
D-4$_2$[4]-7	*((x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.67,.67,.83,.83)	1.51
D-4$_2$[4]-8	*((x'y'zw')+(x'y'zw)+(x'yz'w')+(x'yz'w)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.67,.67,.67,1)	1.53
D-4$_2$[4]-9	*((x'y'zw')+(x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.67,.67,.67,.83)	1.42
D-4$_2$[4]-10	*((x'y'zw')+(x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xyz'w)+(xyz'w')+(xyzw'))*	(.67,.67,.67,.83)	1.42
D-4$_2$[4]-11	*((x'y'zw')+(x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw'))*	(.83,.67,.67,.83)	1.51
D-4$_2$[4]-12	*((x'y'zw')+(x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw'))*	(.67,.67,.67,.83)	1.42
D-4$_2$[4]-13	*((x'y'zw')+(x'y'zw)+(x'yz'w')+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw)+(xyz'w')+(xyz'w))*	(.67,.67,.67,1)	1.53
D-4$_2$[4]-14	*((x'y'z'w)+(x'y'zw')+(x'yz'w')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.67,.67,.67,.67)	1.33
D-4$_2$[4]-15	*((x'y'z'w)+(x'y'zw')+(x'yz'w')+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.67,.67,.67,.67)	1.33
D-4$_2$[4]-16	*((x'y'z'w)+(x'y'zw')+(x'yz'w')+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.67,.67,.67,.67)	1.33
D-4$_2$[4]-17	*((x'y'z'w)+(x'y'zw')+(x'yz'w')+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw'))*	(.67,.67,.67,.67)	1.33
D-4$_2$[4]-18	*((x'y'z'w)+(x'y'zw')+(x'yz'w')+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w)+(xyzw'))*	(.67,.67,.67,.67)	1.33
D-4$_2$[4]-19	*((x'y'z'w)+(x'y'zw')+(x'yz'w')+(x'yz'w)+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'z'w)+(xy'zw')+(xy'zw)+(xyz'w')+(xyzw))*	(.67,.67,.67,.67)	1.33

Table 2
Distribution of the 18 tested structure families consisting of a total of 84 structures distributed among six groups of subjects. The first column shows a group label, the second column displays the particular structure families in each group. The third column specifies the number of structures in each particular structure family and the fourth column shows the number of subjects assigned per group. The U stands for up-parity and the D stands for down-parity structure families. Asterisks identify the 11 structure families from Feldman (2000). Note that the U-4_2[4] family and its down-parity counterpart, the D-4_2[4] or $\mathbf{4_2[12]}$ family (each of which consists of 19 structures), were split into 3 groups of 6, 6, and 7 structures.

Groups of Tested Families (18 Families in total divided into six groups)		Number of Unique Structures Associated with each individual Family (84 in Total)	Number of Subjects Assigned per Group of Families
Group I	U-2_2[1]	1	30
	2_2[2]	2	
	D-2_2[1]	1	
Group II	U-3_2[1]	1	30
	U-3_2[2]*	3*	
	U-3_2[3]*	3*	
	3_2[4]*	6*	
	D-3_2[3]*	3*	
	D-3_2[2]*	3*	
	D-3_2[1]	1	
Group III	U-4_2[1]	1	30
	U-4_2[2]*	4*	
	U-4_2[3]*	6*	
	D-4_2[3]*	6*	
	D-4_2[2]*	4*	
	D-4_2[1]	1	
Group IV	U-4_2[4]*	6*	30
	D-4_2[4]*	6*	
Group V	U-4_2[4]*	6*	30
	D-4_2[4]*	6*	
Group VI	U-4_2[4]*	7*	30
	D-4_2[4]*	7*	

were capable of processing this kind of categorical stimuli (i.e., the GCM, ACM, GISTM, and GISTM-SE). Hence, our discussion focuses on these four. Finally, Experiment 3 aimed to show that relative magnitude judgments of concept learning difficulty are consistent with results on categorization performance as measured by error rates on parainformative tasks. Recall that both the GISTM and GISTM-SE attempt to predict perceived degree of learning difficulty. Yet, Experiments 1 and 2 are designed to test classification performance in terms of proportion of classification errors. This is not an issue as long as we accept the assumption that degree of learning difficulty may be operationalized in terms of proportion of classification errors. Experiment 3 aimed to corroborate this way of operationalizing perceived degree of concept learning difficulty by directly testing people's magnitude judgments on the learning difficulty of the 3_2[4] structures.

5.1. Experiment 1: Difficulty ordering of 84 category structures

5.1.1. Subjects

A total of 180 Ohio University undergraduates participated in the experiment. Six groups of 30 subjects each were assigned to the following six sets of structure families: {2_2[1], 2_2[2], 2_2[3]}, {3_2[1], 3_2[2], 3_2[3], 3_2[4], 3_2[5], 3_2[6], 3_2[7]}, {4_2[1], 4_2[2], 4_2[3], 4_2[15], 4_2[14], 4_2[13]}, {4_2[4]–(1–6), 4_2[12]–(1–6)}, {4_2[4]–(7–12), 4_2[12]–(7–12)}, and

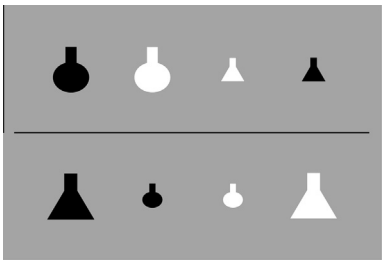

Fig. 5. Example of a stimulus shown on a computer display to subjects in Experiment 1. The categorical stimulus above the line is a structure instance from one of the six structures in the $3_2[4]$ family of structures, below the line is its logical complement.

$\{4_2[4]–(13–19), 4_2[12]–(13–19)\}$. In total, 18 structure families (including their down parity counterparts) were tested for a grand total of 84 structures. These 84 structures are specified in column 2 of Table 1 in terms of their Boolean concept function descriptions (see Section 1.2 for an explanation). Column 3 of the same table displays the structural manifold of each structure. On the other hand, Table 2 shows the number of structures in each of the 18 structure families and how the structures were assigned to the six distinct groups of 30 subjects.

Each of the 84 structures was tested using four distinct categorical stimuli conforming to the structure (i.e., four distinct structure instances per structure). Fig. 1b shows six structure instances, each corresponding to one of the six structures in the $3_2[4]$ structure family. Structure instances were sampled at random from the entire population of possible structure instances of each particular structure. Finally, as illustrated in Table 2, the $4_2[4]$ structure family was spliced into three sets of structures due to their large number of structures (19 in total).

5.1.2. Materials

An HP XW4600 workstation with a Dell 1708FP 15 in. flat panel LCD monitor (5 ms response time) was used to display the stimuli. Categorical stimuli were sets of flat flasks that varied in color (black or white) and size (large or small) for two dimensional stimuli, in color (black or white), size (large or small), and shape (triangular or rectangular) for three dimensional stimuli, and in color (black or white), size (large or small), shape (triangular or rectangular), and neck width (narrow or wide) for four dimensional stimuli. Each stimulus consisted of two sets of spatially separated flasks: one on top of a line in the middle of the digital display and the other (its complement) below the line as exemplified in Fig. 5.

5.1.3. Procedures

Prior to the start of the experiment, the classification task was explained to each subject. Subjects were told that an art collector likes collecting flasks and that for a period of 20 s they would be shown the flasks that the art collector likes to collect above the horizontal line in the middle of the screen and, below the line, those that the art collector does not like to collect. Moreover, the subjects were told that after the 20 s learning period, they would be shown each of the flasks seen above and below the line, one at a time and at random, for about three seconds. They then were to determine and specify within the three seconds, by pressing one of two mouse buttons (the left one labeled "Y" and the right one labeled "N"), which flask was liked by the art collector and which was not. They were also told that failure to press a button within three seconds counted as a classification error.

After receiving these verbal instructions, each subject sat in front of the digital display so that their eyes were approximately 2.5 feet away. The experiment, a program written in Psychophysics toolbox (version 3), began upon the press of the space bar on the keyboard. The first screen of the experiment contained the same instructions that had been given verbally. After the instructions were read by the subject, the first block of classification trials began by the press of the space bar. Per the given verbal instructions, each block consisted of the following sequence of events: (1) the categorical stimulus was presented for 20 s during the learning phase, (2) afterward, each object from the categorical stimulus seen during the learning phase was displayed at random and one at a time for a period of three seconds, (3) for each block of the 2^D trials (corresponding to the number of possible objects in the two displayed categorical stimuli above and below the line), subjects had up to 3 s to make a classification decision by pressing one of two mouse buttons that were clearly la-

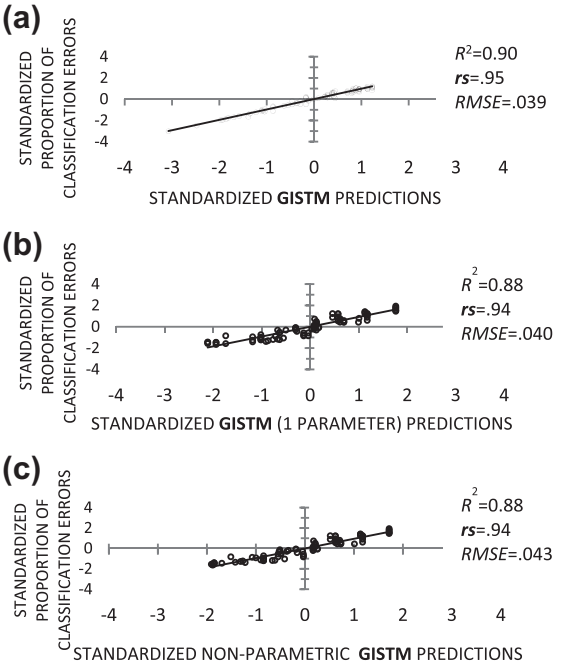

Fig. 6. (a) GISTM fit to data from the 84 classification tasks involving 84 category structures. Three values of the scaling parameter k were used in total, one for all the two dimensional categories stimuli (k_2 = .69), one for the three dimensional stimuli (k_3 = .72), and one for the four dimensional stimuli (k_4 = .55). *rs* stands for Spearman's Rho. (b) Using a single parameter accounts for about 88% of the variance but is contrary to the homogeneous subspaces hypothesis proposed in Section 4. (c) Using the non-parametric variant also accounts for about 88% of the variance (for more details, see the support documents website).

Table 3
Approximate R^2 (first number), Spearman's Rho (second number), and Root Mean Square Error (third number) for the GISTM (top of first column) and GISTM-SE (top of second column) using data from the 84 structures studied by the author and the 76 structures studied by Feldman. Scaling parameter k values are also included where the subscript on each k indicates the dimensionality of the ideotype subspace.

	GISTM $pe^{-k\Phi^2(X)}$	GISTM-SE $pe^{-k\Phi^2(X)} / \eta$
VEXPRO-76	R^2=.90; ***rs*=.95**; *RMSE*=.038 **k_3=.69; k_4=.54**	R^2=.91; ***rs*=.96**; *RMSE*=.037 **k_3=1.02; k_4=.68**
VEXPRO-84	R^2=.90; ***rs*=.95**; *RMSE*=.039 **k_2=.69; k_3=.72; k_4=.55**	R^2=.91; ***rs*=.97**; *RMSE*=.037 **k_2=1.3; k_3=1.02; k_4=.69**
FEXPRO-76	R^2=.69; ***rs*=.83**; *RMSE*=.05 **k_3=.32; k_4=.50**	R^2=.66; ***rs*=.81**; *RMSE*=.05 **k_3=.55; k_4=.63**

beled as "Y" (left button) and "N" (right button). There were a total of four blocks of classification trials, one per structure instance. The program recorded the percentage of classification errors per block of trials and the average percentage of errors across each set of four blocks of classification trials corresponding to each of the 84 structures tested.

5.2. Methodological design of Experiment 1

Our first experiment tested classification performance on parainformative classification tasks involving 84 distinct categorical stimulus structures. Although Feldman (2000) used a similar parainformative paradigm, there are significant differences between the two experiments that we believe account for the major differences found in their corresponding results. In the remainder of this discussion, FEXPRO refers to Feldman's experimental protocol and VEXPRO to the author's. To start with, FEXPRO utilizes different training periods for categorical stimuli of different dimensionality. The assumption being that the higher the dimensionality of the categorical stimulus is, the more time it will take to learn it. More specifically, subjects are given 5p seconds of training time (where p is the number of objects in the categorical stimulus). However, application of the 5p seconds rule may have introduced a bias. Indeed, there are higher dimensional stimuli that take less time to learn than lower dimensional stimuli. In contrast, VEXPRO assigns the same amount of time per categorical stimulus regardless of its dimensionality: namely, 20 s.

The second methodological difference between VEXPRO and FEXPRO is that in VEXPRO, we have sampled from the entire population of structure instances. That is, VEXPRO samples at random from the entire class of possible categorical stimuli or instances of a particular structure; in contrast, FEXPRO does not sample from the entire population of all possible structure instances according to the Supplementary Support Webpage of experimental details cited by Feldman (2000). This, again, could introduce some significant biases in the experiment. The third methodological difference concerns the nature of the categorical stimuli presented to subjects. FEXPRO uses a world of "amoeba". This world may have been perceived as considerably more abstract than the world of art collectors collecting sets of flasks used in VEXPRO. Thus, it may have been harder to communicate the nature of the task at hand. On the same line of argument, we point at the way that instructions were delivered to subjects. Under VEXPRO, subjects were explained the experimental task in a rather thorough fashion both verbally and in writing. A greater amount of briefing time can influence an experiment by eliminating noise associated with a subject's uncertainty as to what he/she is supposed to do. The fourth methodological difference between the two experiments concerns the way that the stimuli were assigned per group of subjects. FEXPRO assigns to each group of subjects (six groups in total) a family of structures and their down parity counterparts. Therefore, some groups of subjects are getting a much higher number of structure instances than others. In VEXPRO, larger families of structures (e.g., the $4_2[4]$ family) are split and distributed among several groups as described under Section 5.1.1. The value of this is to reduce what would be a two hour experiment to a less than one hour experiment. We believe that this strategy reduces noise in the data by reducing subject fatigue.

Finally, in addition to the methodological differences discussed above, we also note two facts that lead us to believe that the data obtained under VEXPRO is less noisy than the data obtained under FEXPRO. First, Feldman (2000) did not observe the SHJ ordering in his data. In fact, error rates per type in his experiment were as follows: I(.06), II(.17), III(.16), IV(.23), V(.19), VI(.28). In contrast, our data clearly reflected the SHJ ordering: I(.02), II(.07), III(.19), IV(.18), V(.17), VI(.32) (pairwise t-tests indicated no significant differences between types III, IV, and V). Finally, the correlation between our data and Feldman's data on the 76 category structures that he tested was not very high ($r = .66, p \leqslant .0001$; $rs = .65$). All of the above points suggest that the VEXPRO data is probably less noisy and more reliable than the FEXPRO data.

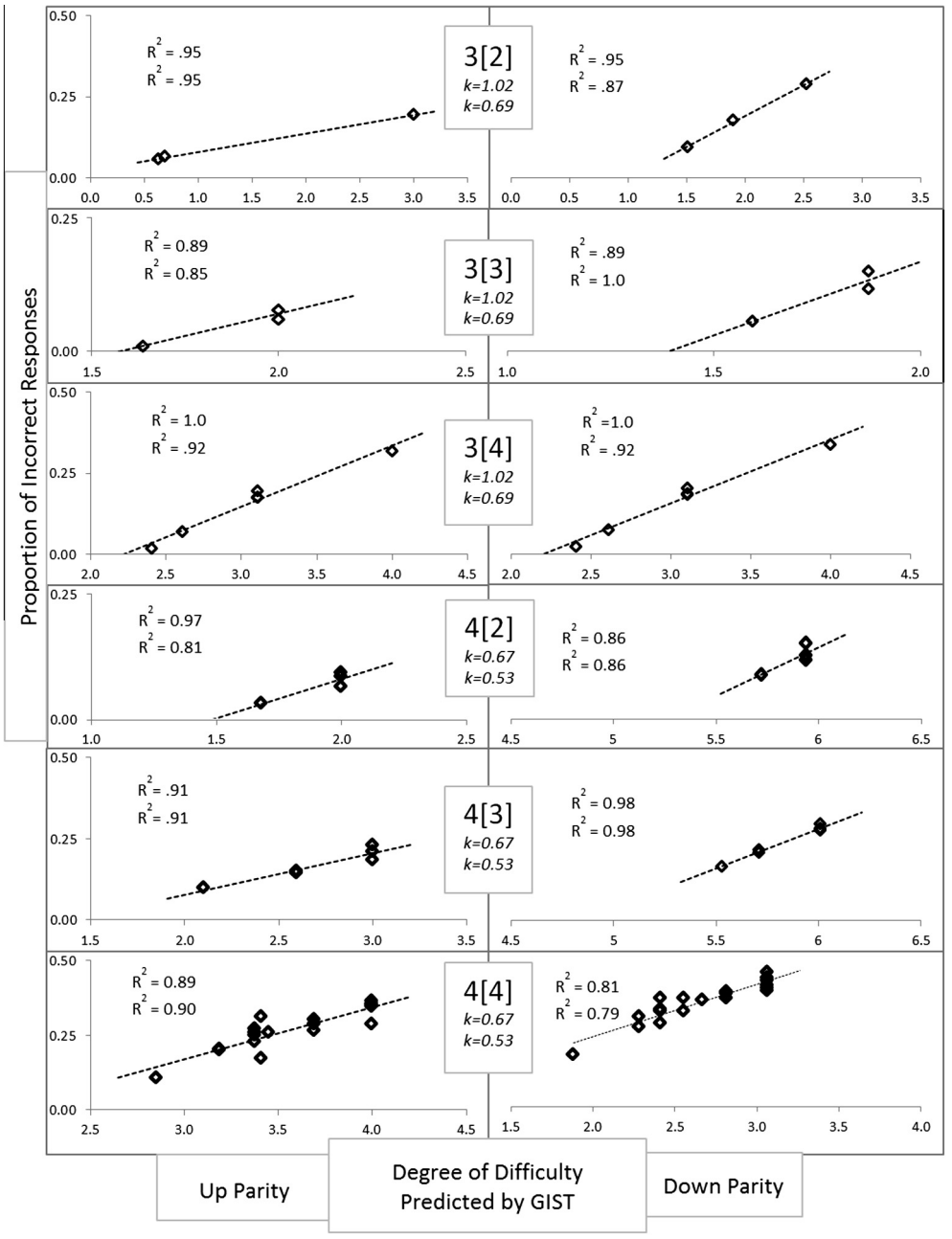

Fig. 7. GISTM and GISTM-SE fits for 11 families of categorical stimulus structures using data from Experiment 1. The first R^2 value corresponds to the GISTM-SE, the second R^2 value corresponds to the GISTM. Likewise, the top scaling parameter k value corresponds to the GISTM-SE, the bottom to the GISTM. Also note that the scaling parameter k values are the same for all structures of the same dimensionality regardless of their family of origin.

5.2.1. Model fits

In this section we examine how well the GISTM and the GISTM-SE fit the VEXPRO-84 data (data from all 84 structures tested in Experiment 1), the VEXPRO-76 data (data from Experiment 1 on the 76 structures tested in Feldman, 2000), and the FEXPRO-76 data (data from Feldman's experiment on 76 structures; Feldman, 2000). The model fitness analysis was conducted with the following statistics: the coefficient of determination R^2, the root mean square error (*RMSE*), Spearman's Rho (*rs*) which measures ordinal correlation, and the Wilke-Shapiro normality test on the distribution of the residuals. The scaling parameter k estimates were computed using the RISK SOLVER ENGINE, a program that implements classical methods of constrained optimization (such as the Lagrangian) using gradient descent algorithms (see Daellenbach & George, 1978 for a theoretical explanation) to maximize the coefficient of determination or, equiva-

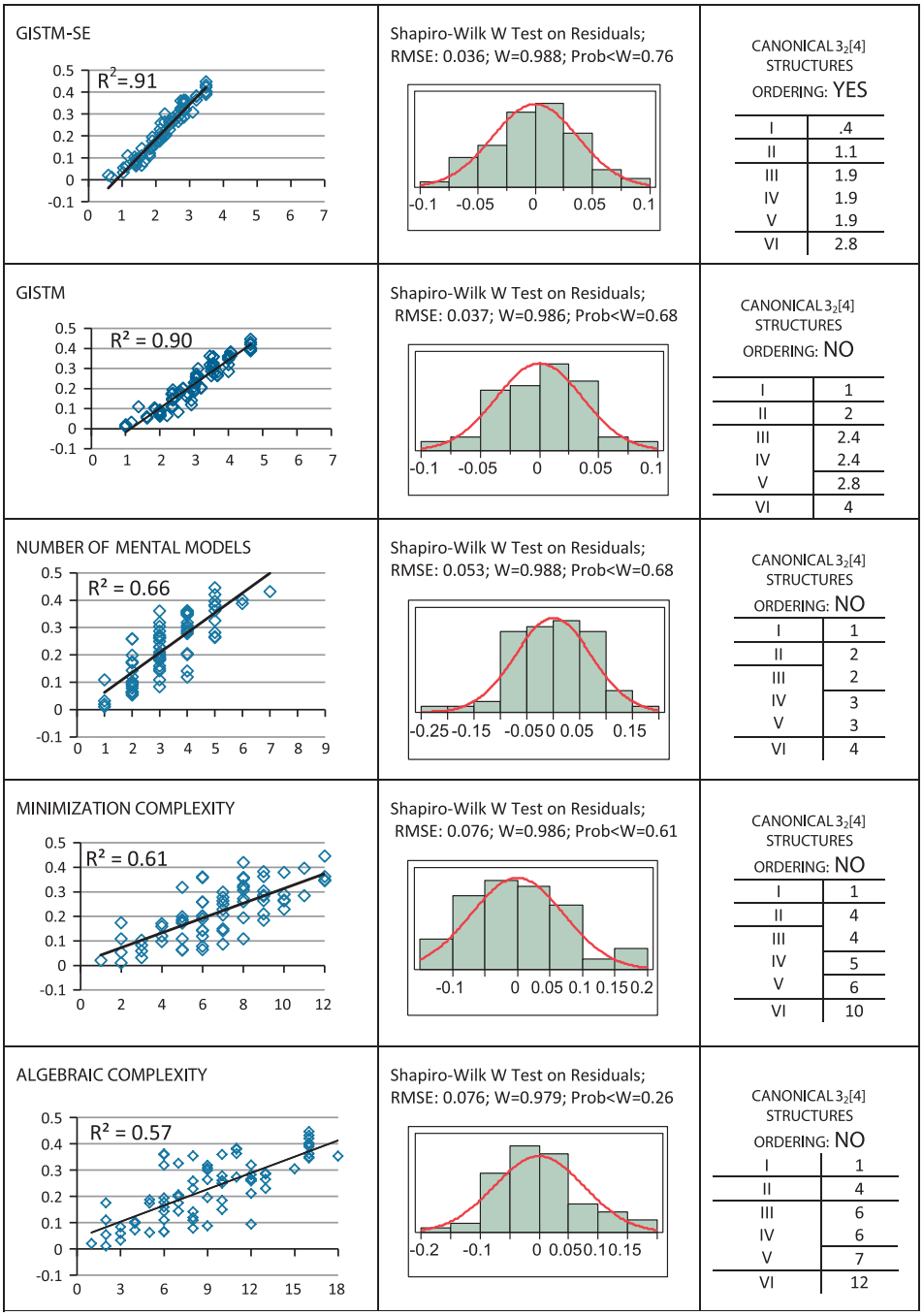

Fig. 8. Model fits to our VEXPRO-76 data involving the 76 structures studied in Feldman (2000). Note that the quality of each fit is reflected by the distribution of the residuals. Note the significantly lower RMSE for the GISTM and the GISTM-SE. Also, note that structural equilibrium accounts for the SHJ ordering. The Shapiro–Wilk *W* Test is a measure of the extent to which the residuals are normally distributed (e.g., $W = 1$ indicates a perfectly normal distribution).

lently, minimize the sum of squared differences between the model predictions and the empirical data. Parameter estimates for the scaling parameter k were performed at two levels: dimensional classes of structures and all structures combined. So, for example, at the dimensional level, the 76 structures tested by Feldman (a subset of the 84 tested by the author) were divided into 3-dimensional and 4-dimensional stimuli. Thus, two parameter values were generated (one value, labeled k_3, for all the three dimensional stimuli and one value, labeled k_4, for all the four dimensional stimuli). Similarly, three scaling values were used on the 84 structures (with an additional estimated value, labeled k_2, for all the two-dimensional stimuli).

Table 4
Model comparisons discussed in this section are summarized in the table above. There are a total of five performance benchmarks reported, one per row. *rs* stands for the Spearman rho coefficient which is a measure of correlation at an ordinal level. Note that the GISTM, GISTM-NP (non-parametric GISTM) and the GISTM-SE hold a significant advantage over the rest of the models and that they perform similarly well (with a slight advantage held by the GISTM-SE in being able to precisely predict the SHJ ordering with a single scaling parameter value k=1.02)

	GCM Nosofsky (1984)	MinC Feldman (2000); Vigo (2006)	ACM (Feldman, 2006)	NOMM (Goodwin & Johnson-Laird, 2011)	GISTM-NP	GISTM	GISTM-SE
SHJ Ordering 1<2<[3,4,5]<6	*YES: rs*=1 1<2<[3,4,5]<6	*NO: rs*=.89 1<[2,3]<4<5<6	*NO: rs* =.96 1<2<[3,4]<5<6	*NO: rs* =.91 1<[2,3]<[4,5]<6	*NO: rs*=.96 1<2<[3,4]<5<6	***NO: rs*=.96** **1<2<[3,4]<5<6**	***YES: rs* =1** **1<2<[3,4,5]<6**
84 Structures	*rs=.58* R^2=.27 *RMSE=.10*	*rs =.80* R^2=.64 *RMSE=.07*	*rs=.79* R^2=.63 *RMSE=.07*	*rs =.78* R^2=.60 *RMSE=.08*	*rs =.94* R^2=.88 *RMSE=.04*	***rs =.95*** ***R^2=.90*** ***RMSE=.04***	***rs =.96*** ***R^2=.91*** ***RMSE=.04***
Fits on 24 Gradient Substructures	R^2=.27 *RMSE=.10*	*N/A*	R^2=.57 *RMSE=.08*	*N/A*	R^2=.83 *RMSE=.03*	***R^2=.87*** ***RMSE=.03***	***R^2=.76*** ***RMSE=.03***
Individual Differences Potential	**YES**	NO	NO	NO	YES	**YES**	**YES**
Continuous & Dichotomous Dimensions	**YES**	NO	YES	NO	YES	**YES**	**YES**

Moreover, a single scaling parameter value was estimated for all 84 structures. We first report fits for this single parameter value as a reference point to appreciate the robustness of the model. Estimating for a single best scaling parameter value (k = .54) accounted for about 88% of the variance in the VEXPRO-84 data (R^2 = 0.88, p < .0001, *RMSE* = .04, *rs* = .94) as shown in Fig. 6b. However, the learning difficulty ordering of the 3_2[4] family of category structures was only approximated: I(1.3) < II(2.3) < [III(2.7) = IV(2.7) = V(3)] < VI(4). On the other hand, the GISTM-SE with a single scaling parameter estimate (k = .71) accounted for 83% of the variance (R^2 = 0.83, p < .0001, *RMSE* = .05, *rs* = .91), but predicted the learning difficulty ordering of the 3_2[4] family of category structures: I(.78) < II(1.45) < [III(2.3) = IV(2.3) = V(2.3)] < VI(2.83). When using three scaling parameter values (one per dimensional subspace), as is called for by the heterogeneous psychological space proposed in Section 4, the fits to the data are a bit better. For example, Fig. 6a shows how well the GISTM captures the variance of the VEXPRO-84 data (R^2 = 0.90, p < .0001, *RMSE* = .037, *rs* = .95). Likewise, the GISTM-SE fitted the data about as accurately with (R^2 = 0.91, p < .0001, *RMSE* = .037, *rs* = .97). Accordingly, with respect to the VEXPRO-76 data, the fits by the GISTM and the GISTM-SE were equally accurate (GISTM: R^2 = 0.90, p < .0001, *RMSE* = .038, *rs* = .95; GISTM-SE: R^2 = 0.91, p < .0001, *RMSE* = .037, *rs* = .96) as shown in Table 3 which also specifies the scaling parameter estimates.

To test the robustness of both models, and to determine whether they are over fitting the data, we applied a 10-fold cross-validation test with 100 iterations using a program written in Matlab 7.7. The test yielded nearly identical average R^2s and root mean square errors to those reported above and in Table 3 with respect to the VEXPRO-76 data. This seems to indicate that both models are highly robust and that neither model is over fitting the data (GISTM, avg. R^2 = 0.90, avg. *RMSE* = .036; GISTM-SE, avg. R^2 = 0.91, avg. *RMSE* = .037). Furthermore, the average k parameter estimates for all of the cross-validation tests closely matched those reported in Table 3 (i.e., for the three dimensional data the actual k = .69, the cross-validation k = .63; for the four dimensional data the actual k = .54, the cross validation k = .54). The same performance levels were achieved by both models using the same cross-validation procedure with respect to the VEXPRO-84 data.

As expected, the GISTM and the GISTM-SE fitted the FEXPRO-76 data less accurately than it fitted the VEXPRO-76 data (GISTM: $R^2 = 0.69$, $p < .0001$, $RMSE = .049$, $rs = .83$; GISTM-SE: $R^2 = 0.66$, $p < .0001$, $RMSE = .05$, $rs = .81$). These results, along with their corresponding parameter values, are specified in Table 3. We hypothesize that this is due to more noise in Feldman's data and to differences in the protocols employed by each experiment (for a detailed discussion see Section 5.2). However, the obtained fits were still superior to the leading models proposed thus far as demonstrated in the following section.

In addition to these tests, we divided our data in terms of structure families as shown in Fig. 7 to determine how accurately the data was fitted on a per family basis. Note that Fig. 7 specifies the scaling parameter value that was estimated. This one value was used for all the families of the same dimensionality. Note that, on a per family basis, the fits of both the GISTM and the GISTM-SE were very accurate, accounting for about 80–100% of the variance in the data depending on the given family.

5.2.2. Model comparisons

In Section 1 we discussed several theories of concept learning and their core models. These included ALCOVE (Kruschke, 1992), MinC (Feldman, 2000), SUSTAIN (Love, Medin, & Gureckis, 2004), Algebraic Complexity (Feldman, 2006), and Categorical Invariance (Vigo, 2009). In this section we compare the performance of their core models with the GISTM and GIST-SE. However, we do not consider ALCOVE nor SUSTAIN because they have already been compared unfavorably with the NOMM and the ACM across a wide range of Boolean concepts (see Feldman, 2006 and Goodwin & Johnson-Laird, 2011). However, we do consider the core model underlying ALCOVE, the GCM (Nosofsky, 1984), because like the GISTM and unlike the rest, it is a "phenomenological model" (Luce, 1995) of concept learning. In other words, the model is a high level description of a phenomenon using a mathematical equation that exhibits overall properties of the phenomenon, but not its internal structure. To explain, Luce (1995, p. 3) writes: "This approach is like that of classical physics, in which objects have properties, e.g. mass, charge, temperature-but no explicit molecular or atomic structure is attributed to them." In the GCM, the degree of learning difficulty of a concept is operationalized as is done in the appendix of Nosofsky (1984) by taking the average of the probabilities corresponding to each object-stimulus being classified correctly, and converting these to a percentage of correct responses for the categorical stimulus as a whole.

To generate the NOMM (Goodwin & Johnson-Laird, 2011) and the ACM (Feldman, 2006) predictions we used the same computer programs used by the authors and available for download from their supporting documents websites (see the cited articles for the URLs). Regarding the NOMM, the model predictions for the 76 structures were also posted on the authors' (Goodwin & Johnson-Laird) supporting document's website, so we used these values. Furthermore, predictions for the additional eight structures tested were computed manually and, for confirmation, by the LISP program posted in the authors' supporting documents website. For the ACM, we used the "Concept Algebra Toolbox" (CAT) program to generate the algebraic complexity values. For MinC, as Goodwin and Johnson-Laird (2011) did, we used the predictions published in Feldman's catalogue (Feldman, 2003) with the corrections made by Vigo's QMV method (Vigo, 2006).

We compared the performance of the GISTM and the GISTM-SE to the performance of the models discussed under section 2 in terms of their ability to fit and predict the data from our experiment and Feldman's experiment. More specifically, we used six performance benchmarks: namely, (1) the ability to predict the SHJ ordering (Shepard et al., 1961), (2) the ability to predict the learning difficulty ordering of the 84 (and the well-known subset of 76) category structures (in up and down parity) tested in our Experiment 1, (3) the potential to account for individual differences between subjects, (4) the ability to predict classification performance on the 41 structures (i.e., those in up parity and those without parity) tested by Feldman (2000), (5) the ability to handle continuous and multi-valued dimensions, and (6) the ability to predict the learning difficulty ordering of the 24 category structures tested in our Experiment 2 involving three quaternary dimensions. Table 4 summarizes these comparisons.

Both the GISTM and the GISTM-SE excelled at each benchmark. For example, with respect to the first benchmark, only the GIST-SE (with one scaling parameter value ($k = 1.02$) for all 3-dimensional structures) and the GCM (with 13 free parameter values; two parameter values per each of the six structures and one scaling parameter) could predict the canonical $3_2[4]$ structures ordering found in our data precisely (note that in the GCM the attention weight parameters must sum to one so that only two attention weight parameters need to be estimated to determine the third). This ability we think is due to the moderating effect that structural equilibrium has on learnability, making structure V (see Fig. 1) slightly easier to learn than it would otherwise be strictly on the basis of its relatively low degree of homogeneity. Note that the GISTM comes in a close second place (after the GISTM-SE) among all the models tested ($rs = .92$) when predicting the SHJ ordering at an ordinal level.

With respect to the second benchmark, ability to predict classification performance on the 84 structures tested under Experiment 1, both the GISTM and GISTM-SE account for over 90% of the variance in the VEXPRO-84 data as discussed previously. Moreover, with respect to the VEXPRO-76 data, both the GISTM and GISTM-SE fitted the data equally well as illustrated in Fig. 8 ($R^2 = 0.90$, $p < .0001$, $RMSE = .038$, $rs = .95$; $R^2 = 0.91$, $p < .0001$, $RMSE = .037$; $rs = .96$). In contrast, as shown in Table 4 and Fig. 8, the NOMM, ACM and MinC fitted the VEXPRO-84 data (Table 4) and the VEXPRO-76 data (Fig. 8) similarly, accounting for somewhere between 57% and 66% of the variance in the data. This fact is consistent with a suggestion made under Section 1: namely, that the reduction strategies underlying these models are closer than may seem. Also, note that the NOMM, the ACM, and MinC accounted respectively for approximately 57%, 49%, and 31% of the variance in the FEXPRO 76 structures data (Goodwin and Johnson-Laird, 2011). These differences in performance are not surprising given that the FEXPRO-76

Table 5

Twenty-four category structures associated with the $3_4[4]$ family tested in Experiment 2. Column 1 shows their general label, column 2 shows the rule that defines it with the standardized value for each dimension placed next to each dimensional variable within parentheses, column 3 shows its structural manifold, column 4 shows the empirical error rate (what percentage of the classification decision were wrong on average for all subjects), and columns 5 and 6 show the predictions by the GISTM and GISTM-SE.

Group	24 Gradient substructures	Structural manifold	Error rate	GISTM	GISTM-SE
1	$x(.66)y(.34)z(.34) + x(.66)y(.34)z(.66) + x(.34)y(.34)z(.34) + x(.34)y(.34)z(.66)$	(1,0,1)	.16	1.8	0.4
	$x(0)y(.34)z(.34) + x(0)y(.66)z(.34) + x(0)y(.66)z(.66) + x(0)y(.34)z(.66)$	(0,1,1)	.17	1.8	0.4
	$x(0)y(.34)z(0) + x(1)y(.34)z(0) + x(1)y(.66)z(0) + x(0)y(.66)z(0)$	(1,1,0)	.12	1.8	0.4
	$x(1)y(1)z(0) + x(1)y(1)z(1) + x(1)y(0)z(1) + x(1)y(0)z(0)$	(0,1,1)	.13	1.8	0.4
2	$x(.34)y(.66)z(.34) + x(.66)y(.34)z(.66) + x(.34)y(.34)z(.34) + x(.66)y(.66)z(.66)$	(0,1,0)	.34	2.7	1.1
	$x(1)y(.66)z(.34) + x(0)y(.34)z(.34) + x(1)y(.66)z(.66) + x(0)y(.34)z(.66)$	(0,0,1)	.24	2.7	1.1
	$x(.34)y(1)z(0) + x(.34)y(0)z(0) + x(.66)y(1)z(1) + x(.66)y(0)z(1)$	(0,1,0)	.29	2.7	1.1
	$x(1)y(0)z(0) + x(1)y(0)z(1) + x(0)y(1)z(1) + x(0)y(1)z(0)$	(0,0,1)	.32	2.7	1.1
3	$x(.34)y(.34)z(.66) + x(.66)y(.34)z(.66) + x(.34)y(.66)z(.66) + x(.34)y(.66)z(.34)$	(.5,.5,.5)	.32	3.0	1.9
	$x(.66)y(0)z(.34) + x(.66)y(1)z(.66) + x(.66)y(1)z(.34) + x(.34)y(1)z(.66)$	(.5,.5,.5)	.29	3.0	1.9
	$x(1)y(1)z(.66) + x(0)y(0)z(.34) + x(0)y(0)z(.66) + x(1)y(0)z(.66)$	(.5,.5,.5)	.27	3.0	1.9
	$x(1)y(1)z(0) + x(1)y(1)z(1) + x(1)y(0)z(0) + x(0)y(1)z(1)$	(.5,.5,.5)	.35	3.0	1.9
4	$x(.34)y(.34)z(.34) + x(.66)y(.66)z(.34) + x(.66)y(.34)z(.34) + x(.66)y(.34)z(.66)$	(.5,.5,.5)	.30	3.0	1.9
	$x(.34)y(.34)z(1) + x(.66)y(.34)z(0) + x(.66)y(.66)z(1) + x(.66)y(.34)z(1)$	(.5,.5,.5)	.26	3.0	1.9
	$x(1)y(0)z(.66) + x(0)y(0)z(.34) + x(0)y(0)z(.66) + x(0)y(1)z(.66)$	(.5,.5,.5)	.26	3.0	1.9
	$x(0)y(1)z(1) + x(0)y(1)z(0) + x(0)y(0)z(0) + x(1)y(1)z(0)$	(.5,.5,.5)	.34	3.0	1.9
5	$x(.66)y(.66)z(.34) + x(.34)y(.66)z(.66) + x(.34)y(.34)z(.66) + x(.34)y(.34)z(.34)$	(0,.5,.5)	.33	3.3	2.0
	$x(1)y(.34)z(.34) + x(0)y(.66)z(.34) + x(1)y(.66)z(.66) + x(0)y(.66)z(.66)$	(.5,0,.5)	.32	3.3	2.0
	$x(.34)y(0)z(0) + x(.66)y(1)z(1) + x(.66)y(0)z(1) + x(.66)y(1)z(0)$	(0,.5,.5)	.35	3.3	2.0
	$x(0)y(0)z(1) + x(1)y(1)z(0) + x(0)y(1)z(1) + x(0)y(0)z(0)$	(0,.5,.5)	.39	3.3	2.0
6	$x(.34)y(.66)z(.34) + x(.66)y(.66)z(.66) + x(.34)y(.34)z(.66) + x(.66)y(.34)z(.34)$	(0,0,0)	.44	4.0	2.8
	$x(0)y(.66)z(.34) + x(1)y(.34)z(.34) + x(1)y(.66)z(.66) + x(0)y(.34)z(.66)$	(0,0,0)	.42	4.0	2.8
	$x(0)y(.66)z(1) + x(1)y(.34)z(1) + x(0)y(.34)z(0) + x(1)y(.66)z(0)$	(0,0,0)	.40	4.0	2.8
	$x(1)y(1)z(1) + x(1)y(0)z(0) + x(0)y(0)z(1) + x(0)y(1)z(0)$	(0,0,0)	.39	4.0	2.8

and VEXPRO-76 datasets reflect significantly different difficulty orderings for the 76 structures as is evident by a low ordinal correlation between them ($rs = .65$). Indeed, given this latter fact, we expected even bigger differences in the model fits to the two datasets as was the case with the GISTM. Finally, the GCM accounts for only 27% of the variance in the VEXPRO-84 data as shown in Table 4 using a set of uniformly distributed attention weights (per group of 2, 3, and 4-dimensional stimuli) and using optimal value estimates of three scaling parameter values (one parameter value per group of 2, 3, and 4-dimensional stimuli).

Likewise, both the GISTM and GISTM-SE yield the best fits to Feldman's data on the 41 structures (in up parity and no-parity) accounting for about 92% of the variance in the data. In contrast, the GCM accounts for 24% of the variance in the data ($R^2 = 0.24$, $p < .0001$), MinC accounts for about 42% of the variance ($R^2 = 0.42$, $p < .0001$), and NOMM account for about 63% of the variance in the data ($R^2 = 0.63$, $p < .0001$). But again, due to the fact that the $3_2[4]$ structures family ordering was not observed in Feldman's data, these fitness values may be misleading. Notwithstanding, the GISTM achieves a better fit to the FEXPRO-76 data than the competing models by accounting for nearly 70% of its variance (see Table 3). Indeed, none of the discussed competing models accounted for more than about 57% of the variance on the FEXPRO-76 data (Goodwin and Johnson-Laird, 2011).

With respect to the fourth benchmark, as we shall see in the following section, the GISTM and GISTM-SE fitted the data of Experiment 2 (on continuous dimensional values) accurately, accounting for about 91% of the variance. The two other models (GCM and ACM) capable of handling and accounting for this kind of data did not fit the data nearly as well (see next section for a discussion). With respect to the fifth benchmark, among all the models tested, only the GCM, GISTM, and GISTM-SE can potentially account accurately for individual differences mainly because the other models (i.e., the ACM, NOMM, and MinC) are parameter-free. The GCM, GISTM, and GISTM-SE potentially account for individual differences via the scaling parameter k in the GISTM and GISTM-SE and via attention weights and a scaling parameter in the GCM. Because such an analysis would be very elaborate and beyond the scope of this article, we did not test the ability of these models to

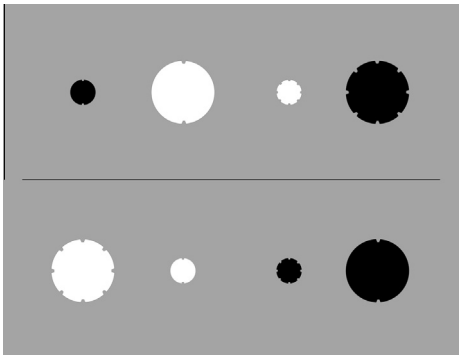

Fig. 9. Structure instance of a category structure with three quaternary dimensions and four objects shown to subjects on a computer screen during the second condition of Experiment 3: note that only four objects comprise the negative or complementary category below the horizontal line.

account for individual differences among subjects, although the ability of the GCM to do so has been well documented (Nosofsky, 1986; Nosofsky et al., 1994a; Rehder and Hoffman, 2005) and the GISTM seems to have similar potential (especially if invariance-pattern detection weights per dimension are used as suggested in Vigo, 2011a). Finally, with respect to the sixth benchmark, note that MinC and NOMM do not function on categorical stimuli defined over continuous dimensions: we regard this fact as a significant limitation.

5.3. Experiment 2: Application to the $3_4[4]$ class of categorical stimuli defined over quaternary continuous dimensions

5.3.1. Subjects

A total of 50 Ohio University undergraduates participated in the experiment which consisted of two parts. In the first part of the experiment, subjects were assigned to 24 categorical stimuli (one per trial) consisting of four objects defined over three quaternary dimensions. In the second part of the experiment, each subject was assigned to four randomly generated structure instances of each of the 24 structures in Table 5.

5.3.2. Materials

An HP XW4600 workstation with a Dell 1708FP 15 in. flat panel LCD monitor (5 ms. response time) was used to display the categorical stimuli. We used categorical stimuli consisting of four round geometric shapes that varied on a: (1) color brightness gradient (black (RGB: 0, 0, 0), medium gray (RGB: 85, 85, 85), light gray (RGB: 170, 170, 170), or white (RGB: 255, 255, 255); (2) a continuity gradient (two bumps in the circular outline, four bumps in the circular outline, six bumps in the circular outline, and eight bumps in the circular outline), and (3) a size gradient (.5 in. in diameter, 1 in. in diameter, 1.5 in. in diameter, and 2 in. in diameter). These values were chosen so that they would map (i.e., standardize) into four evenly distributed and linearly increasing values in the [0,1] real number interval (i.e., [0,.34,.66,1]). There are a total of $4^3 = 64$ possible objects that are definable on these four values. Under each object a text field appeared for subjects to enter a magnitude judgment for each of the four dimensional values of an indicated particular dimension.

On the other hand, Part 2 aimed at testing classification performance with respect to some of the structures of the $3_4[4]$ class of categorical stimuli. Table 5 lists the 24 categorical structures utilized. The number and nature of the possible structures associated with this class of stimuli has not been determined by Boolean algebraists. Thus, the structures were generated bottom-up by selecting four objects (i.e., a categorical stimulus consisting of four objects) at random from a possible $4^3 = 64$ objects. Each generated categorical stimulus conformed to a structure which was visually inspected to ensure its uniqueness among the rest. The structures are organized in Table 5 in six groups with shared degree of invariance. Column 1 shows their general label, column 2 shows the rule that defines it (with the standardized value for each dimension placed next to each dimensional variable within parentheses), and column 3 shows the structural manifold of the structure as computed by GISTM. Subjects were tested on four randomly generated structure instances conforming to each of the 24 structures for a total of 96 trials. As in Experiment 1, each stimulus was comprised of a structure instance (i.e., the positive categorical stimulus consisting of four objects) and a negative or complementary categorical stimulus consisting also of four objects not in the positive categorical stimulus (these were also selected at random but this time from the remaining 60 out of the 64 possible objects from which the positive categorical stimuli were selected). Again, this meant that only a small subset of the true complement of the positive categorical stimulus (consisting of 60 objects) was used as the negative category (see Fig. 9 for an example).

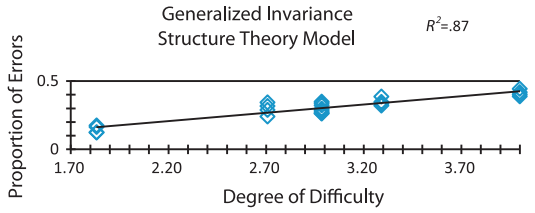

Fig. 10. The GISTM predictions of degree of concept learning difficulty with respect to categorical stimuli consisting of four object-stimuli defined over three quaternary dimensions ($k = .39$).

5.3.3. Procedure

Each subject participated in both parts of the experiment. Prior to the start of the first part of the experiment, the experimenter told the facts of the experiment to each participant. Per the instructions given, participants viewed sets of four objects with three quaternary dimensions (see Section 1.2 for an explanation). With respect to two of the three dimensions, dimensional values were identical for all four objects. However, for the remaining dimension, each object was assigned a unique dimensional value out of the four values defined on an intensity-gradient. For example, for a particular trial, all objects were gray in color and 1.5 in. in diameter, but each possessed a different number of bumps in its circular outline: two, four, six, and eight. Each set of four objects was accompanied by the following written instructions at the top of the screen: "You have 30 s to type a number from 1 to 10 in the blank field under each of the four objects to indicate the perceived degree of its *x* dimension" (where *x* denotes brightness, size, or contour discontinuity). After entering a value for each of the four objects, participants were presented with the next trial. In total, participants performed 24 trials (eight sets of objects for each of the tested three dimensions).

After the first part of the experiment was completed and prior to the start of the second part, each participant was told the facts of the experiment: first, that an art collector likes collecting decorative plates and that for a period of 20 s they would be shown a set of these plates above a line in the middle of the computer display that the art collector likes and another set of four plates below the same line that the art collector does not like. Moreover, the subjects were told that after the 20 s learning period, they would be shown each of the decorative plates seen during the

training period (in the two sets) one at a time and at random for about three seconds. They were also told that their task was to determine and specify within the three seconds, by pressing one of two mouse buttons (the left one labeled "Y" and the right one labeled "N"), which plate was liked by the art collector and which was not; Failure to press a button within the three seconds would count as a classification error. After receiving these verbal instructions, each subject sat in front of the digital display so that their eyes were approximately 2.5 feet away from the screen. The experiment, a program written in Psychophysics toolbox (version 3), began upon the press of the space bar on the keyboard. The first screen of the experiment contained the same instructions that had been given verbally. After the instructions were read, the first block of classification trials began by the press of the space bar.

Per the verbal and written instructions, each block consisted of the following sequence of events: (1) the random stimulus, consisting of a categorical stimulus and its complement, was presented for 20 s during the learning phase, (2) for each classification trial, each of the eight objects seen during the learning phase was displayed at random and one at a time for a period of three seconds, and (3) for each trial, subjects had up to 3 s to make a classification decision. Non responses counted as errors. Subjects were tested on four randomly generated structure instances per structure, so they executed four classification blocks per each structure of Table 5 for a total of $24 \times 4 = 96$ blocks. After the completion of each block, a new block would begin after about one second. The program recorded the percentage of classification errors per block of trials and the average percentage of errors across each set of four blocks of trials for each of the 24 structures.

5.3.4. Results of Experiment 2

The purpose of Experiment 2 was to show how effectively the GISTM and the GISTM-SE fit classification data involving categorical stimuli that are defined by dimensions with a gradient of values. First, we examined the results from the first part of the experiment. Recall that the aim of part one was to validate our theoretical standardized numerical assignment of 0, .33, .67, and 1 to the four qualitative states represented by each dimension. The data showed that the values assigned by subjects on a 1–10 scale to each of the four possible dimensional values for the dimensions of brightness, size, and continuity were on average 0, .30, .66, and 1 after normalizing. These results confirmed that the objective assignment of numerical values to the dimensions of our chosen stimuli (i.e., 0, .33, .67, and 1) were consistent with the subjective assignments made by the 50 subjects.

The classification performance data (error rates) from part 2 of the experiment is shown in the fourth column of Table 5. The GISTM, with $k = .39$, accounted for nearly 90% of the variance in the data of Experiment 2 ($R^2 = .87$, $p < .0001$; $RMSE = .03$) as illustrated in Fig. 10. On the other hand, the GISTM-SE, with $k = 1$, was able to account for 76% of the variance ($R^2 = .76$, $p < .0001$; $RMSE = .04$) in the data. In contrast, the ACM accounted for 57% of the variance ($R^2 = .57$, $p < .0001$; $RMSE = .08$) with three bound (non-free) parameters assigned by the CAT program (Feldman, 2003) and corresponding to three levels of decomposition (one level per dimension), while the GCM accounted for about 27% of the variance in the data ($R^2 = .27$, $p < .0001$; $RMSE = .10$) using a scaling parameter for all the structures and two attention-weight free parameters.

6. Conclusion and research directions

In this paper, we have presented a new theory of concept learning based on invariance pattern detection that addresses some of the shortcomings of the well-known theories to date. In particular, we introduced a general model of conceptual behavior that is equally adept at predicting classification behavior with respect to stimuli interpreted on binary, multivalued, and continuous dimensions. Data from three new experiments and from two previous key experiments provided empirical support for the theory and model.

Furthermore, model comparisons showed that the GISTM and its structural equilibrium variant (GISTM-SE) make more accurate quantitative and qualitative predictions than the leading alternatives. In addition to: (1) accurately fitting the data from every key dataset considered, (2) predicting the SHJ learnability ordering, and (3) providing a plausible explanation for human classification performance, GIST bridges two core constructs on which theories of concept learning are based: complexity reduction and attention regulated similarity assessment. It does this by showing how the invariance structure detection process that is central to it can be explained in terms of a simple cognitive mechanism involving the core capacities of similarity assessment, discrimination, and goal-directed attention. We think this finding is significant not only to concept learning research but to further our understanding of the cognitive nature of invariance structure in general.

Beyond bridging structural and similarity constructs in human concept learning, GISTM, via the use of its scaling parameter k, may potentially account for individual differences in classification performance among humans. Empirical tests need to be conducted to attest to this important feature of the model. Moreover, the scaling/discriminability parameter k could be construed as a bound parameter in GIST by assuming that observers learn the optimal discrimination weights (i.e., one that maximizes classification performance) for categorical stimuli of different dimensionality. This point is similar to the idea proposed by Nosofsky (1984) that the $3_2[4]$ ordering can be predicted a priori by the GCM if one assumes that observers learn the "optimal" attention weights for each dimension of the stimuli (Nosofsky, 1984), thereby not requiring estimation of these weights as free parameters. Again, further empirical tests are necessary to assess the validity of the optimality assumption underlying the GIST. Furthermore, in spite of fundamental differences, there are some theoretical connections that emerge from the principles underlying GIST and some of the assumptions of rival theories (e.g., the notion of discarding "irrelevant" dimensions in the NOMM) which should be further explored.

Notwithstanding these promising results, GIST and GISTM also introduce a number of theoretical and empirical challenges: for example, we did not describe, nor

modeled, the possible mechanisms underlying the encoding of ideotypes (and their distal relationships in ideotype space) into rules and magnitude judgments. Exactly how these representations emerge from the proposed system of pattern detection is an important future direction for the current research. Also, other questions emerge regarding the generality of GIST: can it predict classification performance with respect to natural categories? Can it be as effective in predicting the learnability of categorical stimuli of higher dimensionality and cardinality? These are open questions that will require a considerable amount of empirical research to answer adequately.

Finally, some of the core assumptions of GIST, such as dimensional binding, need to be further investigated, perhaps with the aid of eye tracking technology as it has been done effectively in other areas of concept learning research (e.g., Rehder & Hoffman, 2005; Vigo, Zeigler, and Halsey, 2013). In spite of these open challenges, we have shown that GIST provides a robust and general alternative to other theories of concept learning and classification performance based on its potential to unify competing theories and to account accurately for the learnability of a wide range of categorical stimuli beyond the Boolean variety.

Author's Note:

A program written in Matlab (version 7) that is able to compute the structural manifold of any dimensionally-defined categorical stimulus (with binary, multivalued, or continuous dimensions) is available from the author at http://www.scopelab.net/resources.htm. Additional supporting materials, including tables and figures, are available on the same webpage in a PDF document.

Acknowledgements

I wish to thank Karina-Mikayla Barcus, Sunil Carspecken, Charles Doan, Andrew Halsey, and Derek Zeigler for assisting me with the data collection and data analysis of the reported experiments.

Appendix A. Derivation of the exponential law of invariance

If we let subjective maximal baseline complexity or degree of learning difficulty be measured by cardinality, then consider a categorical stimulus X with |X| object stimuli (|X| stands for the cardinality or the number of elements in X). Now let's assume that the percentage change in the perceived degree of learning difficulty of X (denoted by $\psi(X)$) is negatively proportional to the degree of perceived invariance of the stimulus X. For example, if |X| = 4 then we begin with a maximal perceived degree of difficulty of 4 at the baseline of 0 degree of perceived gestalt homogeneity or invariance. Then as the perceived homogeneity increases to 1, and then 2, and 3, the percentage change in perceived raw complexity will decrease systematically as follows: ((4–0)/4) = 1, ((4–1)/4) = 3/4, ((4–2)/4) = 1/2, ((4–3)/4) = 3/4, ((4–4)/4) = 0. This can be formally articulated with the following expression where the negative sign means that percentage change in subjective degree of difficulty decreases as perceived invariance increases:

$$\frac{\Delta\psi(\widehat{\Phi}(X))}{\psi(\widehat{\Phi}(X))} = -k\,\widehat{\Phi}(X) \tag{7.1}$$

Via algebraic manipulation, we can derive the following equation from Eq. (7.1):

$$\frac{\Delta\psi(\widehat{\Phi}(X))}{\widehat{\Phi}(X)} = -k\psi(\widehat{\Phi}(X)) \tag{7.2}$$

This equation can be rewritten as the rate of change of ψ with respect to $\widehat{\Phi}$ by adding a delta to the denominator of (7.2). This is entirely consistent with our original assumption. We then get:

$$\frac{\Delta\psi(\widehat{\Phi}(X))}{\Delta\widehat{\Phi}(X)} = -k\psi(\widehat{\Phi}(X)) \tag{7.3}$$

Now, if we further assume that ψ and $\widehat{\Phi}$ are differentiable functions (note that in GIST they are not), then we get:

$$\frac{d\psi(\widehat{\Phi}(X))}{d\widehat{\Phi}(X)} = -k\psi(\widehat{\Phi}(X)) \tag{7.4}$$

The solution to Eq. (7.4) is the exponential rate of change:

$$\psi(\widehat{\Phi}(X)) = \psi_0 e^{-k\widehat{\Phi}(X)} \tag{7.5}$$

Here $\psi(\widehat{\Phi}(X))$ is the quantity at degree of invariance $\widehat{\Phi}(X)$ and $\psi_0 = \psi(0)$ is the initially perceived degree of difficulty of the categorical stimulus X or its baseline maximal degree of difficulty which is defined by its cardinality when its degree of invariance $\widehat{\Phi}(X) = 0$. Since our theory is discrete due to the fact that all points in psychological space represent only categorical stimuli as ideotypes, then there will be "real-world gaps" in the functions ψ and $\widehat{\Phi}$. However, we assume that these two functions are special cases of the differential functions above: that is, they apply to a subset of points in their respective continuous domains, and as such, some of their desired properties are also valid for these relatively few discrete "real-world points". The scaling parameter k preserves relative distances in ideotype space.

Appendix B. Generalization to continuous domains using the invariance-similarity equivalence principle

Note: A more detailed account of this generalization is given in the supporting documents website.

In the upcoming discussion we shall employ the following notation:

(1) Let X be a categorical stimulus and |X| stand for the cardinality (i.e., the number of elements) of X.
(2) Let the object-stimuli in X be represented by the vectors $\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2, \ldots, \vec{\mathbf{x}}_n$ (where $n = |X|$).
(3) Let the vector $\vec{\mathbf{x}}_j = (x_1, \ldots, x_D) \in X$ be the j-th D-dimensional object-stimulus in X (where D is the number of dimensions of the stimulus set).

(4) Let $\vec{\mathbf{x}}_{ji}$ be the value of the i-th dimension of the j-th object-stimulus in X. We shall assume throughout our discussion that all dimensional values are real numbers greater than or equal to zero.
(5) Let $S(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k)$ stand for the similarity of object-stimulus $\vec{\mathbf{x}}_j \in \mathrm{X}$ to object-stimulus $\vec{\mathbf{x}}_k \in \mathrm{X}$ as determined by the assumption made in multidimensional scaling theory that stimulus similarity is some monotonically decreasing function of the psychological distance between the stimuli.
(6) Let μ be a standardization operator that transforms the values of a square matrix to values in the [0, 1] closed real number interval. The operator is precisely defined in the supplementary documents.

We begin by describing formally the hypothetical processes of dimensional binding and partial similarity assessment. To do so, we will introduce a new kind of distance operator. But first, let's define the generalized Euclidean distance operator Δ^r (a.k.a. *Minkowski* distance) between two object-stimuli $\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k \in \mathrm{X}$ with attention weights ω_i as:

$$\Delta^r(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k) = \left[\sum_{i=1}^{D} \omega_i \cdot |\vec{\mathbf{x}}_{ji} - \vec{\mathbf{x}}_{ki}|^r\right]^{1/r} \tag{7.6}$$

As in the GCM (Nosofsky, 1984), the inclusion of a parameter ω_i represents the selective attention allocated to dimension i such that $\sum_i \omega_i = 1$.We use this parameter family to represent individual differences in the process of assessing similarities between object-stimuli at this level of analysis. For the sake of simplifying our explanation and examples below, we shall disregard this parameter.

Next we introduce a new kind of distance operator termed the *partial psychological distance operator* $\Delta^r_{[d]}$ to model dimensional binding and partial similarity assessment.

$$\begin{aligned}\Delta^r_{[d]}(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k) &= \left[\sum_{i \neq d} \omega_i |\vec{\mathbf{x}}_{ji} - \vec{\mathbf{x}}_{ki}|^r\right]^{1/r} \\ &= \sqrt[r]{\left[\sum_{i=1}^{D} \omega_i |\vec{\mathbf{x}}_{ji} - \vec{\mathbf{x}}_{ki}|^r\right] - \omega_d[|\vec{\mathbf{x}}_{jd} - \vec{\mathbf{x}}_{kd}|^r]}\end{aligned} \tag{7.7}$$

Eq. (7.7) computes the psychological distance between two stimuli ignoring their dth dimension ($1 \leqslant d \leqslant D$). In other words, it computes the partial psychological distance between the exemplars corresponding to the object-stimuli $\vec{x}_j, \vec{x}_k \in \mathrm{X}$, by excluding dimension d in the computation of the Minkowski generalized metric. For example, if the categorical stimulus X consists of four object-stimuli, we represent the partial pairwise distances between the four corresponding exemplars with respect to dimension d with the following partial distances matrix:

$$\mathbf{D}^r_{[d]}(\mathrm{X}) = \begin{bmatrix} \Delta^r_{[d]}(\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_1) & \Delta^r_{[d]}(\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2) & \Delta^r_{[d]}(\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_3) & \Delta^r_{[d]}(\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_4) \\ \Delta^r_{[d]}(\vec{\mathbf{x}}_2, \vec{\mathbf{x}}_1) & \Delta^r_{[d]}(\vec{\mathbf{x}}_2, \vec{\mathbf{x}}_2) & \Delta^r_{[d]}(\vec{\mathbf{x}}_2, \vec{\mathbf{x}}_3) & \Delta^r_{[d]}(\vec{\mathbf{x}}_2, \vec{\mathbf{x}}_4) \\ \Delta^r_{[d]}(\vec{\mathbf{x}}_3, \vec{\mathbf{x}}_1) & \Delta^r_{[d]}(\vec{\mathbf{x}}_3, \vec{\mathbf{x}}_2) & \Delta^r_{[d]}(\vec{\mathbf{x}}_3, \vec{\mathbf{x}}_3) & \Delta^r_{[d]}(\vec{\mathbf{x}}_3, \vec{\mathbf{x}}_4) \\ \Delta^r_{[d]}(\vec{\mathbf{x}}_4, \vec{\mathbf{x}}_1) & \Delta^r_{[d]}(\vec{\mathbf{x}}_4, \vec{\mathbf{x}}_2) & \Delta^r_{[d]}(\vec{\mathbf{x}}_4, \vec{\mathbf{x}}_3) & \Delta^r_{[d]}(\vec{\mathbf{x}}_4, \vec{\mathbf{x}}_4) \end{bmatrix}$$

Similarly, we can define the partial similarity between the two exemplars corresponding to the two object-stimuli – as is done in the GCM (Nosofsky, 1984) and in multidimensional scaling (Shepard et al., 1972; Kruskal & Wish, 1978) – as a monotonically decreasing function F of the partial distance between the two exemplars corresponding to the two object-stimuli.

$$S_{[d]}(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k) = F(\mu(\Delta^r_{[d]}(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k))) \tag{7.9}$$

As in Shepard (1987), we define subjective similarity as the negative exponent of the partial distance measure $\Delta^r_{[d]}(\vec{x}_j, \vec{x}_k)$ and set $r = 1$ (i.e., we use the city block metric in our example) as shown in Eq. (7.10).

$$S_{[d]}(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k) = e^{-\Delta^1_{[d]}(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k)} \tag{7.10}$$

In spite of using the above metric, we acknowledge the possibility that a different kind of function may be playing a similar role in the computation of partial similarities. Next we can construct the matrix of the pairwise partial psychological similarities between all four exemplars corresponding to the four object-stimuli in X as seen in (7.12):

$$\mathbf{S}_{[d]}(\mathrm{X}) = \begin{bmatrix} - & S_{[d]}(\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_2) & S_{[d]}(\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_3) & S_{[d]}(\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_4) \\ S_{[d]}(\vec{\mathbf{x}}_2, \vec{\mathbf{x}}_1) & - & S_{[d]}(\vec{\mathbf{x}}_2, \vec{\mathbf{x}}_3) & S_{[d]}(\vec{\mathbf{x}}_2, \vec{\mathbf{x}}_4) \\ S_{[d]}(\vec{\mathbf{x}}_3, \vec{\mathbf{x}}_1) & S_{[d]}(\vec{\mathbf{x}}_3, \vec{\mathbf{x}}_2) & - & S_{[d]}(\vec{\mathbf{x}}_3, \vec{\mathbf{x}}_4) \\ S_{[d]}(\vec{\mathbf{x}}_4, \vec{\mathbf{x}}_1) & S_{[d]}(\vec{\mathbf{x}}_4, \vec{\mathbf{x}}_2) & S_{[d]}(\vec{\mathbf{x}}_4, \vec{\mathbf{x}}_3) & - \end{bmatrix}$$

Again, as a process assumption, we have excluded reflexive or self-similarities in the diagonal of the partial distances matrix shown in (7.11). However, we include symmetric comparisons since we assume that they are processed by humans when assessing the overall homogeneity of a stimulus; besides, they add to the homogeneity of the stimulus as characterized by the categorical invariance (see Fig. 4) principle and the categorical invariance measure, and we wish to be consistent with both of these constructs (see Section 2).

Adding the values of the similarity matrix that correspond to differences within a chosen discrimination threshold τ_d for each dimension d we derive the following expression which is functionally analogous to the local homogeneity or local invariance operator defined in Section 2 (for any pair of objects $(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k)$ where $\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k \in \mathrm{X}$, $j \neq k$, and $j, k \in \{1, 2, \ldots, |\mathrm{X}|\}$):

$$H_{[d]}(\mathrm{X}) = \frac{\sum_{0 \leqslant \Delta^r_{[d]}(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k) \leqslant \tau_d} S_{[d]}(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k)}{|\mathrm{X}|} \tag{7.12}$$

Note that, in this article, $r = 1$ and $\tau_d = 0$ for all subjects and any dimension d; however, the latter threshold may also be treated as a free parameter that accounts for individual differences in classification performance. The assumption is that humans vary in their capacity to discriminate between stimuli and in their criterion for discriminating.

Lastly, we define the generalized structural manifold by Eq. (7.13). This construct is analogous to the global homogeneity construct defined in Eq. (2) of Section 2, except that it applies to both binary and continuous dimensions and is equipped with a distance discrimination threshold. It measures the perceived degree of global homogeneity of any stimulus set.

$$\Lambda(\mathrm{X}) = (H_{[d=1]}(\mathrm{X}), H_{[d=2]}(\mathrm{X}), \ldots, H_{[d=D]}(\mathrm{X})) \tag{7.13}$$

The overall degree of perceived global homogeneity or invariance of a categorical stimulus X defined over $D \geqslant 1$ dimensions and for any pair of objects $(\vec{x}_j, \vec{x}_k)$ (such that $\vec{x}_j, \vec{x}_k \in X$, $j \neq k$, and $j,\ k \in \{1, 2, \ldots, |X|\}$) is then given by Eq. (7.14) as follows:

$$\widehat{\Phi}(X) = \left[\sum_{d=1}^{D} [H_{[d]}(X)]^2\right]^{\frac{1}{2}} \tag{7.14}$$

Appendix C. Supplementary material

Supplementary documentation associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2013.05.008 and at http://www.scopelab.net/resources.htm. Computer programs associated with this article may be downloaded at http://www.scopelab.net/programs.htm.

References

Allport, A. (1987). Selection for action: some behavioural and neurophysiological considerations of attention and action, In: Perspective on perception and action, H. Heuer and A. Sanders (Ed.), Erlbaum.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442–481.

Bonatti, Luca. (Oct 1994). Propositional reasoning by model? *Psychological Review, 101*(4), 725–733.

Bourne, L. E. (1966). *Human conceptual behavior*. Boston: Allyn and Bacon.

Bush, R. R., Luce, R D., and Rose, R. M. Learning models for psychophysics. In R. C. Atkinson (Ed), Studies in mathematical psychology. Stanford: Stanford University Press, 1964. Pp. 201–217.

Daellenbach, H. G., & George, A. (1978). *Introduction to operations research techniques*. Boston: Allyn & Bacon.

Estes, W. K. (1994). Classification and cognition. *Oxford psychology series* (Vol. 22). Oxford: Oxford University Press.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature, 407*, 630–633.

Feldman, J. (2003). A catalog of Boolean concepts. *Journal of Mathematical Psychology, 47*(1), 98–112.

Feldman, J. (2006). An algebra of human concept learning. *Journal of Mathematical Psychology, 50*, 339–368.

Garner, W. R. (1963). Goodness of pattern and pattern uncertainty. *Journal of Verbal Learning and Verbal Behavior, 2*, 446–452.

Garner, W. R. (1974). *The processing of information and structure*. New York: Wiley.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1*, 225–241.

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

Goodwin, P. G., & Johnson-Laird, P. N. (2011). Mental models of Boolean concepts. *Cognitive Psychology, 63*, 34–59.

Haygood, R. C., & Bourne, L. E. Jr., (1965). Attribute-and-rule learning aspects of conceptual behavior. *Psychological Review, 72*, 175–195.

Higonnet, R. A., & Grea, R. A. (xxxx). *Logical design of electrical circuits*. New York, NY, USA: McGraw-Hill.

Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22–44.

Kruskal, Joseph. B., & Wish, Myron. (1978). *Multidimensional Scaling*. Beverly Hills: Sage.

Lafond, D., Lacouture, Y., & Mineau, G. (2007). Complexity mnimization in rule-based category learning: Revising the catalog of Boolean concepts and evidence for non-minimal rules. *Journal of Mathematical Psychology, 51*(2), 57–74.

Leyton, M. (1992). *Symmetry, causality, mind*. The MIT Press.

Love, B. C., & Medin, D. L. (1998). SUSTAIN: A model of human category learning. *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98), USA, 15*, 671–676.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*, 309–332.

Luce, D. (1995). Four Tensions Concerning Mathematical Modeling in Psychology. *Annual Review of Psychology, 46*, 1–27.

Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action*. Hillsdale: Erlbaum.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(1), 104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39–57.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. G. (1994a). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory and Cognition, 22*(3), 352–369.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994b). Rule-plus-exception model of classification learning. *Psychological Review, 101*, 53–79.

O'Brien, D. P., Braine, M. D., & Yang, Y. (1994). Propositional reasoning by mental models? Simple to refute in principle and in practice. *Psychological Review, 101*(4), 711–724.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26*, 303–343.

Prinz, W. (1983). Asymmetrical control areas in continuous visual search. In R. Groner, C. Menz, D. F. Fisher, & R. A. Monty (Eds.), *Eye movements and psychological functions: International views* (pp. 85–100). Hillsdale, N.J.: Erlbaum.

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology, 51*, 1–41.

Schneider, W. X. (1993). Space-based visual attention models and object selection: Constraints, problems, and possible solutions. *Psychological Research, 56*, 35–43.

Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review, 91*(4).

Shepard, R. N. (1987). Towards a universal law of generalization for psychological science. *Science, 237*, 1317–1323.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied, 75*(13), 1–42.

Shepard, R. N., Romney, A. K., & Nerlove, S. B. (Eds.). (1972). *Multidimensional scaling: Theory and applications in the behavioral sciences. Vol. I: Theory*. New York: Seminar Press.

Stevens, S. S. (1955). On the psychophysical law. *Psychological Review, 64*(3), 153–181.

Vigo, R. (2006). A note on the complexity of Boolean concepts. *Journal of Mathematical Psychology, 50*(5), 501–510.

Vigo, R. (2009). Categorical invariance and structural complexity in human concept learning. *Journal of Mathematical Psychology, 53*, 203–221.

Vigo, R. (2011a). Towards a law of invariance in human concept learning. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the Cognitive Science Society* (pp. 2580–2585). Austin, TX: Cognitive Science Society.

Vigo, R. (2011b). Representational information: A new general notion and measure of information. *Information Sciences, 181*, 4847–4859.

Vigo, R., Zeigler, D. E., & Halsey, P. A. (2013). Gaze and informativeness during category learning: Evidence for an inverse relation. *Visual Cognition, 21*(4), 446–476.