

Family Resemblance in Unsupervised Categorization: A Dissociation Between Production and Evaluation

John. D. Patterson (jpatter4@binghamton.edu)
Sean Snoddy (ssnoddy1@binghamton.edu)
Kenneth J. Kurtz (kkurtz@binghamton.edu)

Department of Psychology, Binghamton University (SUNY)
Binghamton, NY 13902 USA

Abstract

A plurality of the categories we hold exhibit family resemblance (FR; i.e., many characteristic but few defining features), suggesting FR may occupy a central role in human category formation. However, research in unsupervised learning has shown that when people are asked to sort an array of novel items into categories, they ubiquitously use a unidimensional (UNI) rule – despite the availability of a FR solution. This work suggests that, perhaps, FR similarity is *not* a core tendency in category formation. Here, we question whether the UNI bias is a result of the sorting paradigm. Specifically, we speculate the paradigm conflates two components vital for category formation: production and evaluation. Across three experiments we show that when evaluation is separated from generation – by using a novel forced-choice task that pits different category organizational schemes against one another – people exhibit a FR over UNI preference. The implications of these results are discussed.

Keywords: unsupervised categorization; similarity; family resemblance; unidimensional bias; category construction

Introduction

Understanding the cognitive basis on which we create novel categories in the absence of feedback is foundational to understanding human category learning more broadly. One way to address this question is simply by studying the categories we already hold. Theoretical and behavioral work has shown that the natural categories are described by a family resemblance (FR), or overall similarity, structure wherein members of a category share many characteristic features but share few or no defining features (Rosch & Mervis, 1975, Wittgenstein, 1953). Given the prevalence of FR among natural categories, an intuitive hypothesis is that overall similarity is the preferred or default basis on which we form novel categories.

Under this hypothesis, Medin, Wattenmaker, and Hampson (1987) sought to investigate unsupervised category formation more directly by using a sorting paradigm in which the participant was given an array of novel, multi-dimensional stimuli and asked to sort them into two equal-size categories. Critically, the examples could be sorted based on FR or, alternatively, based on a unidimensional (UNI) rule (e.g., ‘red things in one category, blue things in another’). Contrary to their expectations, Medin et al. (1987) demonstrated across several experiments that people overwhelmingly preferred to construct categories described by a UNI rule; they *very*

seldom created categories adherent to FR, despite UNI solutions having less within-category, and more between-category, similarities than those based on FR (Medin et al., 1987). Much work has subsequently replicated the strong UNI bias under the full-array sorting task (e.g., Ahn & Medin, 1992; Lassaline & Murphy, 1999; Regehr & Brooks, 1995; Wattenmaker, 1992).

The inconsistency between the UNI bias in the full-array sorting task and the tendency for natural categories to be described by FR has puzzled the field, and much research has been devoted to understanding why such an inconsistency exists. Coarsely, this work has two central themes: feature and task effects. Research on feature effects has shown that, generally speaking, changing the quality or the number of features is ineffective at reversing the UNI bias – and in some cases can exacerbate it (Regehr & Brooks, 1995). This of course comes with one notable exception: prior knowledge. People produce more FR sorting when the features of the FR categories map onto known concepts (e.g., the features of extro- vs. introversion), or novel concepts taught to participants, that explain and relate the features to one another (e.g., Ahn, 1999; Medin et al., 1987). This work is important for understanding category formation; however, we consider it to address a fundamentally different question. Instead of asking, “what is the organizational basis used when forming a completely novel/artificial category,” it asks, “given a latent or manifest conceptual/causal basis in the features, do people construct categories by it?”

Research on task effects has shown two critical findings. First, Lassaline and Murphy (1996) showed that an inference task (e.g., ‘if it has feature A on dimension 1, what feature is it likely to have on dimension 2’) prior to the sorting task increased FR responding. This experiment shows that encoding feature co-prediction is vital for generating FR. However, it leaves open the question of whether people do this kind of feature encoding spontaneously during category formation.

Second, Regehr and Brooks (1995) used a novel Match-to-Standards (MTS) task in which participants sorted each item, one at a time, by matching them to one example item from each FR category; each sorted example covered up the previous example that was sorted into that category and the standards remained visible throughout the task. The MTS task led to much greater FR responding, relative to the full-array sorting task. This is suggestive that FR structure in

natural categories might be an emergent property of many item-item matches, though this connection has never been empirically drawn. Several subsequent studies have followed up on this approach to studying unsupervised categorization, examining factors such as feature separability, time pressure, and working memory load (Milton, Longmore, & Wills, 2008; Milton & Wills, 2004; Wills, Milton, Longmore, Hester, & Robinson, 2013).

However, we have two concerns with the MTS task, as the data currently stand. One, it is based on local item-to-item matching in which only the most recent item sorted into each category is visible (along with the standards); thus, while people produced more FR responding, it is unclear if that responding is attributable to the participant's appreciation of similarity structure generated across examples in each category or if instead FR responding was simply a product of making local matches (without appreciating category-level structure). Two, the task does not measure unsupervised category formation. Given the supervision, in the form of standards from each category, the task is instead a measure of semi-supervised category formation (Patterson & Kurtz, 2018; Vong, Navarro, & Perfors, 2016).

The goal of the current work was to investigate task effects in unsupervised category formation from a novel perspective that aims to address some of the limitations of previous research on task effects. We ask whether UNI similarity is indeed a deep organizational preference in category formation or if, instead, it is a direct product of the standard full-array sorting paradigm (e.g., Medin et al., 1987). We identify three aspects of the sorting task that independently or in conjunction could encourage UNI solutions. First, the task presents a whole array of multi-dimensional stimuli to the participant simultaneously. Intuitively, complexity in both the number of items and number of features might encourage problem simplification in the form of sample or dimensionality reduction. As sample reduction is not an option (participants must include all items in the solution), dimensionality reduction may be utilized.

Second, the goal of the sorting task is to produce two categories; this goal is decidedly intentional and discriminative – i.e., goal is to predict/separate class. Research has shown that intentional and discriminative learning leads to greater rule focus relative to either incidental learning or learning where class-prediction is softened (Levering & Kurtz, 2015; Love, 2002). Third, and critically, the standard sorting task conflates two intuitively essential components for category formation: the *generation* of a candidate category structure and the *evaluation* of that structure relative to possible alternatives. Given the goal of the task is to generate *one* candidate structure, the evaluation of this structure is likely to be inadequate due to insufficient alternative structures with which to compare it to. Furthermore, we expect candidate structures in naturalistic settings are generated not cold (as in the standard paradigm), and not *necessarily* with prior top-down

knowledge, but through feature statistics that accrue with incidental experience (Lassaline & Murphy, 1996; Love, 2002). As such, we expect the structure hypotheses generated by participants in the sorting paradigm to be immature.

In the experiments that follow, we introduce a novel Structure Choice Task (SCT) in which two candidate structures are presented side-by-side and the participant chooses which they prefer. The task thus obviates the need to generate structure hypotheses (which we believe are undermined in the standard sorting paradigm) and isolates structure evaluation. If FR is a preferred organizational principle in category formation (and if the UNI bias is a product of the standard sorting paradigm), we should expect people to choose the FR structure more frequently than the UNI one.

We point out that the SCT resolves limitations from previous task effects research in two ways. First, the task does not rely on any pre-task encoding manipulations; participants encode the items/structures however they wish and make a judgment. As such, the preferences produced are spontaneous. Second, the task does not restrict the number of stimuli that are under consideration, as in the MTS task. Because the SCT presents whole categories, organized in two different ways, it should reflect the participant's category-level similarity preference (rather than local matching).

In Experiment 1A, we pit FR against UNI in the SCT and provide first-ever evidence of a spontaneous FR preference in a category-level task that uses knowledge-poor features. In Experiment 1B, we replicate E1A and extend it by comparing SCT results to full-array sorts completed either before or after the SCT; despite replicating the FR preference in the SCT, effectively nobody produced FR sorts – suggesting the standard sorting paradigm encourages UNI solutions. In Experiment 2, we address potential critiques to FR supremacy in the SCT.

Experiment 1A

In a within-subjects design, FR and UNI organizations of the same items were pitted against one another in the SCT. Without having to generate hypotheses, we predicted participants would prefer FR organizations.

Method

Participants 108 undergraduate students at Binghamton University participated.

Materials and Design The stimuli were based on a five-dimensional variant of the abstract FR category structure from Medin et al. (1987); each binary dimension is represented as a pair features (see Figure 1). Each category consisted of a prototype – containing all five characteristic features of the category – and five 'one-off' items that differed from the prototype by a feature that was consistent with the opposite prototype. Five 12-item stimulus sets, from distinct domains (see Figure 1 for prototypes), were

created from this abstract structure.

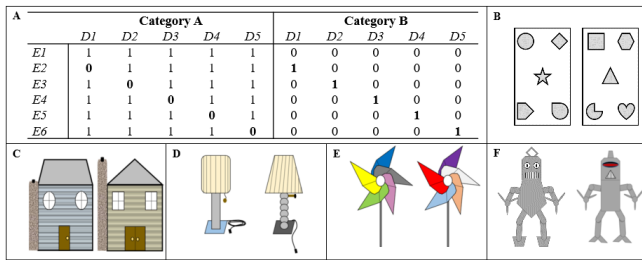


Figure 1: (A) The abstract FR structure [prototypes: row 1]; (B-F) Prototypes from all five stimulus sets: and prototypes of each category structure for all five stimulus sets, respectively: Card, House, Lamp, Pinwheel, Robot.

On each trial of the SCT, two completed sorts from the same stimulus set were presented side-by-side on the computer screen, one a FR solution and the other a UNI solution (see Figure 2 for an example trial). The UNI solution was generated on each trial of the experiment by swapping one-offs of a randomly selected row in the abstract structure, between category. Stimulus sets were presented in a randomized order. The order of items in each category of each completed sort, as well as the side on which FR was shown (left vs. right), were also randomized.

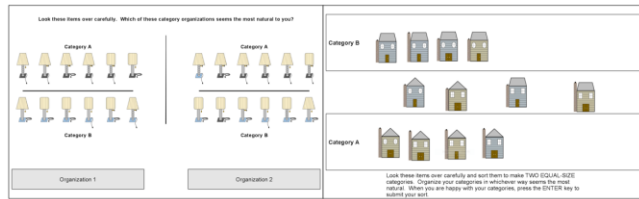


Figure 2. The left panel contains an example trial of the SCT containing a FR (right) and UNI organization (left; base color). The right panel contains an in-progress example of the sort task.

Procedure The SCT was programmed and administered via computer. Immediately prior to the task, participants were given instructions: “In this study you will be presented with a single set of items that is organized in two different ways. An example trial is pictured below. On each trial, carefully look at both organizations and select whichever one seems the most natural to you.” The example trial shown with the instructions was identical to Figure 2, except the lamps were replaced with naturalistic ducks and bats – one organization had ducks and bats separated as categories A and B, while the other had them intermixed between categories. After the instructions, participants sequentially performed the SCT on each of the five domains. Participants selected their preference by clicking a button located below the organizations with the mouse. After responding, participants proceeded to the next trial/stimulus set.

Results & Discussion

We obtained a difference score for each participant that reflected their net preference (total FR selections minus total

UNI selections). As there were five trials, difference scores could range from -5 (all UNI selections) to 5 (all FR selections). The analysis showed that FR organizations were selected reliably more frequently than UNI ones (the ordinal, non-normal data were subjected to a Wilcoxon signed-rank test: $Mdn = 1, Z = 3.787, p < .001$; see Figure 3). Supplemental analyses¹ provide a histogram of difference scores (pp. 1) and an analysis of preference by stimulus domain (pp. 2-3).

These results show clearly that when people’s structural preferences are assayed – in the absence of being asked to produce candidate structures – they prefer FR. This provides inaugural evidence that FR may be a deep organizational basis that is sought in naïve, unsupervised category formation. Compared to the UNI bias typically seen in the full-array sorting paradigm, these findings represent a massive divergence – suggesting the affinity for UNI organizations in the sorting task is related to shortcomings in the generation of candidate structures. However, these stimuli have never been subjected to a direct comparison between the SCT and the full-array sorting task. It is possible that the stimuli we used are somehow generally more prone to be categorized by FR. This potential critique is addressed in Experiment 1B.

Experiment 1B

In this experiment, we sought to replicate the FR advantage and relate the outcomes of the SCT and the full-array sorting task using the same stimulus sets. To this end, each participant completed both the SCT and the full-array sorting task for all the stimulus sets in a task-blocked format (e.g., SCT for all sets, then sorting task for all sets). The order of the tasks was balanced across participants. We predicted participants would display a profound UNI bias in the sorting task, but that people would readily choose FR in the SCT.

Method

Participants 140 undergraduate students at Binghamton University participated. Participants were randomly assigned to the SCT first (SCT-SORT; $N = 71$) or the sort first (SORT-SCT; $N = 69$) condition.

Materials and Design The materials were identical to Experiment 1A. The same stimuli were used in both the SCT and sort task.

Procedure Both tasks were administered through a computer program. The SCT procedure was identical to that of E1A. In the SORT task, participants were instructed that there were many ways to create two equal-size categories, but their goal was to sort them in a way they thought most natural; an example sort, using the same demo images used

¹ The supplemental analyses can be viewed at this link: https://osf.io/jr2wu/?view_only=de559c73f1ef4b4da3781f4ef680f74f

for the SCT demo, was provided with the instructions. After instructions, participants then sorted each stimulus set in random order. For each stimulus set, the stimuli initialized to a row in the middle of the screen in a random order. Above and below that row were the two category bins. Participants used the mouse to drag and drop items into either category bin, and were able to reclassify the items freely. When finished, participants hit the enter key to submit. Sorts were coded as FR, UNI, or OTHER. OTHER sorts reflect any type of category produced by participants that is not FR or UNI; these sorts lack any interpretable structure. In the SORT-SCT group, participants sorted each of the randomly ordered stimulus sets before then completing the SCT for each set. The SCT-SORT group was the same, but the order of the two tasks was swapped.

Results & Discussion

The primary goal of the experiment was to replicate the FR advantage in the SCT; we use the same SCT analysis as in E1A. The SCT difference scores did not differ as a function of condition (SORT-SCT: $Mdn = 1$; SCT-SORT: $Mdn = 1$; $Z = -0.41$, $p = .682$). As such, SCT data from the two conditions were combined. Our analysis using the magnitude of the difference scores showed that participants selected the FR structure reliably more frequently than the UNI one (Wilcoxon signed-rank: $Mdn = 1$, $Z = 2.98$, $p = .003$; see Figure 3). Thus, we replicated the effect found in E1A and illustrate that when the task is constrained to the evaluation of candidate structures, FR is preferred over UNI solutions (see supplemental analyses pp. 2-3 for an analysis of preference by stimulus domain).

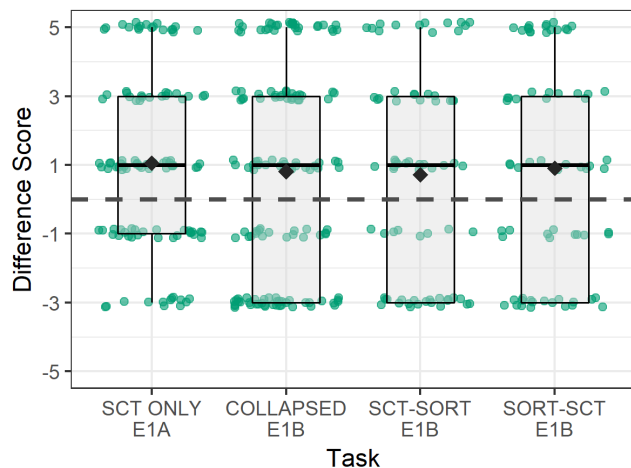


Figure 3. SCT preferences in Experiments 1A (left) and 1B (right three). COLLAPSED reflects an aggregation of both SCT-SORT and SORT-SCT conditions. Green dots reflect participant difference scores. Diamonds show means. Dashed line shows a difference score of 0 (FR = UNI).

The second goal of Experiment 1B was to compare the SCT to the standard full-array sorting paradigm, using the same stimuli as in the SCT. A potential concern from E1A

is that the preponderance of FR responding in the SCT is a result of the stimuli, rather than a result of separating evaluation from production. As with the SCT data, we collapsed sort task data across conditions; the conditions did not reliably differ in the number of UNI solutions provided (SORT-SCT: 99%, SCT-SORT: 97%; $\chi^2(1, N = 593) = 0.002$, $p = .967$), and there were too few alternative solutions generated (FR/OTHER) to compare across order conditions. For the collapsed data, we replicated the prevalent UNI bias; participants produced significantly more UNI sorts (98%) than FR (0.2%) and OTHER sorts (1.8%) [UNI vs. FR: $\chi^2(1, N = 594) = 590.1$, $p < .001$; UNI vs. OTHER: $\chi^2(1, N = 604) = 560.8$, $p < .001$]. Moreover, more OTHER sorts were produced than FR sorts, reflecting the rarity of FR solutions produced in the sorting paradigm [$\chi^2(1, N = 12) = 8.333$, $p = .004$].

Overall, these findings show a successful replication of the FR preference under the SCT. By contrast – and consistent with previous research using the array sorting task – the sorting task led to an overwhelming number of UNI sorts. This highlights two points. First, there were only 12 non-UNI sorts produced; this means the very same people who produced consistent UNI sorts found FR to be more compelling than UNI in the SCT, under the same stimuli. This strongly suggests that the UNI bias in the sorting task arises not because participants evaluate UNI as a superior organizational principle (which the SCT data shows), but instead because some element(s) of the sorting task encourages it. Second, regarding the concern that our stimuli might generally be more prone to FR preference, the strong UNI bias in the full-array sorting task indicates that is not the case.

Experiment 2

The purpose of this experiment is to address an alternative account of the FR advantage observed and replicated in E1. Specifically, we were concerned that participants might have failed to notice the UNI rule and, in the (perceived) absence of a UNI rule, chose FR as a best of bad options (perhaps without appreciating the FR similarity).

In this experiment, we use a three-condition (within-subjects) version of the SCT. The first condition is the same FR vs. UNI choice task as in the previous experiments. However, we introduce two new conditions: UNI vs. OTHER and FR vs. OTHER. Note that the UNI vs. OTHER condition should make UNI a compelling option – provided people do notice the UNI rule in the SCT; a UNI over OTHER preference would thus suggest participants in the previous experiments did notice the UNI rule, but preferred FR. A potential concern about this design is that, by elevating UNI to an ‘optimal’ choice on those trials, it might invite a demand characteristic (e.g., ‘UNI is better option here, maybe they want me to find/select the UNI option elsewhere’) or make UNI organizations more appealing than they might otherwise be. The FR vs. OTHER condition was included to minimize these effects, as it serves to balance the number of times FR and UNI were ‘optimal’ vs.

OTHER groupings. In addition, it allows us to examine if FR is preferred over OTHER organizations and affords us a global measure of preference by making each response type equally probable due to chance (total FR vs. UNI, across all trial types).

Method

Participants 363 undergraduate students from Binghamton University participated.

Materials and Design The materials were the same as in Experiment 1. The design was like E1A, but expanded to include UNI vs. OTHER and FR vs. OTHER trial types, all within-subjects. OTHER organizations were created by taking a FR organization and swapping three randomly-chosen, non-prototype items from one category with three items occupying the same rows in the opposite category, according to the abstract structure shown in Figure 1; these arrangements had no discernible structure. In sum, there were three trial types: 1) FR-UNI, 2) FR-OTHER, and 3) UNI-OTHER. The particular OTHER and UNI groupings that were created for each subject were held constant across trial types within a domain (e.g., if the UNI rule was on the base color for the Lamp set [see Figure 2] in the FR-UNI trial, base color would also be used to form the UNI rule in the UNI-OTHER trial for Lamps). This was done to ensure a consistent comparison across trial types in a stimulus set.

Procedure The procedure was like Experiment 1A. Participants were presented with each stimulus set sequentially, in a random order. For each stimulus set, the participant was presented with each trial type sequentially, in a random order. In each trial type, the participant selected which of the two organizations they preferred. Unlike in the previous experiments, the participants were elicited for an explanation of their choice for each trial type in the first and last stimulus sets.

Results & Discussion

Separate difference scores were calculated for each trial type: UNI minus OTHER, FR minus OTHER, and FR minus UNI. Wilcoxon signed-rank tests were conducted on the magnitudes of the difference scores. Supplemental analyses include a histogram of difference scores (pp. 4) an analysis of preference by stimulus domain (pp. 5-6) and an analysis of FR-UNI trials based on the previous trial (pp. 7).

Our primary concern in this experiment was to determine whether people do in fact detect UNI rules in the SCT. The critical trial type for assessing this was UNI-OTHER; a UNI preference would suggest participants do detect the UNI rule. The analysis of the UNI-OTHER condition yielded a reliable UNI over OTHER preference (Wilcoxon signed-rank: $Mdn = 3, Z = 12.487, p < .001$). Importantly, the UNI preference provides strong evidence that people do in fact notice UNI rule embedded in UNI organizations and suggests that the FR over UNI preferences observed in E1A

and E1B are not derived from participants simply failing to notice the rule.

The same pattern was observed in the FR-OTHER condition; participants selected significantly more FR organizations than OTHER organizations (Wilcoxon signed-rank: $Mdn = 3, Z = 13.195, p < .001$). As a complement to the UNI-OTHER analysis, this finding shows that people are sensitive to FR as a coherent organizational basis and find it compelling relative to less coherent options.

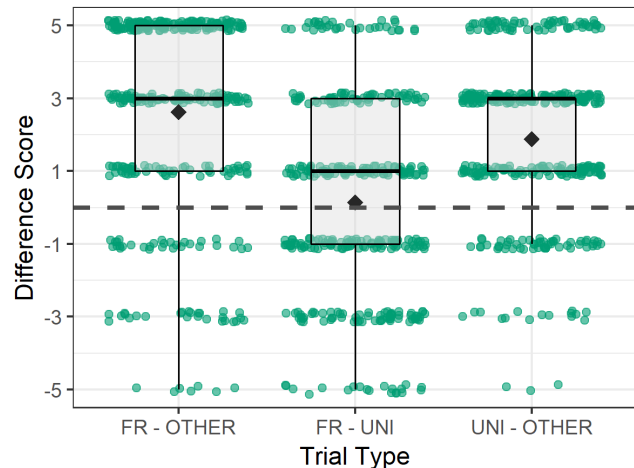


Figure 4. SCT preferences by trial type in Experiment 2. Green dots show each participant's difference score. Diamonds show means. Dashed line shows a difference score of 0. Positive scores in the FR-OTHER and FR-UNI trials reflect a FR preference; positive scores reflect a UNI preference in the UNI-OTHER trials.

Looking to the FR-UNI condition – a replication plus extension (given the novel trial types) – we failed to find the FR preference observed in E1A and E1B. The difference score magnitude did not differ from zero (Wilcoxon signed-rank: $Mdn = 1, Z = 0.851, p = .395$; see Figure 4). The failure to replicate is curious. One possibility is that the FR preference observed across E1A and E1B is a Type 1 error. However, both of those experiments were well-powered and the effect was replicated. Another possibility is that by introducing: (1) the new trial types; and/or, (2) the verbal explanations for preferences on the first (and last) stimulus set altered participants' behavior in the task. The use of OTHER organizations as a comparator effectively set up both UNI and FR organizations as 'correct' answers to the task. Then, on FR-UNI trials, the participant must decide which 'correct' answer is 'more correct'. Given UNI rules lend themselves more easily to verbal description (e.g., Zeithamova & Maddox, 2006), and given we asked participants for verbal descriptions, participants may have surmised that UNI was the 'more correct' choice and chose it more frequently than in the previous experiments. Upcoming studies will seek to disentangle these possibilities.

Lastly, we consider the global preference measure (all FR minus all UNI, across all trial types within subject). Consistent with the FR preference observed in the previous

two experiments, we found that people chose FR reliably more frequently than UNI organizations (Wilcoxon signed-rank: $Mdn = 1$, $Z = 2.543$, $p = .011$). This suggests that, despite not showing a preference for FR over UNI organizations on FR-UNI trials, participants did demonstrate an overall preference for FR when collapsing across all trial types.

In sum, we provided evidence that the FR over UNI preference observed across E1A and E1B is not attributable to people failing to notice the rule. Moreover, we provided additional evidence that people are sensitive to FR. Although the FR over UNI advantage did not replicate (potentially due to manipulations introduced in the current experiment), the FR over OTHER advantage indicates that people view FR as a meaningful organizing principle for categories.

General Discussion

The widespread UNI bias in unsupervised category formation – and its inconsistency with the FR structure of natural categories – has remained a question mark in the field for decades. In the experiments above, we approach the question from the perspective that the standard sorting paradigm encourages UNI responding by virtue of being an intentional, production-focused task that does not afford the requisite incidental exposure for learners to generate candidate structures as they might otherwise in naturalistic settings. To circumvent these issues, we introduced the SCT – a task that requires only the evaluation of provided candidate structures rather than both generation and evaluation.

Across two high-powered experiments, we observed a FR over UNI preference in the SCT – contesting the prevalent UNI bias under the sorting paradigm. This preference emerged even despite the use of knowledge-poor features (Medin et al., 1987), a full-array format (Regehr & Brooks, 1995), and despite the omission of a pre-task encoding phase (Lassaline & Murphy, 1996). Experiment 1B showed that the FR preference observed in the SCT is not due to the stimuli being generally FR-prone, as evidenced by the sorting task, and showed that people sort according to UNI rules, regardless of their SCT preference. In an extended form of the SCT, we showed in Experiment 2 that the FR preferences observed in the preceding experiments did not arise from a failure to identify the UNI rule in UNI organizations and provided further evidence that people view FR as a meaningful way of organizing categories.

These results are compelling for a number of reasons. First, they suggest that, at one extreme, people prefer FR over UNI structures in category formation (E1A & E1B). At the other extreme, they suggest that people do not have a preference between UNI and FR structures (E2). Regardless of which is true, the data show that people are sensitive to and appreciative of both types of category-level similarity, and this represents a massive departure from the strict UNI bias observed in – to our awareness – every unsupervised category learning study with domain-naïve participants and

no additional encoding tasks (though see Pothos & Close, 2008; Pothos et al., 2011 which address multidimensional vs. UNI sorting – though not the specific tension between FR and UNI). As such, these findings present a potential alignment with the basis of similarity that apparently underlies natural categories (Rosch & Mervis, 1975).

Second, these results highlight the importance of task in the behaviors that are produced, afforded, and encouraged. Although we found evidence that people appreciate FR, in E1B we showed that people – many of which displayed a FR preference in the SCT – uniformly produced UNI sort solutions. Thus, by isolating a sub-component of the overall formation task, we found radically different outcomes. This suggests the discrepancy between natural and sort-task categories is tied to the generation of candidate structures and reinforces a need for researchers to examine phenomena with an array of task formats.

We note a few limitations of the present work that motivate future studies. First, though we found and replicated the FR over UNI advantage, this advantage failed to replicate in E2. In future work, we aim to disentangle if and how the additional manipulations are attributable. Second, while our data show that people prefer FR when given candidate structures, our study does not speak to how people might initially generate FR as a candidate structure in the first place. We speculate this might occur through incidental experience (without class prediction) that leads to knowledge of feature statistics, as is hinted at by previous work (Lassaline & Murphy, 1996; Love, 2002). In future work, we intend to assess if a novel, repeated, item-matching task (like MTS, but without supervision) leads to such knowledge that transfers to the sorting task. Finally, the SCT involves an absolute, forced judgment between candidate structures. In future work, we plan to: (1) assess the degree to which people prefer one structure over another using a rating scale to determine if FR is viewed as a compelling way to structure categories as opposed to the better of two poor options; and, (2) afford the option of ‘no judgment’ to gain greater fidelity in our results.

References

- Ahn, W. K. (1999). Effect of causal structure on category construction. *Memory & Cognition*, 27(6), 1008-1023.
- Ahn, W. K., & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science*, 16(1), 81-121.
- Lassaline, M. E., & Murphy, G. L. (1996). Induction and category coherence. *Psychonomic Bulletin & Review*, 3, 95-99.
- Levering, K. R., & Kurtz, K. J. (2015). Observation versus classification in supervised category learning. *Memory & Cognition*, 43(2), 266-282.
- Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review*, 9(4), 829-835.
- Medin, D. L., Wattenmaker, W.D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness,

- and category construction. *Cognitive Psychology*, 19, 242-279.
- Milton, F., Longmore, C. A., & Wills, A. J. (2008). Processes of overall similarity sorting in free classification. *Journal of Experimental Psychology: Human Perception and Performance*, 34(3), 676-692.
- 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.
- Patterson, J.D., & Kurtz, K.J. (2018). Semi-supervised learning: A role for similarity in generalization-based learning of relational categories. In C. Kalish, M. Rau, J. Zhu, T. Rogers (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp 2211-2217). Austin, TX: Cognitive Science Society.
- Pothos, E. M., & Close, J. (2008). One or two dimensions in spontaneous classification: A simplicity approach. *Cognition*, 107, 581-602.
- Pothos, E. M., Perlman, A., Bailey, T. M., Kurtz, K., Edwards, D. J., Hines, P., & McDonnell, J. V. (2011). Measuring category intuitiveness in unconstrained categorization tasks. *Cognition*, 121, 83-100.
- Regehr, G., & Brooks, L. R. (1995). Category organization in free classification: The organizing effect of an array of stimuli. *The Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 347-363.
- Rosch, E., & Mervis, C. G. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- Wattenmaker, W. D. (1992). Relational properties and memory-based category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1125-1138.
- Wills, A. J., Milton, F., Longmore, C. A., Hester, S., & Robinson, J. (2013). Is overall similarity classification less effortful than single-dimension classification?. *The Quarterly Journal of Experimental Psychology*, 66(2), 299-318.
- Wittgenstein, L. (1953). *Philosophical Investigations*. Philosophische Untersuchungen. Oxford, England: Macmillan.
- Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition*, 34(2), 387-398.