“The GIST of concepts” Supporting Materials

A. Experiment 3

General Description

In experiment 3 we obtained learnability ratings on ten trials consisting of the simultaneous presentation of six categorical stimuli. Each of the six categorical stimuli was a structure instance of one of the six category structures in the $3_2[4]$ structure family, and defined over the same set of dimensions. Also, each of the ten trials featured a unique set of structure instances (i.e., each of the six was generated by random selection from the entire population of all the possible instances associated with the particular structure).

Subjects

A total of 30 Ohio University undergraduates participated in the experiment. All subjects were assigned to all six structures associated with the $3_2[4]$ class of categorical stimuli.

Materials

An HP XW4600 workstation with a Dell 1708FP 17 inch flat panel LCD monitor (5 msec. response time) was used to display the stimuli. Stimuli were sets of flat flasks or bottles that varied in color (black or white), size (large or small), and shape (triangular or rectangular). Each stimulus consisted of the simultaneous display of six categorical stimuli, each conforming to one of the six structures in the $3_2[4]$ family of structures. Each categorical stimulus was placed within a rectangular boundary. Under each rectangle, a prompt for entering a magnitude judgment appeared. Moving from prompt to prompt was achieved using the tab key on the keyboard.

Procedure

During a training phase before the experiment, subjects performed three blocks of the classification task described under Experiment 1 (see main article) involving three randomly chosen structures from the $3_2[4]$ family of structures. After completing the training phase, subjects performed the experiment. Each trial consisted of the simultaneous presentation of six categorical stimuli. Each of the six categorical stimuli was a structure instance of one of the six category structures in the $3_2[4]$ structure family. Also, each of the ten trials featured a unique set of structure instances (i.e., each of the six was generated by random selection from the entire population of all the possible structure instances associated with the particular structure). As in Experiment 1, structure instances of each structure were randomly generated but in such a way that the assignment of dimensions was consistent across all six instances displayed in each trial.

Also, the location of each of the six structure instances was randomly determined per trial. Figure 1 below shows one of the trials generated by the computer program used in the experiment.

Before the experiment began, subjects were told to rank each of the six structure instances in terms of their relative degree of difficulty from the standpoint of correctly classifying their members. Subjects understood what we meant by “classification difficulty” from having experienced three classification blocks during the training phase. Specifically, subjects were asked to enter a number from 1 to 10 under each of the six categorical stimuli indicating how difficult, relative to the other five sets, they believed it would be to classify its elements. They were also told that they had to enter their degree of classification difficulty ratings using the prompts underneath each of the six categorical stimuli and type in their ratings. Furthermore, they were told that they had 30 seconds to do so. Figure 1 below shows one of the presented stimuli after the ratings were entered. After receiving these verbal instructions, subjects sat in front of the digital display so that their eyes were at approximately 2.5 feet from it. 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. The first of 10 trials began by the press of the space bar. The program recorded the six ratings per trial.

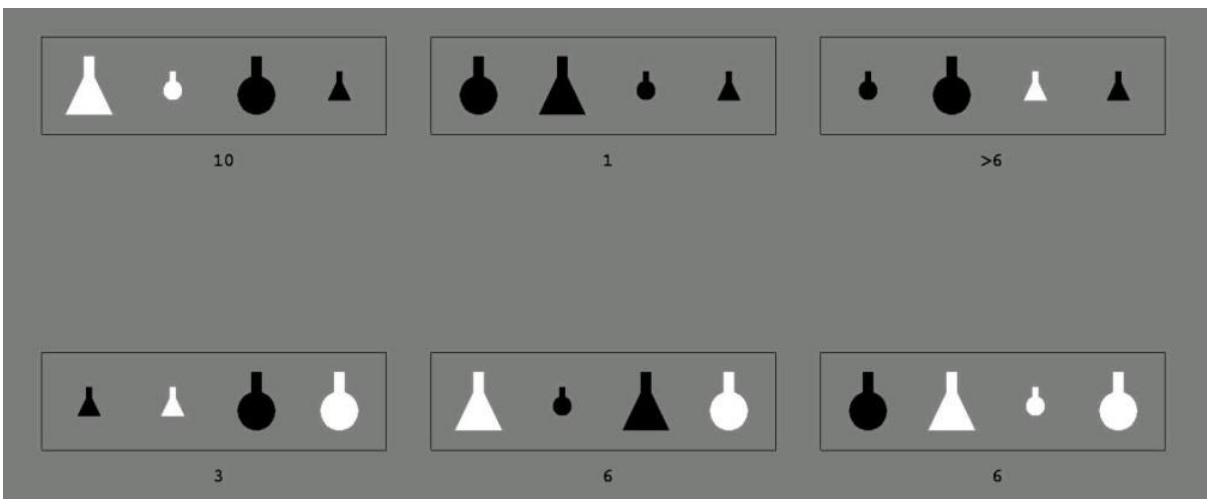


Figure 1. A computer display image from Experiment 3 is shown. Note that instances of all six structures corresponding to the $3_2[4]$ class of categorical stimuli are shown in each trial.

Results

The observed average magnitude judgments for each of the six $3_2[4]$ structures on a learning difficulty scale of 1 through 10 (where 10 indicated the highest degree of learning difficulty) are shown in Table 1.

3_2[4] Structure Type	I	II	III	IV	V	VI
Avg. Rating	1.7	4.3	5.5	5.6	5.5	6.3

Table 1. Difficulty ratings for the six 3_2[4] category structures

Pairwise t-tests showed no significant differences between the average classification difficulty ratings assigned to types III (5.5), IV (5.6), and V (5.5) by the 30 subjects, while there were significant differences between the rest. Thus, as seen in Table 1 above, the SHJ ordering of I < II < [III, IV, V] < VI was clearly present. We take this result as evidence in support of the thesis that humans have the capacity to form metajudgments about the learnability of concepts that are consistent with their classification performance. More specifically, humans are able to assess degree of concept learning difficulty by analyzing the structure of a stimulus set. In addition, we take this result to be consistent with the idea that the same pattern detection precursor (i.e., an ideotype) that accounts for learning difficulty judgments also informs the formation of concept representations such as rules and prototypes.

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

In the upcoming discussion we shall employ the following notation:

1) Let X be a categorical stimulus and $|\mathrm{X}|$ 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 $n = |\mathrm{X}|$).
3) Let the vector $\vec{\mathbf{x}}_j = (x_1, \ldots, x_D) \in \mathrm{X}$ 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 equation 5 below.

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:

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

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.

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

Equation 2 computes the psychological distance between two stimuli ignoring their *d*-th dimension $(1 \leq d \leq D)$. In other words, it computes the partial psychological distance between the exemplars corresponding to the object-stimuli $\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k \in \mathrm{X}$, 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:

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

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.

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

In equation 4 above, we have standardized the value $\Delta^r_{[d]}(\vec{\mathbf{x}}_j,\vec{\mathbf{x}}_k)$ in the [0, 1] interval using the following linear transformation μ where the $\mathtt{max}$ and $\mathtt{min}$ of a matrix are respectively its largest and smallest element, and the $\mathtt{max}(\mathbf{D}^r_{[d]}(\mathrm{X})\neq \mathtt{min}(\mathbf{D}^r_{[d]}(\mathrm{X}))$ for any *d* and *r*.

(5) $$\mu(\Delta^r_{[d]}(\vec{\mathbf{x}}_j,\vec{\mathbf{x}}_k))=\frac{\Delta^r_{[d]}(\vec{\mathbf{x}}_j,\vec{\mathbf{x}}_k)-\mathtt{min}(\mathbf{D}^r_{[d]}(\mathrm{X}))}{\mathtt{max}(\mathbf{D}^r_{[d]}(\mathrm{X}))-\mathtt{min}(\mathbf{D}^r_{[d]}(\mathrm{X}))}$$

This standardization will prove useful when we introduce the discrimination threshold parameter later in this section. As in Shepard (1987), we define subjective similarity as the negative exponent of the partial distance measure $\Delta^r_{[d]}(\vec{\mathbf{x}}_j,\vec{\mathbf{x}}_k)$ and set *r*=1 (i.e., we use the city block metric in our example) as shown in equation 6 below.

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

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 below:

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

Note that we have excluded reflexive comparisons from the diagonal of the partial distances matrix shown in 7 above even though we believe in their existence; however, we believe that they do not contribute, unlike symmetric comparisons, to detecting invariants. 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 principle and the categorical invariance measure, and we wish to be consistent with both of these constructs.

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 of the main article. 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 $j, k \in \{1, 2, ..., |\mathrm{X}|\}$:

(8) $$H_{[d]}(\mathrm{X}) = \frac{\sum\limits_{0 \leq \Delta^r_{[d]}(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k) \leq \tau_d} S_{[d]}(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k)}{|\mathrm{X}|}$$

The equation above defines the perceived degree of local homogeneity $H_{[d]}$ of a *D*-dimensional categorical stimulus X with respect to dimension *d*. $H_{[d]}(\mathrm{X})$ is the ratio between the sum of the similarities corresponding to distances that are zero or close to zero (depending on the value of the discrimination resolution threshold) in the matrix $\mathrm{D}^r_{[d]}$ (for a particular anchored dimension *d*) and the number of items in the categorical stimulus X. In other words, $H_{[d]}(\mathrm{X})$ is the ratio between: 1) the sum of the similarities in the matrix $\mathbf{S}^{\mathrm{X}}_{[d]}$ (for a particular anchored dimension *d*) that correspond to distances in the $[0, \tau_d]$ discrimination resolution interval, and 2) the number of items in the dataset X. When the partial distances are close to zero, the points are for all intent and purpose treated as perfectly similar or identical.

For example, take a categorical stimulus consisting of four binary dimensions and four objects as seen in Table 2 below and represented by $\mathrm{A} = \{1110,\ 1101,\ 1100,\ 1111\}$. Equation 9 below shows the matrix used to calculate the degree of partial homogeneity (with respect to dimension 3) of A when we let $\tau_1 = 0$ and *r*=1.

	D1	D2	D3	D4
O1	1	1	1	0
O2	1	1	0	1
O3	1	1	0	0
O4	1	1	1	1

Table 2. Matrix representing a categorical stimulus structure with four object-stimuli O1-O4 of four dimensions D1-D4.

$$(9)\qquad \mathbf{S}_{[3]}(\mathrm{A}) = \begin{bmatrix} - & S_{[3]}(\vec{\mathbf{x}}_1,\vec{\mathbf{x}}_2) & S_{[3]}(\vec{\mathbf{x}}_1,\vec{\mathbf{x}}_3) & S_{[3]}(\vec{\mathbf{x}}_1,\vec{\mathbf{x}}_4) \\ S_{[3]}(\vec{\mathbf{x}}_2,\vec{\mathbf{x}}_1) & - & S_{[3]}(\vec{\mathbf{x}}_2,\vec{\mathbf{x}}_3) & S_{[3]}(\vec{\mathbf{x}}_2,\vec{\mathbf{x}}_4) \\ S_{[3]}(\vec{\mathbf{x}}_3,\vec{\mathbf{x}}_1) & S_{[3]}(\vec{\mathbf{x}}_3,\vec{\mathbf{x}}_2) & - & S_{[3]}(\vec{\mathbf{x}}_3,\vec{\mathbf{x}}_4) \\ S_{[3]}(\vec{\mathbf{x}}_4,\vec{\mathbf{x}}_1) & S_{[3]}(\vec{\mathbf{x}}_4,\vec{\mathbf{x}}_2) & S_{[3]}(\vec{\mathbf{x}}_4,\vec{\mathbf{x}}_3) & - \end{bmatrix} = \begin{bmatrix} - & .37 & 1 & .37 \\ .37 & - & .37 & 1 \\ 1 & .37 & - & .37 \\ .37 & 1 & .37 & - \end{bmatrix}$$

Note that the computed matrix above contains 4 ones that represent four identical pairs of exemplars corresponding to four pairs of object-stimuli. Applying equation 9 above, we get 10.

$$(10)\qquad H_{[3]}(\mathrm{A}) = \frac{\displaystyle\sum_{0\le\Delta^{1}_{[3]}(\vec{\mathbf{x}}_j,\vec{\mathbf{x}}_k)\le 0} S_{[3]}(\vec{\mathbf{x}}_j,\vec{\mathbf{x}}_k)}{|\mathrm{A}|} = \frac{1+1+1+1}{4} = 1$$

Lastly, we define the generalized structural manifold by equation 11. This construct is analogous to the global homogeneity construct defined in equation 2 of section 2 (in main article), 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.

$$(11)\qquad \Lambda(\mathrm{X}) = \left(H_{[d=1]}(\mathrm{X}), H_{[d=2]}(\mathrm{X}), \ldots, H_{[d=D]}(\mathrm{X})\right)$$

We hypothesize that for every dimension d the discrimination resolution threshold τ_d will be a relatively small number dependent on the discriminatory capacities of the observer. Also, the above equation assumes that, for any d and any r, $\Delta^{r}_{[d]}(\vec{\mathbf{x}}_j,\vec{\mathbf{x}}_k) \in [0,\tau_d]$ are the only partial deltas that partake in determining the partial similarity matrices. Finally, since we standardized the partial distance metric in [0, 1], then we can also say that $\tau_d \in [0,1]$. To simplify our discussion, **in the remaining computations in this paper we shall let $\tau_d = 0$ for all subjects and any dimension** d; however, this value 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 (in this paper we shall not investigate this latter option: that is, we shall not try to derive estimates for τ_d). In either case, we assume that the primary goal of the human conceptual system is to optimize classification performance via the detection of invariants.

The overall degree of perceived global homogeneity or invariance of a categorical stimulus X defined over $D \geq 1$ dimensions and for any pair of objects $(\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k)$ (such that $\vec{\mathbf{x}}_j, \vec{\mathbf{x}}_k \in \mathrm{X}$, $j \neq k$, and $j,k \in \{1,2,...,|\mathrm{X}|\}$) is given by the generalized metric (Minkowski metric) as follows (where $\alpha_d \in [0,1]$ is the SK detection weight used to account for individual differences in perceived degree of difficulty; Vigo, 2011a):

$$\widehat{\Phi}(\mathrm{X}) = \left[\sum_{d=1}^{D} \left[\alpha_d H_{[d]}(\mathrm{X}) \right]^s \right]^{\frac{1}{s}} \tag{12}$$

And more specifically, in the GISTM, where α_d was not featured nor tested and where s=2 transforming the generalized metric into the Euclidean metric:

$$\widehat{\Phi}(\mathrm{X}) = \left[\sum_{d=1}^{D} \left[H_{[d]}(\mathrm{X}) \right]^2 \right]^{\frac{1}{2}} \tag{13}$$

C. <u>Non-parametric (i.e., no free parameters) Variants of the GISTM</u>

<u>Non-parametric Variant 1: *k* defined</u>

In GIST, categorical stimuli are represented by their primal sketches or ideotypes. Ideotypes are precursors to other concept representations. They carry the key structural information needed to form the various well-known concept representations. The distance between any ideotype and the **0** ideotype (see article) represents the degree of categorical invariance detected in the particular categorical stimulus. It, along with a scaling parameter *k*, determines the perceived homogeneity of the category structure that the ideotype encodes. The parameter *k* is called an ideotype discrimination parameter because its value is a function of the number of dimensions of the space.

Discrimination in ideotype space refers to the ability to discriminate between relational or structural information and not between object-exemplars as is the case in the GCM. For the latter model, an increase in dimensionality is consistent with higher discrimination. The opposite is true in GIST, where an increase in dimensionality is consistent with lower discrimination between an ideotype and the zero point in ideotype space (i.e., the **0** ideotype). Thus, the relationship between ideotype discriminability and number of dimensions is an inverse one. To explain, consider two categorical stimuli A and B of different dimensionality but with the same degree of categorical invariance or gestalt homogeneity. Let's say that A is two-dimensional and consists of a black triangle and a black square. On the other hand, B is three-dimensional and consists of a large black triangle and a small black square. Observers generally find A to be more homogeneous than B even though they both have the same unique diagnostic dimension of color. This can be easily explained by the fact that B has more relevant dimensions. In general, all other things being equal, the higher the dimensionality of a categorical stimulus, the less homogeneous it will appear to be, and the closer its ideotype will be to the **0** ideotype. That is, the more difficult it will be to discriminate it from the **0** ideotype.

More specifically, the degree of discrimination in ideotype space (as characterized by the parameter *k*) is inversely proportional to the number of dimensions associated with the ideotypes. In other words, the greater the dimensionality of the ideotype space, the smaller the discrimination parameter will be. To capture this capacity in terms of an ideal observer, we define the discrimination parameter *k* deterministically as D_0/D where D_0 represents a lower bound for the dimensionality of any ideotype and *D* is the number of dimensions of the ideotype ($D \geq 2$). The lower bound D_0 is set to two because, in principle, this is the smallest number of dimensions of any ideotype. This makes sense because an ideotype with only two dimensions has a discrimination index of 2/2=1 which represents the lowest impact on the perceived gestalt homogeneity (as measured by categorical invariance) of the categorical stimulus. Accordingly, as the number of dimensions increase, the discrimination index decreases, as does the impact that the number of dimensions have on perceived homogeneity: for example, for three dimensions the index is 2/3=.67, for four is 2/4=.5, and so on.

Thus, the GISTM becomes the non-parametric GISTM-NP by replacing the free scaling parameter *k* with D_0/D as follows:

$$\psi(\mathrm{X}) = pe^{-\frac{D_0}{D}\hat{\Phi}^2(\mathrm{X})}$$

On the other hand, when we let *k* be a free parameter, we can also learn to what extent the dimensionality of the categorical stimulus is directly affecting the classification performance of a particular subject. Indeed, for any number of relevant dimensions *D*, lower than ideal D_0/D values are consistent with an inability to process (i.e., an inability to detect SKs of) categorical stimuli defined over *D* dimensions. Also, note that D_0/D values are meant to describe the behavior of an *ideal* observer. They are rough approximations of the true values that characterize individual differences.

Furthermore, note that the value *k* is also a function of exposure time: that is, the more time a subject is exposed to a categorical stimulus, the more time the subject has to accurately detect its SKs in spite of its number of dimensions. Thus, the parameter *k* may also be consistent with how long subjects are exposed to a particular categorical stimulus. This latter interpretation is discussed in detail in the book entitled *Mathematical Principles of Human Conceptual Behavior*

(Vigo, 2014). However, although consistent with our findings, it is still subject to empirical verification.

The Non-Parametric GISTM-SE

Recall that structural equilibrium facilitates the processing of categorical stimuli. Thus, the non-parametric counterpart to the GISTM-SE should reflect this fact. It may be reasonable to assume that, consequently, for an ideal observer, there may be a slight gain in discriminating ideotypes so that a more accurate value for k would be $k=(D_0+1)/D$, where, as before, $D_0 = 2$. In other words, the baseline dimension in the discrimination index has increased to three. This is shown in the equation below (where η is the degree of structural equilibrium of the categorical stimulus X):

$$\psi(\mathrm{X}) = \frac{pe^{-\frac{(D_0+1)}{D}\hat{\Phi}^2(\mathrm{X})}}{\eta}$$

Using the λ "core definition" of structural equilibrium (see explanation in the following section) we get:

$$\psi(\mathrm{X}) = \frac{pe^{-\frac{(D_0+1)}{D}\hat{\Phi}^2(\mathrm{X})}}{\lambda}$$

Thus, in the context of structural equilibrium, this second bound characterization of k makes more sense and is consistent with our findings. Tests with categorical stimuli of five dimensions or more will determine the validity of this k value. In the interim, we have confirmed that both, the non-parametric GISTM-NP and the nonparametric GISTM-SE-NP, account for approximately 85% of the variance in the data of Experiment 1 and predict the $3_2[4]$ family ordering precisely.

Non-parametric Variant 2: the form of "raw (initial) perceived complexity" redefined

It was shown in section 5 that both the GISTM and the GISTM-SE make remarkably accurate predictions with the application of a single free scaling parameter. Many laws of physics function in this manner. For example, the law of radioactive decay, which also has an exponential decay form, involves a single scaling parameter value per substance type. The GISTM and GISTM-SE are far more frugal, requiring a single scaling parameter value per dimensional space and fitting the data nearly as accurately with a single scaling parameter value for all of the dimensionality spaces. Notwithstanding, one can dispense with free parameters altogether by approximating the role that the scaling parameter plays in the GISTM. For example, suppose that the subjective raw complexity of a stimulus is not defined in terms of the number of objects in the categorical stimulus but, instead, in terms of the number of potential similarity comparisons between the objects of a categorical stimulus as described in the process account under section 3. The total number of such comparisons is equal to the number of items *p* in the categorical stimulus squared (if we count reflexive comparisons). While the scaling parameter *k* compensated for any possible way of scaling the initial perceived raw complexity of a categorical stimulus (see technical appendix A) before its decay, we can now eliminate it in the following nonparametric version of the model as follows:

$$\psi(\mathrm{X}) = \frac{p^2}{e^{\hat{\Phi}^2(\mathrm{X})}}$$

In spite of achieving accurate fits, this model does not have the ability to account for individual differences nor of handling different scales of initial raw complexity. However, the model is shown in order to illustrate that, in principle, one can achieve accurate fits with the exponential decay function of invariance, even without having to estimate a single free parameter.

Note that, with respect to the GISTM, the estimate of k=.54 from the data on all 84 structures of experiment 1 is consistent with, and approximately equivalent to, the GISTM-NP2 which measures initial raw complexity in terms of the number of possible pairwise comparisons between the objects in a categorical stimulus (i.e., p^2 when including reflexive comparisons). The following simple derivation shows this:

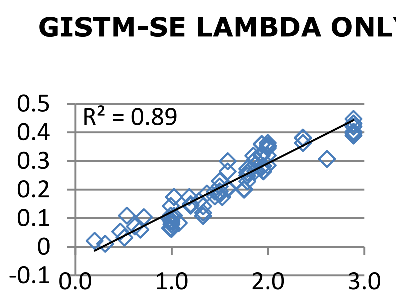
$$\psi(\mathrm{X}) = \psi(\mathrm{X}) = \left[\frac{p}{e^{k\hat{\Phi}^2(\mathrm{X})}}\right]^2 = \frac{p^2}{\left[e^{.5\hat{\Phi}^2(\mathrm{X})}\right]^2} = \frac{p^2}{e^{\hat{\Phi}^2(\mathrm{X})}}$$

This GISTM-NP accounted for about 86% of the variance in the 84 structures data of experiment 1 (R^2=0.86, p<.0001, *RMSE*=.05, rs =.93), about 84% of the variance in the 76 structures data (R^2=0.84, p<.0001, *RMSE*=.05, rs =.93), and again, less accurately, for about 64% of the variance in Feldman's data on the 76 structures (R^2=0.63, p<.0001, *RMSE*=.05, rs =.80).

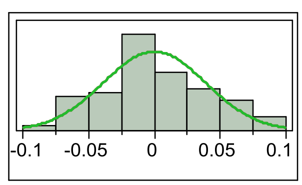
D. <u>Structural Equilibrium as a Moderator</u>

In this section we discuss two measures of degree of *structural equilibrium* (SE): $\lambda(\mathrm{X})$ and $\eta(\mathrm{X})$. The reader is referred to the main article for an explanation of structural equilibrium. Formally, we define the degree of structural equilibrium λ of the ideotype of a categorical stimulus X (written $\lambda(\mathrm{X})$) as the proportion of essential or diagnostic dimensions in the ideotype of X (i.e., dimensions whose structural manifold kernel values are zero) plus one (in order to avoid potential problems with division by zero). For example, the ideotype corresponding to a categorical stimulus X conforming to the $3_2[4]$-VI structure type has three diagnostic dimensions since its structural manifold is (0, 0, 0). Hence, the proportion of zero kernels (3) to number of kernels (3) in its ideotype is 1. To this number we add 1, making the degree of SE $\lambda(\mathrm{X})=2$. The ideotype corresponding to this categorical stimulus is in *absolute* structural equilibrium, a state where its dimensions relate minimally to one another and are structurally stable (i.e., play an equal structural role). On the other hand, the ideotype of a categorical stimulus X conforming to the $3_2[4]$-V structure type has a structural manifold value of (0, ½, ½), and its proportion of zero SKs is 1/3. Once again, to this number we add 1 so that $\lambda(\mathrm{X})\cong 1.33$. Note that, henceforth, we shall abbreviate the degree of structural equilibrium $\lambda(\mathrm{X})$ of X as λ whenever there is no ambiguity as to the categorical stimulus in question. Incorporating $\lambda(\mathrm{X})$ (abbreviated as λ) into Equation 5 of the main article results in the following variant of the GISTM which we shall refer to as the λ-based GISTM-SE. Immediately below the equation (in Figure 2) are the predictions made by this λ-based GISTM-SE on 76 out of the 84 category structures of Experiment 1.

$$\psi(\mathrm{X})=\frac{p}{\lambda e^{k\hat{\Phi}^2(\mathrm{X})}}=\frac{pe^{-k\hat{\Phi}^2(\mathrm{X})}}{\lambda}$$



Shapiro-Wilk W Test on Residuals; **RMSE: 0.039**; W=0.982; Prob<W=0.39



3[4] FAMILY ORDERING (k3=1.35; k4=.80):

YES

I	.2
II	.6
III	1.5
IV	1.5
V	1.5
VI	2

Figure 2. Fitness of the λ-based GISTM-SE on 76 out of the 84 structures studied in Experiment 1. Please, note that the model performed virtually identically with respect to the 84-structures data.

Finally, to make the definition of λ "more psychological", we constrain it further by assuming that its contribution in lowering concept learning difficulty is nonlinear. In particular, we assume that the ease of determining the diagnostic role of the dimensions in categorical stimuli increases rapidly at first as the degree of structural equilibrium increases and then continues to increase but slows down greatly asymptotically. This makes intuitive sense when one considers that a single zero SK value in an ideotype would have a relatively greater impact in lowering its dimensional confoundedness than two, even though, in general, as the number of zeros increase, the more the dimensional confoundedness is lowered. The function that seems to best capture this asymptotic relationship in the real number interval $[1,2]$ is $x^{1/x}$ as shown below (Figure 3), although other functions may be feasible.

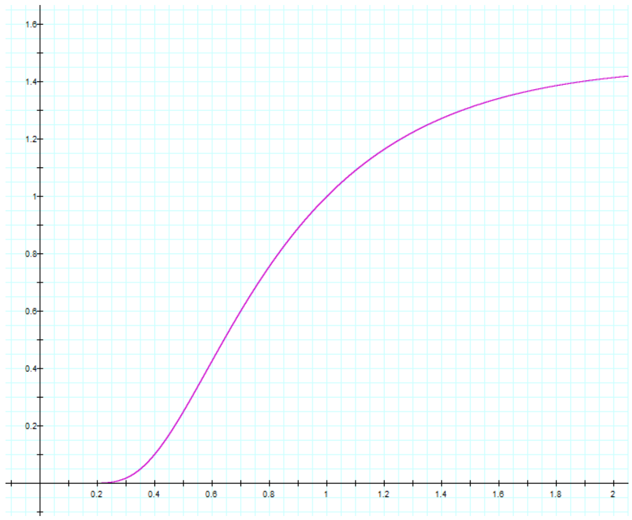


Figure 3. The function $x^{1/x}$ constrained in the [1, 2] interval.

Thus, we define the *coefficient of structural equilibrium* as follows: $\eta(\mathrm{X}) = \lambda^{1/\lambda}$ where $\lambda \in [1,2]$ is a real number. Using our previous two examples, for a 3_2[4]-VI structure type, $\eta(\mathrm{X}) = \lambda^{1/\lambda} = 2^{1/2} \cong 1.41$. Likewise, for a 3_2[4]-V structure type categorical stimulus, $\eta(\mathrm{X}) = \lambda^{1/\lambda} \cong 1.33^{1/1.33} \cong 1.24$. Incorporating $\eta(\mathrm{X})$ (abbreviated as η) into Equation 5 of the main article results in the following variant of the GISTM which we shall refer to as the GISTM-SE; in the main article, we test both the GISTM and the GISTM-SE.

$$\psi(\mathrm{X}) = \frac{p}{\eta e^{k\hat{\Phi}^2(\mathrm{X})}} = \frac{pe^{-k\hat{\Phi}^2(\mathrm{X})}}{\eta}$$

Lastly, it should be noted that in using λ as the coefficient of structural equilibrium instead of η in the GISTM-SE, we computed nearly identical R^2s for all 84 and 76 structures tested in our data and achieved the same key predictions (including precisely predicting the canonical 3_2[4] family ordering; see Figure 4 below); however, the assumption of nonlinearity underlying η seem more plausible. Lambda and eta values for each of the 84 structures tested appear in Table 3 below. Although only the η definition of structural equilibrium is tested in the article, we feel that the more parsimonious (i.e., without extra assumptions) λ version is more elegant and attractive. Notwithstanding, we believe that both definitions of structural equilibrium merit further study.

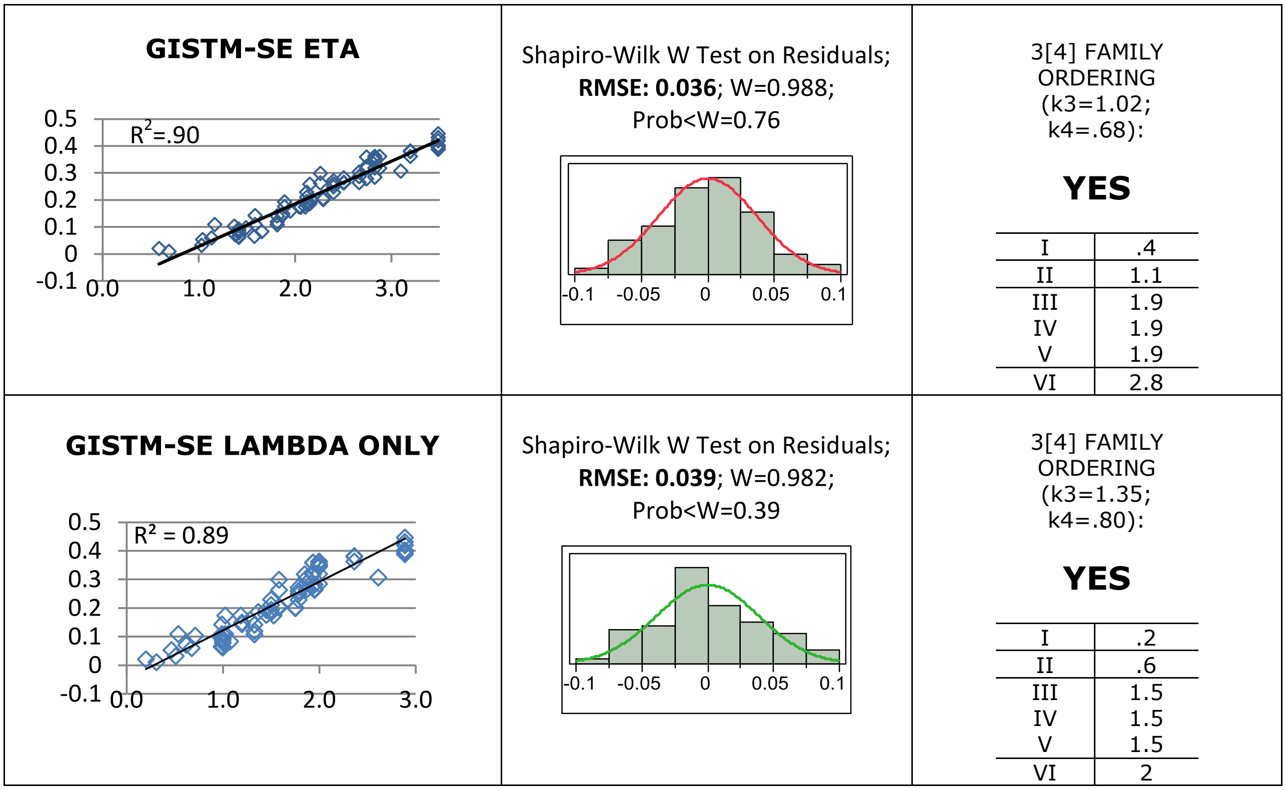


Figure 4. Fitness comparison of the GISTM-SE with the lambda versus the eta coefficient on 76 out of the 84 structures studied in Experiment 1. Note that their performance is very similar. Please, note that the two models also performed virtually identically with respect to the 84-structures data.

Table 3. Lambda (λ) and Eta (η) values for the GISTM-SE (k parameter values are k_2=1.3, k_3=1.02, and k_4=.69):

TYPE	BOOLEAN FUNCTIONAL DESCRIPTION OF STRUCTURE	STRUCTURAL MANIFOLD	$\hat{\Phi}$	λ	η
U-2_2[1]-1*	*(x′y′)*	(0,0)	0.00	2	1.41
U-2_2[2]-1*	*((x′y′)+(x′y))*	(0,1)	1.00	1.5	1.31
U-2_2[2]-2*	*((x′y′)+(xy))*	(0,0)	0.00	2	1.41
U-3_2[1]-1*	*(x′y′z′)*	(0,0,0)	0.00	2	1.41
U-3_2[2]-1	*((x′y′z′)+(x′y′z))*	(0,0,1)	1.00	1.67	1.36
U-3_2[2]-2	*((x′y′z′)+(x′yz))*	(0,0,0)	0.00	2	1.41
U-3_2[2]-3	*((x′y′z′)+(xyz))*	(0,0,0)	0.00	2	1.41
U-3_2[3]-1	*((x′y′z′)+(x′y′z)+(x′yz′))*	(0,.67,.67)	0.94	1.33	1.24
U-3_2[3]-2	*((x′y′z′)+(x′y′z)+(xyz′))*	(0,0,.67)	0.67	1.67	1.36
U-3_2[3]-3	*((x′y′z′)+(x′yz)+(xy′z))*	(0,0,0)	0.00	2	1.41

3_2[4]-1	*((x′y′z′)+(x′y′z)+(x′yz′)+(x′yz))*	(0,1,1)	1.41	1.33	1.24
3_2[4]-2	*((x′y′z′)+(x′y′z)+(xyz′)+(xyz))*	(0,0,1)	1.00	1.67	1.36
3_2[4]-3	*((x′y′z′)+(x′y′z)+(x′yz′)+(xy′z))*	(.50,.50,.50)	0.87	1	1.00
3_2[4]-4	*((x′y′z′)+(x′y′z)+(x′yz′)+(xy′z′))*	(.50,.50,.50)	0.87	1	1.00
3_2[4]-5	*((x′y′z′)+(x′y′z)+(x′yz′)+(xyz))*	(0,.50,.50)	0.71	1.33	1.24
3_2[4]-6	*((x′y′z′)+(x′yz)+(xy′z)+(xyz′))*	(0,0,0)	0.00	2	1.41
U-4_2[1]-1 *	*(x′y′z′w′)*	(0,0,0,0)	0.00	1.75	1.38
U-4_2[2]-1	*((x′y′z′w′)+(x′y′z′w))*	(0,0,0,1)	1.00	1.75	1.38
U-4_2[2]-2	*((x′y′z′w′)+(x′y′zw))*	(0,0,0,0)	0.00	2	1.41
U-4_2[2]-3	*((x′y′z′w′)+(x′yzw))*	(0,0,0,0)	0.00	2	1.41
U-4_2[2]-4	*((x′y′z′w′)+(xyzw))*	(0,0,0,0)	0.00	2	1.41
U-4_2[3]-1	*((x′y′z′w′)+(x′y′z′w)+(x′y′zw′))*	(0,0,.67,.67)	0.94	1.5	1.31
U-4_2[3]-2	*((x′y′z′w′)+(x′y′z′w)+(x′yzw′))*	(0,0,0,.67)	0.67	1.75	1.38
U-4_2[3]-3	*((x′y′z′w′)+(x′y′z′w)+(xyzw′))*	(0,0,0,.67)	0.67	1.75	1.38
U-4_2[3]-4	*((x′y′z′w′)+(x′y′zw)+(x′yz′w))*	(0,0,0,0)	0.00	2	1.41
U-4_2[3]-5	*((x′y′z′w′)+(x′y′zw)+(xyz′w′))*	(0,0,0,0)	0.00	2	1.41
U-4_2[3]-6	*((x′y′z′w′)+(x′y′zw)+(xyz′w))*	(0,0,0,0)	0.00	2	1.41
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	1.5	1.31
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	1.25	1.20
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	1.25	1.20
U-4_2[4]-4	*((x′y′z′w′)+(x′y′z′w)+(x′y′zw′)+(x′yzw))*	(0,0,.50,.50)	0.71	1.5	1.31
U-4_2[4]-5	*((x′y′z′w′)+(x′y′z′w)+(x′y′zw′)+(xyz′w′))*	(0,0,.50,.50)	0.71	1.5	1.31
U-4_2[4]-6	*((x′y′z′w′)+(x′y′z′w)+(x′y′zw′)+(xyz′w))*	(0,0,.50,.50)	0.71	1.5	1.31
U-4_2[4]-7	*((x′y′z′w′)+(x′y′z′w)+(x′y′zw′)+(xyzw))*	(0,0,.50,.50)	0.71	1.5	1.31
U-4_2[4]-8	*((x′y′z′w′)+(x′y′z′w)+(x′yzw′)+(x′yzw))*	(0,0,0,1)	1.00	1.75	1.38
U-4_2[4]-9	*((x′y′z′w′)+(x′y′z′w)+(x′yzw′)+(xy′zw′))*	(0,0,0,.50)	0.50	1.75	1.38
U-4_2[4]-10	*((x′y′z′w′)+(x′y′z′w)+(x′yzw′)+(xy′zw))*	(0,0,0,.50)	0.50	1.75	1.38
U-4_2[4]-11	*((x′y′z′w′)+(x′y′z′w)+(x′yzw′)+(xyzw′))*	(.50,0,0,.50)	0.71	1.5	1.31
U-4_2[4]-12	*((x′y′z′w′)+(x′y′z′w)+(x′yzw′)+(xyzw))*	(0,0,0,.50)	0.50	1.75	1.38
U-4_2[4]-13	*((x′y′z′w′)+(x′y′z′w)+(xyzw′)+(xyzw))*	(0,0,0,1)	1.00	1.75	1.38
U-4_2[4]-14	*((x′y′z′w′)+(x′y′zw)+(x′yz′w)+(x′yzw′))*	(0,0,0,0)	0.00	2	1.41
U-4_2[4]-15	*((x′y′z′w′)+(x′y′zw)+(x′yz′w)+(xy′z′w))*	(0,0,0,0)	0.00	2	1.41
U-4_2[4]-16	*((x′y′z′w′)+(x′y′zw)+(x′yz′w)+(xy′zw′))*	(0,0,0,0)	0.00	2	1.41
U-4_2[4]-17	*((x′y′z′w′)+(x′y′zw)+(x′yz′w)+(xyzw′))*	(0,0,0,0)	0.00	2	1.41

U-4_2[4]-18	((x'y'z'w')+(x'y'zw)+(xyz'w')+(xyzw))	(0,0,0,0)	0.00	2	1.41
U-4_2[4]-19	((x'y'z'w')+(x'y'zw)+(xyz'w)+(xyzw'))	(0,0,0,0)	0.00	2	1.41
D-3_2[1]-1 *	((x'y'z)+(x'yz')+(x'yz)+(xy'z')+(xy'z)+(xyz')+(xyz))	(.86,.86,.86)	1.48	1	1.00
D-3_2[2]-1	((x'yz')+(x'yz)+(xy'z')+(xy'z)+(xyz')+(xyz))	(.67,.67,1)	1.37	1	1.00
D-3_2[2]-2	((x'y'z)+(x'yz')+(xy'z')+(xy'z)+(xyz')+(xyz))	(.67,.67,.67)	1.15	1	1.00
D-3_2[2]-3	((x'y'z)+(x'yz')+(x'yz)+(xy'z')+(xy'z)+(xyz'))	(.67,.67,.67)	1.15	1	1.00
D-2_2[1]-1 *	((x'y)+(xy')+(xy))	(.67,.67)	0.94	1	1.00
D-3_2[3]-1	((x'yz)+(xy'z')+(xy'z)+(xyz')+(xyz))	(.40,.80,.80)	1.20	1	1.00
D-3_2[3]-2	((x'yz')+(x'yz)+(xy'z')+(xy'z)+(xyz))	(.40,.40,.80)	0.98	1	1.00
D-3_2[3]-3	((x'y'z)+(x'yz')+(xy'z')+(xyz')+(xyz))	(.40,.40,.40)	0.69	1	1.00
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')+(xyzw))	(.93,.93,.93,.93)	1.87	1	1.00
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	1	1.00
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	1	1.00
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	1	1.00
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	1	1.00
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	1	1.00
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	1	1.00
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))	(.77,.77,.77,.92)	1.62	1	1.00
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	1	1.00
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)+(xyzw')+(xyzw))	(.77,.77,.77,.77)	1.54	1	1.00
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')+(xyzw')+(xyzw))	(.77,.77,.77,.77)	1.54	1	1.00
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	1	1.00
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	1	1.00
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	1	1.00
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	1	1.00
D-4_2[4]-5	((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')+(xyzw))	(.67,.67,.83,.83)	1.51	1	1.00
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')+(xyzw')+(xyzw))	(.67,.67,.83,.83)	1.51	1	1.00
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'))	(.67,.67,.83,.83)	1.51	1	1.00
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	1	1.00
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	1	1.00
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')+(xyzw))	(.67,.67,.67,.83)	1.42	1	1.00
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	1	1.00
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	1	1.00
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') +(xy'zw)+(xyz'w')+(xyz'w))	(.67,.67,.67,1)	1.53	1	1.00

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')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.67,.67,.67,.67)	1.33	1	1.00
D-4_2[4]-15	*((x'y'z'w)+(x'y'zw')+(x'yz'w')+(x'yzw')+(x'yzw)+(xy'z'w')+(xy'zw')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw')+(xyzw))*	(.67,.67,.67,.67)	1.33	1	1.00
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	1	1.00
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')+(xy'zw)+(xyz'w')+(xyz'w)+(xyzw))*	(.67,.67,.67,.67)	1.33	1	1.00
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	1	1.00
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	1	1.00

E. General Notes

1. With the exception of the results shown in Figure 7, when using regression statistics to test how well the models fitted the data, we removed one outlier (structure 4_2[4]-11 in down parity) from the Feldman data set (Feldman, 2000) and one outlier (structure 4_2[4]-13 in up parity) from our dataset because they had a studentized residual score of approximately four each – a number indicative of considerable noise. Please, note that, on average, not removing this single outlier from either data set yields very slightly lower R^2s for the models (i.e., about two to three percentage points lower).

2. The program that implements the "mental" Λ operator (i.e., computes the structural manifold of any dimensionally-defined categorical stimulus) may be downloaded from the following webpage.

http://www.scopelab.net/resources.htm

F. Typographical errors in "The GIST of Concepts" (2013), *Cognition*

(Please note that these minor typographical errors may have already been corrected in your copy of the article)

1. On page 142 in the bottom-left paragraph the sentence in the middle of the paragraph should read, "…their members are displayed sequentially before…" and not "…their members are displayed sequential before …"
2. On page 144 at the top-right paragraph the sentence should begin: "In GIST, we …" and not "In the GIST, we …"
3. In the caption of Table 2 on page 150 the last sentence should read, "…the D-4_2[4] family…" instead of "…the D-4_{12}[4] family…" D-4_2[4] is an abbreviation for the more formal notation 4_2[12] (this notation was introduced in section 1.2). Also, note that there are two different types of informal labels used to label the down parity families of Table 2 (page 150). To resolve this inconsistency, keep in mind the following: if, for example, the family in up parity is

the $4_2[4]$ family, its down parity counterpart may be labeled in one of three ways: "D-$4_2[4]$", "D-$4_2[12]$", or using the label "$4_2[12]$" (the latter label follows the precise notation introduced in 1.2).

4. In the same Table 2 on page 150 for the "Number of Unique Structures" column there should be three unique structures (not four) for both the U-$3_2[3]$ and D-$3_2[3]$ structure families and four unique structures (not three) for both the U-$4_2[2]$ and D-$4_2[2]$ structure families.
5. In the caption of Table 5 on page 157 the first sentence should read, "…tested in Experiment 2" and not "…tested in Experiment 3."
6. On page 158 (second column) it should read "As in …" and not "As is…"
7. In Appendix B, it states in the sixth item of the list of notations (page 161) that μ is defined in Equation 7.14. This sentence should be deleted. The precise definition of μ may be found in the supplementary documents.