Volume 63 / Number 5 / 2016

Experimental Psychology

Editor-in-Chief
Christoph Stahl

Editors
Tom Beckers
Arndt Bröder
Adele Diederich
Chris Donkin
Gesine Dreisbach
Kai Epstude
Magda Osman
Manuel Perea
Klaus Rothermund
Samuel Shaki
Sarah Teige-Mocigemba

hogrefe

Contents

© 2016 Hogrefe Publishing

Experimental Psychology

Thank you for choosing to publish your article in a Hogrefe journal. We are pleased to provide this e-offprint of the final published article for your personal use, which includes sharing the e-offprint with your coauthors, with other individuals upon request for their personal use, and as part of a grant application or a thesis.

If you wish to post your article to an institutional or subject-based repository, on your personal website, or within a scholarly communication network (SCN), please use either the submitted manuscript (before peer review) or the accepted manuscript (after peer review) in accordance with the publication release for your article and the document "Guidelines on sharing and use of articles in Hogrefe journals" on the journal's web page at www.hogrefe.com/j/exppsy.

Thank you again for publishing your work with Hogrefe.

hogrefe

Author's personal copy (e-offprint)

Research Article

Constructing and Deconstructing Concepts

On the Nature of Category Modification and Unsupervised Sorting Behavior

Charles A. Doan and Ronaldo Vigo

Consortium for the Advancement of Cognitive Science, Department of Psychology, Ohio University, Athens, OH, USA

Abstract: Several empirical investigations have explored whether observers prefer to sort sets of multidimensional stimuli into groups by employing one-dimensional or family-resemblance strategies. Although one-dimensional sorting strategies have been the prevalent finding for these unsupervised classification paradigms, several researchers have provided evidence that the choice of strategy may depend on the particular demands of the task. To account for this disparity, we propose that observers extract relational patterns from stimulus sets that facilitate the development of optimal classification strategies for relegating category membership. We conducted a novel constrained categorization experiment to empirically test this hypothesis by instructing participants to either add or remove objects from presented categorical stimuli. We employed generalized representational information theory (GRIT; Vigo, 2011b, 2013a, 2014) and its associated formal models to predict and explain how human beings chose to modify these categorical stimuli. Additionally, we compared model performance to predictions made by a leading prototypicality measure in the literature.

Keywords: concepts, representational information, invariance, category, family resemblance

Among the many unsolved problems in human categorization research, an intriguing and relevant one to our everyday lives addresses how observers naturally categorize multidimensional stimuli in their environment without supervision – specifically, how do people group complex stimuli without the aid of corrective feedback? For example, we may interpret a politician's platform from a select few issues and naturally group her into a specific political party. Similarly, as a physician we may identify known markers or symptoms in a patient to determine the likelihood of a particular disease diagnosis. In both of these examples, and more generally, we group in part by comparing our current interpretation of a stimulus to previous interpretations of similar stimuli. With the exception of receiving information from experts or from diagnostic imaging procedures, we rarely receive timely feedback on the accuracy of these and more common "everyday" grouping decisions.

Empirical investigations into the conceptual processes underlying such decisions involve asking participants to divide a disorganized set of related stimuli into exclusive and coherent groups in whatever way seems natural to them (Ahn & Medin, 1992; Medin, Wattenmaker, & Hampson, 1987; Milton & Wills, 2004; Regehr & Brooks, 1995). The research goal is to analyze the composition of the resulting groups and the robust result is that observers prefer using one stimulus dimension (e.g., shape) to help divide stimuli (Medin et al., 1987; Milton & Wills, 2004).

Recently, researchers have investigated whether individuals use this strategy when the task is slightly constrained – specifically, if participants are given a subset of stimuli known to be in the group and then they have to sort remaining stimuli. This is analogous to a physician using test results from previously confirmed examples to help her classify the condition of a current patient. Few studies have constrained the task in this manner, but those that have present examples to aid sorting (Milton & Wills, 2004; Regehr & Brooks, 1995). For example, with the *match-to-standards* procedure we may present a member (e.g., exemplar) of at least two categories and instruct participants to sort a stack of remaining stimuli cards into one of the two shown categories, one-at-a-time (Milton & Wills, 2004; Regehr & Brooks, 1995). Or, for a related task referred to as the *match-from-array* procedure (Regehr & Brooks, 1995), we may present the entire stimulus set (minus a preselected prototype placed at the top of the array) to each participant and instruct them to select an

© 2016 Hogrefe Publishing

DOI: 10.1027/1618-3169/a000337

Author's personal copy (e-offprint)

object from the array that likely belongs (1) to the opposite category or (2) to the same category as the prototype. Empirical results suggest participants may implement strategies consistent with a "family-resemblance" principle (henceforth denoted FR) – whereby participants presumably maximize within-category similarity while minimizing between-category similarity of resulting categories (Milton & Wills, 2004; Regehr & Brooks, 1995, Rosch & Mervis, 1975). Thus, instead of using one stimulus dimension, observers may use multiple dimensions and display sensitivity to the similarity relationships between stimulus cards.

Although these constrained-categorization studies seem to demonstrate participants sorting differently, we provide evidence inconsistent with the hypothesis that participants use FR strategies. Rather, we posit that observers are implicitly detecting invariance patterns inherent to the categorical stimuli (i.e., stimulus sets) presented in the aforementioned tasks. Intuitively, this is consistent with the notion that observers seek to minimize the concept learning difficulty of each resulting categorical stimulus. To test this hypothesis, we conducted a novel constrained-categorization task in which participants were instructed to add (or remove) one object to (or from) a presented categorical stimulus, thereby changing the relational context of each object. Importantly, many empirical investigations demonstrate similarity assessment – a core cognitive process theoretically linked with conceptual processing – to be context-dependent (Gati & Tversky, 1984; Medin, Goldstone, & Gentner, 1993; Tversky, 1977). For this reason, many of the categorical stimuli (i.e., sets of objects) shown to participants in the current tasks were logically equivalent – preserving the contextual relations among the stimuli composing each categorical stimulus. However, we also changed contextual relationships by interspersing logically distinct categorical stimuli during the task. Thus, for theories and models to account for the current experimental results they must be able to account for addition and removal behavior on both contextually similar and contextually distinct categorical stimuli. Toward this aim, we computed predictions from two different measures: a nonparametric measure from Generalized Representational Information Theory (GRIT; Vigo, 2013a, 2013b, 2014) and a popular prototype measure designed to capture family-resemblance intuitions quantitatively (Estes, 1986; Nosofsky & Zaki, 2002; Vigo & Basawaraj, 2013).[1] Next, we briefly discuss the theoretical underpinnings of these measures before describing the empirical tasks and associated results.

Generalized Representational Information Theory

To account for previous sorting results and the results of the current study we shall use a theory and associated measure of information referred to as generalized representational information theory (GRIT; Vigo, 2011b, 2013a, 2014). The theory and measure are based on a theory of categorization, named Generalized Invariance Structure Theory (GIST; Vigo, 2013b, 2014), that has been successful in accounting for classification behavior across a wide variety of categorical stimuli and provides a set of powerful and natural tools for the analysis of the sorting task in this report.[2]

Generalized Invariance Structure Theory is founded upon the core idea that observers implicitly detect specific relational patterns in categorical stimuli, termed "categorical invariants." These reveal key structural information about the category such as which stimulus dimensions are necessary in order to form logical rules and which dimensions are redundant and can be discarded when forming such rules. More specifically, the greater the proportion of categorical invariants detected with respect to a particular dimension, the less diagnostic (and therefore more redundant) the dimension will be perceived to be. This proportion of categorical invariants is known in GIST and GRIT as degrees of partial invariance; thus, zero degrees of partial invariance indicate complete diagnosticity for a dimension, while one degree indicates absolute redundancy. Further, these partial invariance values (named "structural kernels" or "SKs") are components of a vector referred to as a "structural manifold" which determines the location in psychological space of an "ideotype." Ideotypes are concept representations that contain the invariance structure information of categorical stimuli.

As an example, Figure 1 displays a categorical stimulus consisting of three stimulus objects, each being defined over the three binary stimulus dimensions of *shape* (triangle or square), *color* (white or black), and *size* (small or large). To compute the partial invariance values comprising the structural manifold of this categorical stimulus, we perform three perturbations (one for each dimension) where, for each, we assign the opposite dimensional value to each object (see "Perturbation" column of Figure 1). We compute the SKs for each dimension by assessing the proportion of stimuli in the perturbed set that match stimuli in the original categorical stimulus. As shown in Figure 1, this procedure results in SKs of 0/3, 0/3, and 2/3 for the three dimensions, respectively; and upon being arranged as a vector results in a

[1] We computed predictions for the simplicity model (Pothos & Chater, 2002, 2005) and found that it did not perform significantly better than the prototype measure.

[2] The mathematical and technical details of both theories can be found in Vigo (2013a, 2013b, 2014).

 © 2016 Hogrefe Publishing

Author's personal copy (e-offprint)

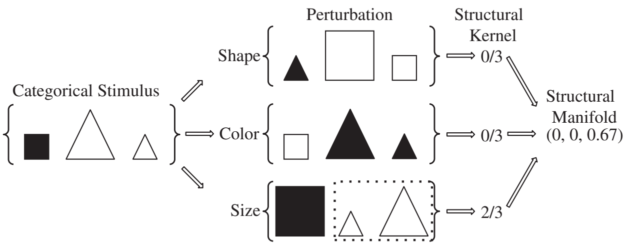

Figure 1. An example categorical stimulus (left) and the perturbation process (middle) are shown for computing the partial invariance values (i.e., "structural kernels" or "SKs") for each stimulus dimension. The SKs are arranged in a vector referred to as a structural manifold (right) that determines the location of the ideotype representation in psychological space. For this categorical stimulus, categorical invariants are only present upon perturbing the "Size" dimension and are enclosed by a dotted box.

structural manifold of (0, 0, .67). Thus, the size dimension is mostly redundant and, therefore, not as diagnostic as the shape and color dimensions for determining category membership – a fact that, according to GIST, observers require when engaged in the formation of logical rules.

In GIST, the above perturbation process is also explained, defined, and generalized to continuous and multi-valued dimensions in terms of a lower level cognitive mechanism of ideotype formation referred to as "dimensional binding." Under dimensional binding, observers assess the "partial similarity" between object-stimuli in a category of objects by binding or suppressing, one-at-a-time, each of their dimensions. This is made possible by rapid goal-oriented attention shifting from one dimension to another along with the sustained and deliberate "unattending" or "suppression" of each dimension. Ultimately, the described mechanism results in the detection of the proportion of invariants associated with each dimension as illustrated by the higher-level process of invariance detection discussed above.

Lastly, the degree of concept learning difficulty of a categorical stimulus (F) is derived as a mathematical function of the structural manifold of the category and the size of the category (|F|). Regarding the former, the overall degree of categorical invariance ($\hat{\Phi}$) of a categorical stimulus is measured by computing the Euclidean distance between the ideotype of the categorical stimulus (F) and the ideotype of the categorical stimulus with SKs equal to 0. The mathematical function shown in Equation 1 below is known as the nonparametric variant of the GISTM (denoted GISTM-NP). The model uses a deterministic quantity to characterize the discrimination between concepts in psychological space: namely, D_0/D where D_0 stands for the minimum number of baseline dimensions needed to define a nontrivial categorical stimulus (D_0 = 2) and D is the number of dimensions that define the categorical stimulus (F). Roughly speaking, holding size constant, the model predicts that categorical stimuli for which observers detect a higher degree of overall invariance are conceptually easier to learn.

$$\Psi(\mathrm{F}) = |\mathrm{F}| \cdot e^{-\left(\frac{D_0}{D}\right)\hat{\Phi}^2(\mathrm{F})}. \qquad (1)$$

Vigo (2013b, 2014) makes a final modification to the GISTM-NP by incorporating a moderating factor referred to as "structural equilibrium" or SE. Structural equilibrium describes the condition or state of total incoherence of a categorical stimulus, where the dimensions of the objects are logically unrelated. The degree of structural equilibrium (λ) of a categorical stimulus is simply the percentage of diagnostic dimensions (i.e., zero SKs) in the structural manifold (+1 to avoid division by zero) and is denoted λ_F. Importantly, a high degree of SE exerts a positive moderating effect in concept learnability by facilitating the identification of diagnostic dimensions that are subsequently processed by a rule-formation system (Vigo, 2014). The model shown below in Equation 2 (the GISTM-NPE) utilizes this measure of SE and is the core model to characterize and measure representational information (see Equation 3) in the current paper.[3]

$$\Psi(\mathrm{F}) = \frac{|\mathrm{F}| \cdot e^{-\left(\frac{D_0}{D}\right)\hat{\Phi}^2(F)}}{\lambda_\mathrm{F}}. \qquad (2)$$

Generalized Representational Information Theory introduces a new construct of information referred to as "representational information" whereby each object in a categorical stimulus is an information carrier in the following sense: it carries representational information about the category containing it. More specifically, the amount of representational information of an object stimulus x in a categorical stimulus (F) is defined in GRIT as the percent rate of change in the perceived complexity (i.e., degree of concept learning difficulty) of a stimulus whenever the object stimulus x is removed from the categorical stimulus (F). Formally, letting x be the only element of the set R, F be the categorical stimulus in question, and G = F-R, using Equation 2, we write:

$$\hbar(\mathrm{R}|\mathrm{F}) = \frac{\Psi(\mathrm{G}) - \Psi(\mathrm{F})}{\Psi(\mathrm{F})} = \frac{\frac{|\mathrm{G}| \cdot e^{-\left(\frac{D_0}{D}\right)\hat{\Phi}^2(\mathrm{G})}}{\lambda_\mathrm{G}} - \frac{|\mathrm{F}| \cdot e^{-\left(\frac{D_0}{D}\right)\hat{\Phi}^2(\mathrm{F})}}{\lambda_\mathrm{F}}}{\frac{|\mathrm{F}| \cdot e^{-\left(\frac{D_0}{D}\right)\hat{\Phi}^2(\mathrm{F})}}{\lambda_\mathrm{F}}}. \qquad (3)$$

In general, a decrease in concept learning difficulty upon the removal of an object from a category means that the

[3] The generalized variant of the model contains one free scaling parameter (k) and appears as such: $\Psi(\mathrm{F}) = \frac{|\mathrm{F}| \cdot e^{-k\hat{\Phi}^2(\mathrm{F})}}{\eta_\mathrm{F}}$ (for details see Vigo, 2013b, 2014).

© 2016 Hogrefe Publishing

Author's personal copy (e-offprint)

object conveys less information about the category than a similar object that results in an increase in concept learning difficulty if removed. Intuitively, removing an object that conveys much information about a category should make that category more complex and more difficult to learn after it is removed. Relating to examples presented previously, this is similar to removing the political candidate that best represents the positions of their political party or by removing the most useful diagnostic images from a physician's referential database. It is easy to imagine that if either of these situations occurred, it would prove more difficult to classify or diagnose newly encountered politicians and patients, respectively.

With respect to the construction task in the current study, the GRIT-NPE predicts observers will add the object from the set of available objects that results in the most positive rate of change to perceived complexity if that object were removed from the category (see "GRIT-NPE Predictions" row of Table 1).[4] Regarding the deconstruction task, the GRIT-NPE predicts observers will remove the object from the category that results in the most negative rate of change to perceived complexity upon the removal of the object (see "GRIT-NPE Predictions" row of Table 2). In general, GRIT predicts subjects selecting the object that minimizes the concept learning difficulty of the resulting categorical stimulus.

Multiplicative-Prototype Model Measure of Prototypicality

In contrast, several classification models containing prototype measures have been developed to predict and explain human classification performance (Estes, 1986; Nosofsky & Zaki, 2002; Rosch & Mervis, 1975). However, because the current experimental task is not a standard classification task, we utilize only the prototype measure at the heart of one of these classification models to calculate predictions. In particular, we utilized the prototypicality measure of the Nosofsky and Zaki (2002) variant of the multiplicative-prototype model, recently referred to as the MPMP-2 (Vigo & Basawaraj, 2013; Vigo, Zeigler, & Halsey, 2013).

The measure assumes that human beings abstract a prototype of the category by assessing the most frequently occurring dimensional values among the category members and determine membership by assessing the similarity between a given member and the abstracted prototype. Quantitatively, Equation 4 shows that the similarity between exemplar (j) and the category prototype (Q) is calculated by taking the negative exponent of the weighted distance between the dimensional values of the exemplar and the prototype (Nosofsky & Zaki, 2002).

$$S(j, Q) = e^{-c\left[\sum_{d}^{D} w_d \cdot \left|x_{jd} - P_{Qd}\right|\right]}. \tag{4}$$

To compare the performance of the MPMP-2 to the performance of the GRIT-NPE, we set the scaling parameter (c) equal to 3 – the number of stimulus dimensions composing the clock stimuli in the current study (Nosofsky, 1984, 1986; Vigo & Basawaraj, 2013). Additionally, we set each of the attention weights (w_d) to 1/3, thus assuming attention would be divided equally among the dimensions across participants.[5]

For the current experimental tasks, the measure provided in Equation 4 was used to calculate the degree of similarity of each category exemplar (clock in the current study) to its category prototype. The measure predicts that the objects most and least similar to the category prototype are those most likely to be added to and to be removed from presented categorical stimuli, respectively (see "MPMP-2 Predictions" row of Tables 1 and 2).

Given explication of the theories underlying the models, we next present the methodology for constructing the categorical stimuli in the current study. Following, we expound on the experimental task used to test modifications to these categorical stimuli. We then present the results and focus on how well each model accounts for the observed choices in each task. Finally, we demonstrate a connection between the results of the current study to results of previous studies implementing various experimental sorting paradigms.

Method

Participants

We recruited 44 undergraduates (29 females, 15 males) from introductory psychology courses at Ohio University.

[4] When calculating a GRIT object (e.g., clock in the current study) prediction for the construction task, we first construe the clock as belonging to the 3-clock categorical stimulus. Then we determine the amount of representational information the clock conveys about the 4-clock categorical stimulus by measuring the rate of change in perceived degree of concept learning difficulty when the clock is removed from the category. Importantly, this procedure is used in GRIT when determining the rate of change in learning difficulty after an object is added to a categorical stimulus. Indeed, in the basic GRIT measure that we are using, the informativeness of an added object is measured by determining the impact of its removal from the new category formed by adding the object. This is consistent with a plausible cognitive mechanism involving observers who first imagine the object in the category (i.e., that first add the object to the category) and then compare this imagined larger category to the original.

[5] We also estimated optimal scaling and attention weight parameters for the data observed in the current study. The analysis can be found in the Discussion section and the parameter estimates can be found in Table 4.

 © 2016 Hogrefe Publishing

Author's personal copy (e-offprint)

Participants were at least 18 years of age and received partial course credit for participating.

Stimuli

An HP XW4600 workstation equipped with a Dell 1708FP 15″ flat panel LCD monitor (5 ms response time; Dell Inc., Round Rock, TX) was used to display sets of realistic clocks defined over separable dimensions (Vigo & Doan, 2015; see Figure 2). Importantly, we followed the operationalization of well-defined objects, categories, and category types as they have been traditionally studied by human categorization researchers (Feldman, 2000, 2003; Goodwin & Johnson-Laird, 2013; Nosofsky, 1984; Shepard, Hovland, & Jenkins, 1961; Vigo, 2009, 2011a, 2013b, 2014). Well-defined objects are assembled from a preset number of dimensions and dimensional values. For the current study, the clocks were defined by factorially combining the three binary dimensions of *shape* (x = round = 0, x' = square = 1), *number of tick marks* (y = few = 0, y' = many = 1), and *color of hands* (z = white = 0, z' = black = 1).

If we logically add objects together using the disjunctive rule of Boolean algebra, we can create a well-defined category description in disjunctive normal form (DNF; Vigo, 2006, 2009, 2011b). The DNF function $xyz + x'yz + x'y'z'$, which may also be represented as {000, 100, 111}, defines a category consisting of a round clock with few tick marks and white hands, a square clock with few tick marks and white hands, and a square clock with many tick marks and black hands. Vigo (2011b) refers to functions in DNF as concept functions and concept functions specifying the same logical structure among a set of objects are equivalent and form a category type. For example, the concept function $xyz + xy'z + xyz' + xy'z'$ is logically equivalent to the concept function $xyz + x'yz + xyz' + x'yz'$. Intuitively, the equivalence of these two concept function instances can be demonstrated by mapping the objects of each function onto the vertices of a three-dimensional Boolean cube; the former occupying the x-dimensional plane and the latter occupying the y-dimensional plane.

Furthermore, classes (e.g., families) of category types are formed depending on the number of dimensions that define their objects (D), number of values per each dimension (n), and number of objects that constitute them (p; Feldman, 2000, 2003; Vigo, 2009, 2011b, 2013b). An adopted notation is the $D_n[p]$ family of category types ($D_n[p]$ – type; Vigo, 2013b). For example, the concept function $xyz + xy'z + x'y'z$ is an instance of category type I of the $3_2[3]$ family of category types (denoted $3_2[3]$ – I) and consists of three objects being defined over three binary-valued stimulus dimensions. The families of category types tested were those consisting of three and five stimulus objects defined over three binary dimensions (denoted $3_2[3]$ and $3_2[5]$). Each of these two families contains three unique category types (cf. Feldman, 2003; Higonnet & Grea, 1958). Tables 1 and 2 display the concept functions that were used to construct the clock stimuli for the six distinct category types shown to participants in the current study.

Procedure

Upon obtaining consent, the researcher communicated following instructions to each participant:

> "You are taking part in a study on category construction and category deconstruction. The two tasks will take place on the computer where you will be shown images of groups of clocks.
>
> For the first task, you will be shown a group of clocks on the top half of the computer screen and a number of individual clocks on the bottom half. Your task is to select the one clock on the bottom half of the computer screen that you believe is the best clock to be added to the group of clocks on the top half. You will have multiple trials, but the task should take no longer than 10 min to complete.
>
> Once you complete this task, you will begin the second task. For the second task, you will be shown a group of clocks in the middle of the computer screen. Your task is to select the one clock from the group that you believe should be removed from the group. You will have multiple trials, but this task should also take no longer than 10 min to complete."

After questions were answered, the researcher guided the participant to the computer where they completed the two tasks. The order of the two tasks was counterbalanced across participants and the dependent variables of interest were the observed choices per category type and associated response times to make such choices. To preserve the unsupervised nature of the two tasks, we had participants determine for themselves how to interpret which clock was the "best" to add or which clock "should be" removed each trial. This protocol is similar to traditional and modified sorting tasks whereby researchers instruct participants to sort stimulus cards into groups by whatever way seems natural to them (Imai & Garner, 1965; Medin et al., 1987; Regehr & Brooks, 1995). And as will be evident below, we observed a consistent pattern of addition and removal decisions across participants – suggesting that participants were interpreting the instructions similarly between the two tasks.

© 2016 Hogrefe Publishing

Author's personal copy (e-offprint)

Category Construction Task

This task consisted of presenting eight randomly selected category instances for each of the three category types belonging to the $3_2[3]$ family of category types twice and in random order. Additionally, the spatial arrangement of the clocks was randomized every instance. The first "Concept Function" column of Table 1 displays DNF descriptions of the 24 randomly selected category instances used in the present study. To clarify, each of these 24 category instances can be constructed by systematically selecting the appropriate three clocks from the eight possible clock stimuli constructed from 3 binary-valued stimulus dimensions (3_2). For example, the first category instance of the $3_2[3]$ – I category type has a DNF description of $xyz + xy'z + x'y'z$. If we assume *shape* (x = round = 0, x' = square = 1), *number of tick marks* (y = few = 0, y' = many = 1), and *color of hands* (z = white = 0, z' = black = 1), then we can construct the following category of clock stimuli: round clock with few tick marks and white hands, round clock with many tick marks and white hands, and square clock with many tick marks and white hands.[6]

For each trial, the three clocks belonging to each category instance were presented on the top half of the computer screen and enclosed by a narrow white box. Simultaneously, the remaining five clocks were displayed individually below the three-object category instance in a randomly determined order (see column 1 of Figure 2). Per trial, participants were given 20 s to select one of the five individual clocks to add to the category above. Between trials, a neutral gray screen (RGB: 127, 127, 127) with a "+" fixation symbol in the middle of the screen appeared for 3 s. The task concluded after participants completed the 48 trials.

Category Deconstruction Task

This task consisted of presenting eight randomly selected category instances for each of the three category types belonging to the $3_2[5]$ family of category structures twice and in random order. As above, the spatial arrangement of the clocks was randomized every instance and the DNF descriptions of the 24 randomly selected category instances can be found in the "Concept Function" column of Table 2.

For each trial, the five objects of each category instance were presented in the middle of the computer screen and enclosed by a narrow white box (see column 2 of Figure 2). The task was the same as the construction task above except participants were instructed to select one of the five objects to remove from the category. Again, the task concluded after participants completed the 48 trials.

Results

General

Statistical analyses were conducted on 39 of the 44 recruited participants. The raw data can be found in the Electronic Supplementary Material, ESM 1. Three of the excluded participants were uncooperative and were "clicking though" the experiment to finish early, while two of the excluded participants were nonnative English speakers not proficient in English and who experienced difficulty understanding the instructions. Despite analyses including the data from these participants resulting in similar conclusions as purported below, we removed these participants from the analyses for two main reasons. First, these participants displayed greater inconsistency with their choices – in that they failed to choose the same object for the second random presentation of a category instance more often than not. They were consistent 43% (95% CI [.35, .52]) of the time in the construction task and 28% ([.20, .36]) of the time in the deconstruction task. This is compared to a consistency measure of 69% ([.66, .72]) and 55% ([.52, .58]) among the remaining 39 participants for both tasks, respectively. Second, these five participants selected objects in the construction task on average in 3.09 s ([2.78, 3.41]) and in 3.50 s ([3.07, 3.94]) for the deconstruction task. As above, these averages are significantly lower when compared to the response time averages observed among the remaining 39 participants for the construction task (M = 5.95 s, SD = 2.98 [5.82, 6.09]) and deconstruction task (M = 6.46 s, SD = 4.03, [6.28, 6.64]). The substantial decreases in consistency and response times support the aforementioned notions of "clicking through" the experiment and experiencing difficulty understanding the instructions.

Interestingly, when asked, participants were unable to linguistically describe a general strategy they used to make addition and removal decisions. A few participants did allude to employing a one-dimensional strategy but could not extrapolate beyond this description. Furthermore, analyses suggest the deconstruction task was more difficult than the construction task. First, the mean response time in the deconstruction task was significantly

[6] Note how the order of selection of the dimensions and dimensional values is arbitrary. For example, we can form structurally equivalent category instances by assuming *number of tick marks* (x = few = 0, x' = many = 1), *shape* (y = square = 0, y' = round = 1), and *color of hands* (z = black = 0, z' = white = 1).

 © 2016 Hogrefe Publishing

Author's personal copy (e-offprint)

Table 1. Concept functions, observed choices, and model predictions for the category construction task

$D_n[p]$ – Type	Concept function	Concept function[a]	O_1	O_2	O_3	O_4	O_5
$3_2[3]$ – I	$xyz + xy'z + x'y'z$	$xyz' + xy'z' + x'yz + x'yz' + x'y'z'$	3	1	68	5	1
	$xyz + x'yz + x'yz'$	$xy'z + x'y'z + xyz' + xy'z' + x'y'z'$	1	1	64	9	3
	$xyz' + xy'z + xy'z'$	$x'yz' + x'y'z' + xyz + x'yz + x'y'z$	4	0	68	5	0
	$xyz' + x'yz + x'yz'$	$xy'z' + x'y'z' + xyz + xy'z + x'y'z$	1	1	68	7	1
	$xy'z + x'yz + x'y'z$	$x'yz' + x'y'z' + xyz + xyz' + xy'z'$	1	3	66	6	2
	$xy'z' + x'yz' + x'y'z'$	$x'yz + x'y'z + xyz' + xyz + xy'z$	5	1	66	5	1
	$x'yz + x'yz' + x'y'z$	$xy'z + xyz + x'y'z' + xy'z' + xyz'$	1	0	74	1	1
	$x'yz + x'y'z + x'y'z'$	$xyz + xy'z + x'yz' + xyz' + xy'z'$	0	0	71	6	1
		Total	16	7	545	44	10
		GRIT-NPE predictions	−.49	−.49	**.57**	−.42	−.49
		MPMP-2 predictions	.14	**.37**	.14	.05	.14
$3_2[3]$ – II	$xyz + xyz' + x'y'z'$	$x'yz' + x'yz + x'y'z + xy'z' + xy'z$	2	12	60	2	2
	$xyz + xy'z' + x'yz$	$xyz' + x'yz' + x'y'z' + xy'z + x'y'z$	4	9	54	4	6
	$xyz + x'y'z + x'y'z'$	$xy'z + xy'z' + xyz' + x'yz + x'yz'$	1	17	51	3	6
	$xyz' + x'yz' + x'y'z$	$x'yz + xyz + xy'z + x'y'z' + xy'z'$	1	9	63	2	3
	$xy'z + xy'z' + x'yz'$	$x'y'z' + x'y'z + x'yz + xyz' + xyz$	1	19	52	4	2
	$xy'z + x'yz + x'yz'$	$xyz + xyz' + xy'z' + x'y'z + x'y'z'$	4	11	58	4	0
	$xy'z + x'yz' + x'y'z$	$x'y'z' + xy'z' + xyz' + x'yz + xyz$	4	8	61	1	4
	$xy'z' + x'yz + x'yz'$	$xyz' + xyz + xy'z + x'y'z' + x'y'z$	2	8	61	3	4
		Total	19	93	460	23	27
		GRIT-NPE predictions	−.45	−.38	**.09**	−.45	−.38
		MPMP-2 predictions	**.37**	.14	.05	**.37**	.14
$3_2[3]$ – III	$xyz + xy'z' + x'yz'$	$xy'z + xyz' + x'y'z + x'yz + x'y'z'$	2	3	61	8	4
	$xyz + xy'z' + x'y'z$	$xyz' + xy'z + x'yz' + x'y'z' + x'yz$	9	2	56	8	2
	$xyz + x'yz' + x'y'z$	$x'y'z' + x'yz + xy'z' + xyz' + xy'z$	6	3	52	12	5
	$xyz' + xy'z + x'yz$	$xy'z' + xyz + x'y'z' + x'yz' + x'y'z$	5	5	53	9	6
	$xyz' + xy'z + x'y'z'$	$xyz + xy'z' + x'yz + x'y'z + x'yz'$	8	2	50	8	9
	$xyz' + x'yz + x'y'z'$	$x'y'z + x'yz' + xy'z + xyz + xy'z'$	8	4	57	5	3
	$xy'z + x'yz + x'y'z'$	$x'yz' + x'y'z + xyz' + xy'z' + xyz$	4	3	61	7	3
	$xy'z' + x'yz' + x'y'z$	$x'yz + x'y'z' + xyz + xy'z + xyz'$	4	4	55	5	9
		Total	46	26	445	62	41
		GRIT-NPE predictions	−.30	−.38	**−.25**	−.30	−.30
		MPMP-2 predictions	.14	**1**	.05	.14	.14

Notes. Each instance was displayed twice, in a randomly determined order, throughout the specific experiment for each participant. To construct the instances with the clock stimuli shown to participants in the current study, use *shape* (x = round = 0, x' = square = 1), *number of tick marks* (y = few = 0, y' = many = 1), and *color of hands* (z = white = 0, z' = black = 1). The values that are bolded represent the objects predicted to be selected by the GRIT-NPE and the MPMP-2.
[a]Concept function for the five objects that were possible additions to the 3-object category instances.

longer than the mean response time in the construction task, $t(3{,}707) = 4.38$, $p < .001$, $d = 0.14$ [.08, .21]. Second, participants were more consistent with their choices in the construction task than in the deconstruction task, $X^2(1, N = 1{,}162) = 13.66$, $p < .001$. Despite the significant differences in consistency and response times between the two tasks, participants were selecting the same clocks and creating the same category types of the $3_2[4]$ family of category types regardless of whether they were adding or removing an object. However, as will be evident below, the proportions corresponding to the most-often chosen clock for the deconstruction task were consistently lower than similar proportions observed for the construction task.

Category Construction

General

Analyses were conducted on 1,864 choices. Although there was the potential for an analysis of 1,872 choices (39 participants × 48 trials), there were eight no responses. We observed 622 addition decisions to category type $3_2[3]$ – I, 622 addition decisions to category type $3_2[3]$ – II, and 620 addition decisions to category type $3_2[3]$ – III.

Author's personal copy (e-offprint)

Table 2. Concept functions, observed choices, and model predictions for the category deconstruction task

$D_n[p]$ – Type	Concept function	O_1	O_2	O_3	O_4	O_5
$3_2[5]$ – I	$x'yz + x'yz' + xy'z + x'y'z + x'y'z'$	1	1	68	3	3
	$x'yz' + xyz' + x'y'z + x'y'z' + xy'z'$	5	7	56	6	4
	$x'y'z + x'y'z' + xyz + xy'z + xy'z'$	4	7	59	3	5
	$xyz' + xyz + x'y'z' + x'yz' + x'yz$	3	8	60	5	2
	$x'yz' + x'yz + xy'z' + xyz' + xyz$	3	3	58	13	1
	$xyz + xyz' + x'y'z + xy'z + xy'z'$	4	3	62	4	4
	$xy'z' + xy'z + x'yz' + xyz' + xyz$	5	6	59	3	4
	$xy'z + xy'z' + x'yz + xyz + xyz'$	6	6	61	2	3
	Total	31	41	483	39	26
	GRIT-NPE predictions	1.43	1.43	**–.21**	1.15	1.43
	MPMP-2 predictions	.37	**.14**	.37	1	.37
$3_2[5]$ – II	$x'y'z + xy'z + xy'z' + x'yz' + xyz'$	7	5	46	13	7
	$x'yz + x'y'z + xy'z + xyz' + xy'z'$	11	5	31	9	18
	$xy'z' + x'y'z' + x'y'z + xyz + x'yz$	11	10	34	11	11
	$xyz + x'yz + x'yz' + xy'z' + x'y'z'$	12	5	39	15	7
	$xy'z + xyz + x'yz + x'y'z' + x'yz'$	14	12	22	8	21
	$x'y'z + x'y'z' + xy'z' + xyz + xyz'$	5	7	40	6	19
	$x'y'z + x'yz + xyz + xy'z' + xyz'$	10	8	28	17	14
	$x'yz' + xyz' + xyz + x'y'z + xy'z$	6	10	44	9	9
	Total	76	62	284	88	106
	GRIT-NPE predictions	1.43	1.15	**.23**	1.43	1.15
	MPMP-2 predictions	**.14**	.37	1	**.14**	.37
$3_2[5]$ – III	$xy'z + xyz' + x'y'z + x'yz + x'y'z'$	12	15	19	18	11
	$xyz' + xy'z + x'yz' + x'y'z' + x'yz$	5	10	32	21	9
	$x'y'z' + x'yz + xy'z' + xyz' + xy'z$	8	14	32	14	5
	$xy'z' + xyz + x'y'z' + x'yz' + x'y'z$	11	8	34	14	8
	$xyz + xy'z' + x'yz + x'y'z + x'yz'$	6	16	29	9	17
	$x'y'z + x'yz' + xy'z + xyz + xy'z'$	7	15	31	9	15
	$x'yz' + x'y'z + xyz' + xy'z' + xyz$	8	7	34	20	9
	$x'yz + x'y'z' + xyz + xy'z + xyz'$	3	14	31	13	16
	Total	60	99	242	118	90
	GRIT-NPE predictions	1.15	1.43	**1**	1.15	1.15
	MPMP-2 predictions	.37	**.05**	1	.37	.37

Notes. Each instance was displayed twice, in a randomly determined order, throughout the specific experiment for each participant. To construct the instances with the clock stimuli shown to participants in the current study, use *shape* (x = round = 0, x' = square = 1), *number of tick marks* (y = few = 0, y' = many = 1), and *color of hands* (z = white = 0, z' = black = 1). The values that are bolded represent the objects predicted to be selected by the GRIT-NPE and the MPMP-2.

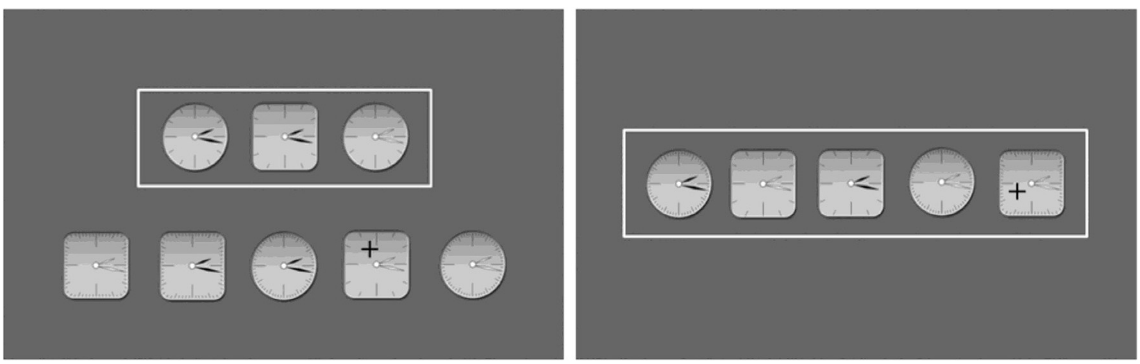

Figure 2. Displayed are example screenshots of an experimental trial for the category construction task (A) and the category deconstruction task (B). The "+" symbol represents the mouse cursor and the participants' selection for these instances.

 © 2016 Hogrefe Publishing

Author's personal copy (e-offprint)

The raw aggregate data for each category type can be found in the "Total" rows of Table 1.

Chi-Square Goodness-of-Fit tests on the observed choices to supplement the three category types were conducted and were highly significant, least favorable $X^2(4, N = 620) = 1{,}044$, $p < .0001$. Thus, participants were not selecting each of the five objects equally to supplement the 3-object category types. Post hoc tests with Bonferroni-Holm corrections revealed O_3 being chosen more often than the other four clocks for instances of each category type, least favorable $X^2(1, N = 553) = 243.6$, $p < .0001$. Further, O_4 was chosen more often than the remaining three objects for type $3_2[3]$ – I, least favorable $X^2(1, N = 60) = 13.07$, $p < .001$. Additionally, O_2 was chosen more often than the remaining three objects for type $3_2[3]$ – II, least favorable $X^2(1, N = 120) = 36.3$, $p < .001$. Lastly, O_4 was chosen more often than O_2 for type $3_2[3]$ – III, $X^2(1, N = 88) = 14.7$, $p < .001$. No other post hoc tests were statistically significant.

Model Comparison

You may recall that the GRIT-NPE predicts that the most likely object to be added to the three-object category types is the object that results in the most positive rate of change in subjective concept learning difficulty if that object would be removed from the category (i.e., .57, .09, and −.25 for the three category types, respectively). Inspecting Table 1 we find that the GRIT-NPE measure predicts the most-often chosen object for each of the three category types. Furthermore, Spearman correlations of .89, .95, and .89 indicate the predicted strong ordinal agreement within the three category types, respectively.

Conversely, the MPMP-2 predicts that the most likely object to be added is the object that is most similar to the category prototype, with higher values indicating higher similarity (i.e., .37, .37, and 1 for the three category types, respectively). Upon inspecting Table 1 we find that the nonparametric MPMP-2 fails to account for the most-often chosen object even once. Additionally, Spearman correlations of −.67, −.95, and −.89 indicate negative ordinal relationships between MPMP-2 predictions and observed choice frequencies.

Consistent with the GRIT-NPE, we found participants choosing to add the object that satisfied the one-dimensional rule for the $3_2[3]$ – I category type. That is, when participants were provided with the categorical stimulus {000, 010, 110} and were asked to choose one object from the set {001, 011, 100, 101, 111} to add to the categorical stimulus, they chose object {100} presumably because it shares the value "0" on the third stimulus dimension with the other three objects. Interestingly, for types $3_2[3]$ – II and $3_2[3]$ – III we observed participants choosing to add the object most dissimilar to the category prototype, which were the object(s) predicted by the GRIT-NPE. In total, the GRIT-NPE outperformed the MPMP-2 in terms of predicting the most-often chosen additions across the three category types, $X^2(1, N = 1{,}525) = 1{,}239.8$, $p < .0001$.

Category Deconstruction

General

Analyses were conducted on 1,845 choices. As above, although there was the potential for an analysis of 1,872 choices (39 participants × 48 trials), there were 27 no responses. In total, there were 620 removal decisions to category type $3_2[5]$ – I, 616 removal decisions to category type $3_2[5]$ – II, and 609 removal decisions to category type $3_2[5]$ – III. The raw aggregate data for each category type can be found in the "Total" rows of Table 2.

Chi-Square Goodness-of-Fit tests on the observed choices to remove from the three category types were conducted and were highly significant, least favorable $X^2(4, N = 609) = 162.7$, $p < .0001$. Thus, participants were not selecting each of the five objects equally to remove from the 5-object category types. Post hoc tests with Bonferroni-Holm corrections revealed O_3 of each of the $3_2[5]$ category types being removed more often than the other four clocks from instances of these category types, least favorable $X^2(1, N = 360) = 42.7$, $p < .0001$. Further, O_5 was chosen more often than O_2 for type $3_2[3]$ – II, $X^2(1, N = 168) = 11.5$, $p < .0001$. Lastly, O_2 and O_4 were selected more often than O_1 for type $3_2[3]$ – III, least favorable $X^2(1, N = 159) = 9.57$, $p < .005$.

Model Comparison

Regarding this task, the GRIT-NPE predicts that the most likely object to be removed from the five-object category types is the object that results in the most negative rate of change in subjective concept learning difficulty if that object would be removed from the category (i.e., −.21, .23, and 1 for the three structure types, respectively). Inspecting Table 2 we find that the GRIT-NPE measure predicts the most-often chosen object for each of these three category types. Spearman correlations of −.67, −.53, and −.45 indicate the predicted negative ordinal relationship within the three category types, respectively.

The MPMP-2 predicts that the most likely object to be removed is the object that is least similar to the category prototype, with lower values indicating lower similarity (i.e., .14, .14, and .05 for the three category types, respectively). The nonparametric MPMP-2 fails to account for the most-often chosen objects. Additionally, Spearman correlations of −.22, .53, and .45 indicate both negative and positive ordinal relationships between MPMP-2 predictions and observed choice frequencies.

Author's personal copy (e-offprint)

Table 3. Resulting $3_2[4]$ structural manifolds for the construction and deconstruction tasks

	$3_2[4]$ structural manifolds					
Construction Task	(1, 1, 0)	(1, 0, 0)	(0, 0, 0)	(1/2, 1/2, 0)	(1/2, 1/2, 1/2)	(1/2, 1/2, 1/2)
$3_2[3]$ – I	O_3			O_4	O_1, O_5	O_2
$3_2[3]$ – II		O_3		O_2, O_5	O_1, O_4	
$3_2[3]$ – III			O_3	O_1, O_4, O_5		O_2
N from Table 1	545	460	445	313	68	33
$3_2[4]$ – Type	I	II	VI	V	III	IV
Deconstruction Task	(1, 1, 0)	(1, 0, 0)	(0, 0, 0)	(1/2, 1/2, 0)	(1/2, 1/2, 1/2)	(1/2, 1/2, 1/2)
$3_2[5]$ – I	O_3			O_4	O_1, O_5	O_2
$3_2[5]$ – II		O_3		O_2, O_5	O_1, O_4	
$3_2[5]$ – III			O_3	O_1, O_4, O_5		O_2
N from Table 2	483	284	242	475	221	140
$3_2[4]$ – Type	I	II	VI	V	III	IV
Total *N*	1,028	744	687	788	289	173
N Possible	1,242	1,238	1,229	3,709	2,480	2,471
Proportion	.83	.60	.56	.21	.12	.07

Notes. The "$3_2[4]$ Structural Manifolds" were created by participants once they either added an object to the categorical stimuli presented in the Construction Task, or if they removed an object from the categorical stimuli presented in the Deconstruction Task.

Similar to the construction results above we observed participants removing the object that did not share the same dimensional value as the other four objects in the category for the $3_2[5]$ – I category type, predicted by the GRIT-NPE. We observed participants choosing to remove the object that may be construed as the category prototype for the remaining two of the three category types (i.e., $3_2[5]$ – II and $3_2[5]$ – III) – exactly the object(s) predicted by the GRIT-NPE. As with the construction task, the GRIT-NPE outperformed the MPMP-2 in predicting the observed choices across the three category types, $X^2(1, N = 1{,}231) = 378.5$, $p < .0001$.

A $3_2[4]$ Category Type Ordering

Upon utilizing a successful categorization model (GISTM-NPE; Vigo, 2013b, 2014) within an established information measure (GRIT-NPE; Vigo, 2013a, 2014), we found support for the hypothesis that individuals seek to minimize the concept learning difficulty of the resulting category. In line with this hypothesis, four of the six most-often created category types in the current study have been empirically associated with lower proportions of error rates in classification tasks (Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994; Shepard et al., 1961; Vigo, 2013b, 2014). More specifically, the family of six category types consisting of four objects defined by three binary dimensions ($3_2[4]$) has a robust empirical classification difficulty ordering in terms of average proportion of incorrect responses (i.e., I < II < [III, IV, V] < VI). As can be seen in Table 3, in the current tasks we observed participants creating type I when adding an object to category type $3_2[3]$ – I and removing an object from $3_2[5]$ – I, and creating type II when adding an object to category type $3_2[3]$ – II and removing an object from $3_2[5]$ – II. Importantly, however, we observed participants creating the most difficult to learn $3_2[4]$ category type (i.e., Type VI) when presented with instances of the $3_2[3]$ – III and $3_2[5]$ – III category types.

To explain why participants were creating types I, II, and VI, we apply the *invariance-parsimony* principle from GIST. This principle states that the human conceptual system favors perfectly redundant and perfectly diagnostic stimulus dimensions (e.g., partial invariance/SK values of 1 or 0, respectively) because the perfectly redundant dimensions (i.e., 1) may be eliminated without loss of information and the perfectly diagnostic dimensions (i.e., 0) are the most precise in determining membership. Considering this principle and the current results, we conducted equality of proportion post hoc tests using the data provided in Table 3 and discovered that categories with corresponding structural manifolds of (1, 1, 0), (1, 0, 1), and (0, 1, 1) – associated with category type $3_2[4]$ – I – were created most often, least favorable $X^2(1, N = 1{,}772) = 155.1$, $p < .0001$.[7] These structural manifolds have a single highly diagnostic dimension and two highly redundant dimensions.

[7] Structural manifolds belonging to the same category type (e.g., $3_2[4]$ – I) may vary in terms of the order of their kernel values. This is dictated by changes in the structural role that each dimension plays in the categorical stimulus. For example, depending on which of the three dimensions (e.g., color, shape, size) is the most diagnostic dimension (expressed by an SK with zero value), instances belonging to category type $3_2[4]$ are characterized by one of the following three structural manifolds: (1, 1, 0), (1, 0, 1), or (0, 1, 1).

 © 2016 Hogrefe Publishing

Author's personal copy (e-offprint)

Table 4. Parameter estimation for the MPMP-2

Category type	Parameters[a]				Model fit	
	α_1	α_2	α_3	c	r	R^2
$3_2[3]$ – I	0 (0)	0 (0)	1 (1)	3 (3.47)	1 (1)	1 (1)
$3_2[3]$ – II	0 (0)	1 (1)	0 (0)	3 (0.19)	.55 (.55)	.30 (.30)
$3_2[3]$ – III	0 (0)	1 (1)	0 (0)	3 (0.35)	.62 (.62)	.39 (.39)
$3_2[5]$ – I	1 (0)	0 (0)	0 (0)	3 (3.41)	−1 (−1)	1 (1)
$3_2[5]$ – II	.01 (.01)	.49 (.49)	.50 (.50)	3 (3.21)	.99 (.99)	.98 (.98)
$3_2[5]$ – III	.35 (.34)	.32 (.12)	.34 (.54)	3 (33.3)	.89 (1)	.79 (.99)

Notes. The estimations were performed on the first category instance for each category type that appear in Tables 1 and 2.
[a]We utilized the Solver platform in Excel that utilizes a gradient descent algorithm when fitting possible combinations of parameter values. We performed two sets of computations: One set fixing the scaling parameter to 3 and estimating the three attention weights, and one set estimating the scaling parameter and the three attention weights (in parentheses).

Conceptually, upon the extraction of these partial invariances we know that by ignoring the two redundant dimensions and focusing attention on the one diagnostic dimension we are able to form a one-dimensional rule for this category type. Following, categories with corresponding structural manifolds of (1, 0, 0), (0, 1, 0), and (0, 0, 1) – associated with category type $3_2[4]$ – II – were created second most often, least favorable $X^2(1, N = 1{,}431) = 4.29$, $p = .038$. Utilizing the two diagnostic dimensions and ignoring the lone redundant dimensions results in a very useful "logical-or" type of rule; for example, *round and few tick marks or square and many tick marks* for the first category instance presented in Table 1. Next, categories with corresponding structural manifolds of (0, 0, 0) associated with type VI were created third most often, least favorable $X^2(1, N = 1{,}475) = 527.5$, $p < .0001$. Fourth, categories with corresponding structural manifolds of (1/2, 1/2, 0), (1/2, 0, 1/2), and (0, 1/2, 1/2) – associated with category type $3_2[4]$ – V – were chosen more often than those with corresponding structural manifolds of (1/2, 1/2, 1/2) associated with types III and IV, least favorable $X^2(1, N = 1{,}077) = 94.5$, $p < .0001$. Lastly, both type III and type IV categories have structural manifolds of (1/2, 1/2, 1/2), but instances belonging to type III were created more often than instances belonging to type IV, $X^2(1, N = 462) = 31.1$, $p < .0001$. Aggregating these findings results in the following ordering for the proportion of times each of the $3_2[4]$ category types was created (whenever creating that Type was a possibility) during an experimental trial: I (.83) > II (.60) > VI (.56) > V (.21) > III (.12) > IV (.07). Note that this ordering is accurately predicted by the invariance-parsimony principle.

Discussion

We implemented two novel constrained categorization tasks with the aim of predicting and explaining why individuals prefer selecting certain objects to add to or to remove from categorical stimuli. We observed participants using "one-dimensional" behavior on approximately 1/3 of experimental trials ($3_2[3]$ – I and $3_2[5]$ – I); however, for the remaining trials we did not observe one-dimensional behavior nor behavior that may be characterized by summed similarity or "family-resemblance" principles. The MPMP-2, which captures family-resemblance intuitions quantitatively, is unable to account for a single empirical result. Upon conducting scaling and attention parameter estimations (see Table 4), we find the parameterized MPMP-2 only accounts for the one-dimensional behavior on category types $3_2[3]$ – I and $3_2[5]$ – I. Additionally, obtained correlations for category types $3_2[5]$ – II and $3_2[5]$ – III are opposite in sign to those predicted by a family-resemblance theoretical account of the model. In general, participants were choosing to add the most dissimilar clocks in the construction task and were choosing to remove the most similar clocks in the deconstruction task.

Explaining Prior Unsupervised Results With GIST and GRIT

Utilizing GIST, GRIT, and the aforementioned invariance-parsimony principle, we can explain the ubiquitous one-dimensional sorting behavior in the traditional unconstrained sorting tasks by Medin et al. (1987) and Regehr and Brooks (1995). In general, for these tasks participants were shown a set of stimuli cards (e.g., cartoon-like bugs, personality traits) and were instructed to sort them into two groups in a way that seemed natural to them. Although sorting consistent with a family-resemblance principle was expected, participants overwhelmingly used only one stimulus dimension to partition the set of stimuli into groups. This one-dimensional bias is robust and is a common empirical result when utilizing sorting classification procedures (Imai & Garner, 1965; Medin et al., 1987; Milton & Wills, 2004; Regehr & Brooks, 1995).

Author's personal copy (e-offprint)

Table 5. Explaining the prevalence of one-dimensional sorts with the GISTM-NPE

	Experiment	Dimensional makeup of category	GISTM-NP	GISTM-NPE
Medin et al. (1987)	1, 2c, d and 3a, b	A = {0000, 1000, 0100, 0010, 0001}	**3.63**	3.63
		B = {1111, 0111, 1011, 1101, 1110}	**3.63**	3.63
		A′ = {0000, 0111, 0100, 0010, 0001}	3.93	**3.15**
		B′ = {1111, 1000, 1011, 1101, 1110}	3.93	**3.15**
	2a,b	A = {000000, 100000, 010000, 001000, 000100, 000010, 000001}	**5.95**	5.95
		B = {111111, 011111, 101111, 110111, 111011, 111101, 111110}	**5.95**	5.95
		A′ = {000000, 011111, 010000, 001000, 000100, 000010, 000001}	6.11	**5.24**
		B′ = {111111, 100000, 101111, 110111, 111011, 111101, 111110}	6.11	**5.24**
Regehr and Brooks (1995)	1A	A = {0000000000, 1000000000, 0100000000, 0010000000, 0001000000, 0000100000, 0000010000, 0000001000, 0000000100, 0000000010, 0000000001}	**10.3**	10.3
		B = {1111111111, 0111111111, 1011111111, 1101111111, 1110111111, 1111011111, 1111101111, 1111110111, 1111111011, 1111111101, 1111111110}	**10.3**	10.3
		A′ = {0000000000, 0111111111, 0100000000, 0010000000, 0001000000, 0000100000, 0000010000, 0000001000, 0000000100, 0000000010, 0000000001}	10.36	**9.42**
		B′ = {1111111111, 1000000000, 1011111111, 1101111111, 1110111111, 1111011111, 1111101111, 1111110111, 1111111011, 1111111101, 1111111110}	10.36	**9.42**

Notes. Family-resemblance sorts are denoted A and B, whereas one-dimensional sorts are denoted A′ and B′. Sorts predicted to be observed based on calculations from the GISTM-NP and GISTM-NPE are bolded.

Table 6. Explaining family-resemblance sorts with the GRIT-NPE

		Category A	Category B
	Prototype	{0000}	{1111}
Match-to-standards (Regehr & Brooks, 1995)	Choice 1 {1000}[A]	{0000, 1000}	{1111}
	Choice 2 {1110}[B]	{0000, 1000}	{1111, 1110}
	Choice 3 {1011}[B]	{0000, 1000}	{1111, 1011}
	Choice 4 {0010}[A]	{0000, 0010}	{1111, 1011}
	Choice 5 {0100}[A]	{0000, 0100}	{1111, 1011}
	Choice 6 {1101}[B]	{0000, 0100}	{1111, 1101}
	Choice 7 {0001}[A]	{0000, 0001}	{1111, 1101}
	Choice 8 {0111}[B]	{0000, 0001}	{1111, 0111}
	Result	{0000, 1000, 0010, 0100, 0001}	{1111, 1110, 1011, 1101, 0111}

Note. The superscripts A and B in each "Choice" row represent the category to which the observer is predicted to place the corresponding object in curly brackets.

Now, upon inspecting the structural manifolds of the categories from these studies provided in Table 5, we find that the family-resemblance categories (e.g., A & B) contain no extreme SK values (i.e., (2/5, …, 2/5) & (2/7, …, 2/7) and (2/11, …, 2/11)), whereas the one-dimensional categories (e.g., A′ & B′) do contain precisely one extreme SK value (i.e., (0, 2/5, …, 2/5) & (0, 2/7, …, 2/7) and (0, 2/11, …, 2/11)). Crucially, because the one-dimensional categories have higher degrees of structural equilibrium (e.g., 1.25, 1.17, and 1.1) than the family-resemblance categories (e.g., 1, 1, and 1) it is easier to find the diagnostic dimension with 0 partial invariance, which reduces the

 © 2016 Hogrefe Publishing

Author's personal copy (e-offprint)

concept learning difficulty of the one-dimensional categories. This reduction in concept learning difficulty can be seen when comparing the GISTM-NP and GISTM-NPE predictions provided in Table 5 for both the one-dimensional and family-resemblance categories and we believe this explanation of the one-dimensional sorting bias integrates well with the results of the current study.

Related, the family-resemblance sorting behavior observed with the match-to-standards and match-from-array ("single-standard-available" condition) procedures may be explained within the GRIT-NPE framework by analyzing behavior at each step of the task. As demonstrated in Table 6, by utilizing the GRIT-NPE to predict the stepwise choices we find that subjects are predicted to form categories consistent with a family-resemblance abstract similarity structure. Applying this same analysis to the "building-array" condition of the match-from-array procedure, it can be shown that the GRIT-NPE accurately predicts the prevalent one-dimensional sorting by subjects found by Regehr and Brooks (1995).

Conclusion and Future Research Directions

In summary, GIST and GRIT unify sorting behavior across tasks by assuming an observer's conceptual system extracts invariance patterns in categorical stimuli. These sub-symbolic patterns are hypothesized to be necessary precursors for symbolic processing – allowing observers to form linguistic rules or concept prototypes in unsupervised, constrained, and supervised categorization tasks employing multidimensional stimuli (Vigo, 2011b, 2013b, 2014). Notwithstanding, there are limitations of the current study that need to be addressed with future research.

First, the generality of the current findings needs validation. An initial step may be performed by testing categorical stimuli consisting of objects defined over multi-valued or continuous stimulus dimensions. Additional experimental modifications may involve multi-object addition and removal decisions on categorical stimuli. Similarity-based measures such as the MPMP-2 have difficulty predicting choices in such cases, whereas predictions obtained via the GRIT-NPE are straightforward. Finally, researchers may compare addition and removal decisions among differing developmental populations, such as children, adolescents, and adults. An investigation by Hayes and Taplin (1993) suggests that 6-year-old children may be constrained to utilize prototype-specific information when making classification decisions. Thus, empirical investigations with children may reveal their addition and removal decisions being accounted for more accurately by a prototype-based measure such as the MPMP-2. Note, however, that GIST has the explanatory framework to account for developmental departures in categorization behavior, as it has been successfully applied to account for categorization behavior in adults, adolescents, and adolescents diagnosed with ADHD (Vigo, Evans, & Owens, 2014).

A second limitation of the current study revolves around the attempt to connect the current research with prior research on unsupervised categorization processes. Specifically, we were deliberately vague with components of the instructions. Despite this, there was clear agreement across participants for the most-often chosen object stimuli in each category type, the individual-level analysis mirrored the group-level analysis for each task, and the clocks participants were choosing to add in the construction task were the same clocks participants were choosing to remove in the deconstruction task. Additionally, participants displayed consistency in their choices for both the construction (69%) and deconstruction (55%) tasks when compared to chance (20%). For these reasons we believe there was agreement among participants regarding how to interpret and complete the tasks. Nonetheless, we suggest future studies implement experimental conditions with varied sets of instructions, isolated presentation of category types, and more informative debriefing protocols. Indeed, presenting instances of only one category type to participants may reveal the use of "exclusive-or" or other multidimensional strategies for particular experimental trials.

Lastly, multiple studies suggest that manipulating stimulus factors (e.g., perceptual/spatial integration; Milton & Wills, 2004, 2008) or the procedure (e.g., time pressure; Milton, Viika, Henderson, & Wills, 2011) may alter sorting strategies adopted by participants in constrained-categorization tasks. Opportunely, there are free parameters in the parametric variants of the GIST-NPE and the GRIT information measure that may be estimated to elucidate what may be occurring with these manipulations. In conclusion, we hope that our account of the results from our experiments using a recent and successful theory of categorization behavior spurs further research on the nature of the cognitive mechanisms underlying unsupervised and constrained classification tasks.

Electronic Supplementary Material

The electronic supplementary material is available with the online version of the article at http://dx.doi.org/10.1027/1618-3169/a000337

ESM 1. Raw data (Excel).
Raw data of the study.

References

Ahn, W., & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science, 16*, 81–121. doi: 10.1207/s15516709cog1601_3

Author's personal copy (e-offprint)

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology, 18*, 500–549. doi: 10.1016/0010-0285(86)90008-3

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature, 407*, 630–633. doi: 10.1038/35036586

Feldman, J. (2003). A catalog of Boolean concepts. *Journal of Mathematical Psychology, 47*, 75–89. doi: 10.1016/S0022-2496(02)00025-1

Gati, I., & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology, 16*, 341–370. doi: 10.1016/0010-0285(84)90013-6

Goodwin, G. P., & Johnson-Laird, P. N. (2013). The acquisition of Boolean concepts. *Trends in Cognitive Sciences, 17*, 128–133. doi: 10.1016/j.tics.2013.01.007

Hayes, B. K., & Taplin, J. E. (1993). Developmental differences in the use of prototype and exemplar-specific information. *Journal of Experimental Child Psychology, 55*, 329–352. doi: 10.1006/jecp.1993.1019

Higonnet, R. A., & Grea, R. A. (1958). *Logical design of electrical circuits*. New York, NY: McGraw-Hill.

Imai, S., & Garner, W. R. (1965). Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology, 69*, 596–608. doi: 10.1037/h0021980

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254–278. doi: 10.1037/0033-295X.100.2.254

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19*, 242–279. doi: 10.1016/0010-0285(87)90012-0

Milton, F., Viika, L., Henderson, H., & Wills, A. J. (2011). The effect of time pressure and the spatial integration of the stimulus dimensions on overall similarity categorization. In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 795–800). Austin, TX: Cognitive Science Society.

Milton, F., & Wills, A. J. (2004). The influence of stimulus properties on category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 407–415. doi: 10.1037/0278-7393.30.2.407

Milton, F., & Wills, A. J. (2008). The influence of perceptual difficulty on family resemblance sorting. In B. C. Love, K. McRae, & M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 2273–2278). Austin, TX: Cognitive Science Society.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 104–114. doi: 10.1037/0278-7393.10.1.104

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57. doi: 10.1037/0096-3445.115.1.39

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. G. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory and Cognition, 22*, 352–369. doi: 10.3758/BF03200862

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 924–940. doi: 10.1037/0278-7393.28.5.924

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26*, 303–343. doi: 10.1207/s15516709cog2603_6

Pothos, E. M., & Chater, N. (2005). Unsupervised categorization and category learning. *The Quarterly Journal of Experimental Psychology, 58A*, 733–752. doi: 10.1080/02724980443000322

Regehr, G., & Brooks, L. R. (1995). Category organization in free classification: The organizing effect of an array of stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 347–363. doi: 10.1037/0278-7393.21.2.347

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573–605. doi: 10.1016/0010-0285(75)90024-9

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied, 75*, 1–42. doi: 10.1037/h0093825

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352. doi: 10.1037/0033-295X.84.4.327

Vigo, R. (2006). A note on the complexity of Boolean concepts. *Journal of Mathematical Psychology, 50*, 501–510. doi: 10.1016/j.jmp.2006.05.007

Vigo, R. (2009). Categorical invariance and structural complexity in human concept learning. *Journal of Mathematical Psychology, 53*, 203–221. doi: 10.1016/j.jmp.2009.04.009

Vigo, R. (2011a). Towards a law of invariance in human conceptual behavior. 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. doi: 10.1016/j.ins.2011.05.020

Vigo, R. (2013a). Complexity over uncertainty in generalized representational information theory (GRIT): A structure-sensitive general theory of information. *Information, 4*, 1–30. doi: 10.3390/info4010001

Vigo, R. (2013b). The gist of concepts. *Cognition, 129*, 138–162. doi: 10.1016/j.cognition.2013.05.008

Vigo, R. (2014). *Mathematical principles of human conceptual behavior: The structural nature of conceptual representation and processing*. New York, NY: Routledge.

Vigo, R., & Basawaraj, B. (2013). Will the most informative object stand? Determining the impact of structural context on informativeness judgments. *Journal of Cognitive Psychology, 25*, 248–266. doi: 10.1080/20445911.2012.755510

Vigo, R., & Doan, C. A. (2015). The structure of choice. *Cognitive Systems Research, 36–37*, 1–14. doi: 10.1016/j.cogsys.2015.02.001

Vigo, R., Evans, S. W., & Owens, J. S. (2014). Categorization behavior in adults, adolescents and ADHD-adolescents: A comparative investigation. *The Quarterly Journal of Experimental Psychology, 68*, 1058–1072. doi: 10.1080/17470218.2014.974625

Vigo, R., Zeigler, D. E., & Halsey, P. A. (2013). Gaze and informativeness during category learning: Evidence for an inverse relation. *Visual Cognition, 21*, 446–476. doi: 10.1080/13506285.2013.800931

Received March 29, 2016
Revision received September 11, 2016
Accepted September 14, 2016
Published online November 11, 2016

Charles A. Doan
Consortium for the Advancement of Cognitive Science
Department of Psychology
200 Porter Hall
Ohio University
Athens, OH 45701
USA
cd634011@ohio.edu

 © 2016 Hogrefe Publishing