Title
What comes to mind? Samples from relevance-based feature spaces

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Mills, Tracey
Phillips, Jonathan

Publication Date
2022

Peer reviewed

What comes to mind? Samples from relevance-based feature spaces

Tracey Mills (tracey.e.mills.22@dartmouth.edu)

Jonathan Phillips (jonathan.s.phillips@dartmouth.edu)
Program in Cognitive Science, Dartmouth College
23 N Main St. Hanover, NH 03755 USA

Abstract

Recent work in judgment and decision making has focused on which actions people consider when solving open-ended problems and found that the actions that come to mind tend to have particular features, such as having a high historical value. Here, we pursue the idea that the process of generating actions for decision-making tasks may actually reflect more general mechanisms for generating kinds of things. We provide evidence that what comes to mind may simply be a reflection of participants sampling from the most relevant part of the representational space they use to encode the type of thing they are generating. In this paper, we (1) introduce an approach for empirically describing a category in terms of the features that people use to represent category members, and for locating category members within that feature space, (2) show that certain locations in a category's feature space predict an item's likelihood of coming to mind, (3) introduce an approach for understanding the relevance of various features to people's representations of category members, and (4) show that features which are most involved in people's representations of category members are also predictors of what comes to mind within a category. We close by proposing that features that are most relevant to our representations of category members and predict coming to mind are those for which it has been historically useful to have information about during past experiences with the category in question.

Keywords: consideration sets; feature representation; categories; sampling; decision making

Introduction

When considering everyday decision making, it is natural to think about our (sometimes painful) deliberation between options. Should I order takeout for dinner, or put the decaying produce in my fridge to use? However, for many mundane decisions, before we can engage in such deliberation, we must first call to mind the options we then go on to deliberate over. While in many decision contexts we cannot possibly call to mind every potential option, we also cannot choose an option that we do not first call to mind, making this preliminary step key to the decision making process.

Prior work on what comes to mind during decision making has shown that people are remarkably skilled at generating candidate options; people can almost immediately generate a small set of options from an effectively infinite option space over a range of decision making contexts, and each of these options are generally good (Phillips, Morris, & Cushman, 2019; Johnson & Raab, 2003; Klein, Wolf, Militello, & Zsambok, 1995). While a container of gummy vitamins might lie somewhere in the option space for dinner tonight, it is much less likely than takeout or last week's groceries to come to mind as an option. More specifically, the options that come to mind have been found to be historically valuable, likely, and semantically accessible. (Morris, Phillips, Huang, & Cushman, 2021; Zhang et al., 2021; Bear, Bensinger, Jara-Ettinger, Knobe, & Cushman, 2020). In fact, Morris et al. (2021) find that even when people are asked to think of options of low value within a certain category, such as "Think of a food you'd least like to have for dinner", they can't help but call to mind options that are generally valuable within the category. Interestingly, in such decisions, people seem to be relying on a representation of an option's value relative to the category of thing being called to mind rather than just relative to the context at hand. If you're trying to think of what you'd least like to have for dinner, tacos may have a low context-specific value (perhaps you had them yesterday), but given that they are a generally highly valued dinner option, they will still be likely to come to mind, even if only to be dismissed.

Here, we pursue the idea that the process of generating options for decision-making tasks may simply reflect much more general mechanisms for generating kinds of things. Specifically, we will argue that the critical capacity for generating options of a certain kind in decision making likely relies on perfectly domain-general mechanisms for calling to mind instances of a category, kind, or concept. If you're trying to decide what kind of pet to get your child, you have to call to mind instances of the concept PET, and if you're trying to decide what to do for summer vacation, you have to call to mind kinds of vacations. Accordingly, work on option generation in decision making stands to benefit from the more general study of how we call to mind instances of a category or concept (for prior reviews on concepts and categories that discuss instance generation, see, (Battig & Montague, 1969; Mervis & Rosch, 1981; De Dayne, Navarro, Perfors, Brysbaert, & Storms, 2019).

Here, our approach will be to empirically demonstrate a domain-general framework for what comes to mind that allows us to account for the findings in the prior research on what comes to mind in decision-making contexts. We begin by examining what comes to participants' mind within different ordinary categories, and then analyze unifying patterns in the kinds of features that predict what comes to mind in each category. While prior decision-making work has made interesting progress in indicating that more generally valu-

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

able items within categories are more likely to come to mind, value can be ill-defined depending on the category in question. When considering crimes, for example, what comes to mind? It is probably not the least bad crimes (e.g., jay walking), which would, historically speaking, have the highest general value; rather, the crimes that intuitively come to mind seem to have a particularly low general value (e.g., murder). Accordingly, it may not be broadly true that general value determines what comes to mind. Rather, the role of value in determining what comes to mind when making decisions may instead illustrate a broader principle according to which features relevant for the category in question may determine what comes to mind.

To ask whether this is correct we seek to demonstrate, within various categories, a correspondence between the features that predict what comes to mind and the features which more naturally coincide with people's representations of members of the given category. On this general view, while the features that predict what comes to mind within each category will differ from each other, the most predictive features will consistently be those which are most central to people's representations of the category's members.

This interpretation of what comes to mind might also explain previous findings that the options generated during decision making are generally valuable. Specifically, this finding may simply be a reflection of how decision makers represent items in a category, with more relevant features likely being indicators of value and also determining what options are considered.

Approach

We propose to make progress on the question of what factors determine what comes to mind when thinking of members of a certain category. We consider 7 familiar categories of items: zoo animals, holidays, jobs, kitchen appliances, chain restaurants, sports, and vegetables. In each, we ask participants to tell us the category members that come to mind and then investigate what factors determine what is called to mind within each of these categories. We introduce a novel experimental technique for constructing the space of relevant features used to represent the members of each category, as well as for locating category members within the resulting feature space. McRae, Cree, Seidenberg, and Mcnorgan (2005) offer related ideas on how to describe and compare category members in terms of certain category-relevant features. For alternate approaches to empirically describe how categories and category members are organized in conceptual space, see (De Dayne et al., 2019) or (Rips, Shoben, & Smith, 1973).

We then demonstrate that an item's position along certain dimensions of a category's feature space, representing the degree to which it is described by various features, predicts its likelihood of coming to mind. Finally, we test a further prediction of this proposal. Specifically, if the set of features that we find to predict what comes to mind are simply part of the way that people represent that category, then they should be especially good at searching along those features (compared to features that do not predict what comes to mind). We find that this is the case: for a given category, the more a particular dimension predicts what comes to mind, the better participants are at generating instances at some end of that dimension. More generally, under the hypothesis we argue for, the process of calling members of a category to mind might be modeled as a search through feature space, weighted towards certain features that are relevant for that category.

What comes to mind

Participants recruited from MTurk ($N = 123$, $M_{age} = 38.0$, $SD_{age} = 10.0$, 62 females, 3 other) were presented sequentially with each of 10 categories (zoo animals, holidays, jobs, kitchen appliances, chain restaurants, sports, vegetables, types of furniture, types of clothing, and breakfast foods) and asked to list 10 items in that category as they came to mind. We excluded all responses where the instance generated was not a member of the given category, and reconciled similar responses into a single response. For example, in the zoo animals category, the responses 'otter,' 'otters,' and 'sea otter' all became 'otter.' Because of difficulty in disambiguating similar responses for 3 categories (types of furniture, types of clothing, and breakfast foods), we excluded these categories from subsequent analyses. From the resulting responses for each of the 7 remaining categories, we selected all items which had been listed by at least 2 participants. In the zoo animals category, we also added to this list animals which appeared in the animal lists of 3 popular U.S. zoos, so that zoo animals that were not called to mind by any participants, but might otherwise be expected to be called to mind, would be included in our analyses. Analogous procedures for the other categories did not result in the addition of items to the category list, because most popularly recognized instances of the category were already largely present in participant responses. These lists serve as the item list for each category in the remainder of our studies, with the following list sizes: (zoo animals: 63, holidays: 29, jobs: 85, kitchen appliances: 32, chain restaurants: 42, sports: 37, vegetables: 40). See **Fig**. 1 for an illustration of the zoo animals that come to mind.

A new method for empirically deriving category-specific representational spaces

Our proposal is that we may be able to make progress on understanding which category members come to mind by understanding their location in participants' representational space for that category. To pursue this idea, we first needed to find a way of empirically determining which features are relevant for each of the 7 categories used, and second we needed to determine where each category member fell along the relevant features. To achieve these goals, we developed a new method for empirically deriving category-specific representational spaces and then locating category members within the resulting space.

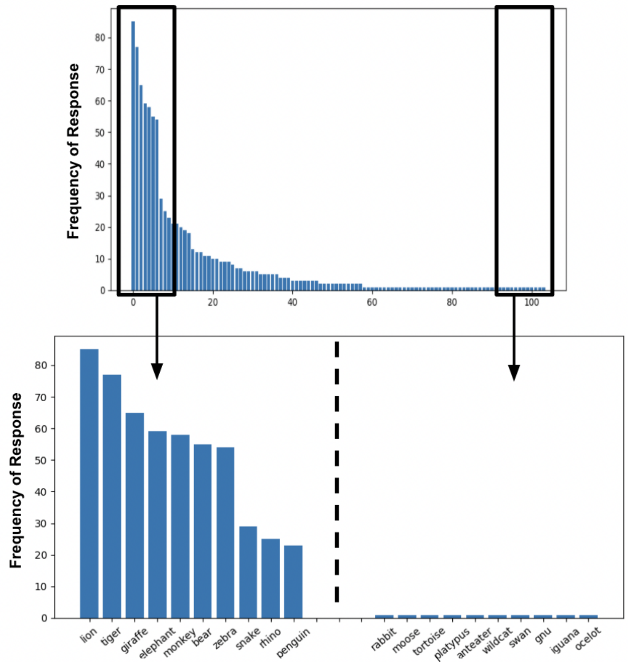

Figure 1: Distribution of the number of times each zoo animal came to mind across participants, from most (left) to least frequently.

Constructing a category's feature space

To get an idea of what features people find relevant when describing members in each category, we recruited participants from MTurk ($N = 147$, $M_{age} = 36.9$, $SD_{age} = 10.4$, 73 females, 4 other), assigned each participant to a category, and presented them with 10 pairs of members of that category derived from the prior experiment. For each pair, we asked them to tell us what made the two members similar or different. For example, participants assigned to the zoo animals category were told they would be comparing different zoo animals, and then for 10 trials were asked to list up to four similarities and four differences between two randomly paired zoo animals from the item list.

To illustrate, when comparing a panther and an owl, one participant remarked that both eat meat, while only an owl can fly. Features that were remarked upon at least twice within a category can be interpreted as relevant features for representing members of that category. We also introduced features that we expected to be less relevant to the category, such as 'has large feet relative to its body size' for the zoo animals category. From this set of variably relevant features, we constructed a given category's feature space with each feature as a dimension in that space. The dimensionality of each category's feature space is as follows: zoo animals: 30, holidays: 16, jobs: 16, kitchen appliances: 16, chain restaurants: 17, sports: 14, vegetables: 14. We can think think of a given category member's location in feature space as indicating how it is thought of in terms of these features relative to other category members, with features likely varying in relevance.

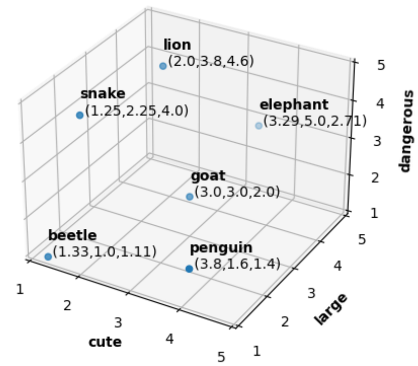

Figure 2: Visualization of a subset of feature space for the zoo animals category. This space has only 3 dimensions: cute, large, and dangerous. Zoo animals are located along each dimension of feature space according to the average participant rating for how well a feature describes that zoo animal.

Locating members of a category in category-specific representational spaces

To determine the location of each member in its category's feature space, we recruited additional participants from MTurk ($N = 292$, $M_{age} = 40.4$, $SD_{age} = 12.1$, 133 females, 2 other) to judge how well each feature described the items in a category's item list. Participants were again assigned to one of the 7 categories and told they would be answering questions about things in that category. They were then presented with members from that category (from the first studies) and asked to rate how well a series of features (taken from the previous study) described the category member, on a scale from 1 ('not well') to 5 ('very well'). The feature statements were of the general form 'This [category member] [has this feature].' For example, a participant assigned to the zoo animals category might be asked to answer questions about a llama in one trial, and would be asked to rate how well the feature statement 'This zoo animal has large feet relative to its body size' describes a llama.

We then took the average rating across participants for how well a feature described each category member as an estimate of that member's location along that dimension in the category's feature space. Thus, a member's location in category-specific feature space can be represented as a vector of ratings for each feature, or a point in the n-dimension feature space, see **Fig**. 2 for an illustration.

Predicting what comes to mind by location in feature space

We next asked whether the different locations of category members along the different dimensions of feature space can

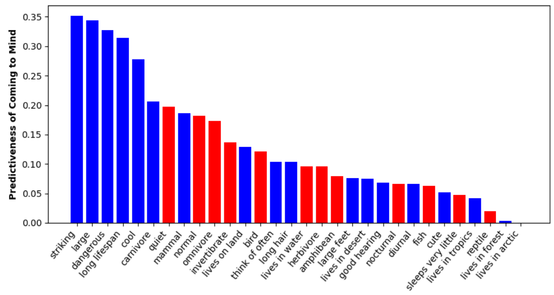

Figure 3: For each dimension in zoo animal feature space, the absolute value of the correlation between the zoo animal's location along that dimension and that zoo animal's likelihood of coming to mind. Directionally negative correlations are indicated in red.

help explain which category members come to mind. To do this, we simply calculated, for each dimension of feature space, how well each member's location along that dimension predicted the frequency with which it came to mind in our first study. A strong positive relationship would indicate that the more a given feature applies to a given category member, the more likely that category member is to come to mind. At the same time, a strong negative correlation shows something similar, since it indicates that the less a given feature applies to a given category member, the more likely it is to come to mind. For example, amongst zoo animals, the feature 'striking' is positively related with coming to mind, indicating that zoo animals that are considered more striking are more likely to come to mind, while the feature 'quiet' is negatively related with coming to mind, indicating that zoo animals that are considered less quiet are more likely to come to mind. Accordingly, because our goal is to identify features that are related to category members' likelihood of coming to mind, we analysed the absolute value of these relationships rather than the directional correlation. This approach revealed two key findings: (1) member location along a number of category-specific features was highly predictive of whether the member would come to mind, and (2) there was a large amount of variance in the predictiveness of the different features, see **Fig**. 3 for an illustration.

Determining relevant features and their relationship with coming to mind

We have seen that different dimensions of category-specific representational space vary in how well they predict whether or not a given category member comes to mind. Importantly, many of the features that predict what comes to mind are either orthogonal to value (e.g., 'diurnal'), or inversely related to it (e.g., 'dangerousness'). So, what makes some features more predictive than others? We hypothesized that features which predict coming to mind may simply be more relevant for our representations of category members in general.

What features are most relevant for representations of category members?

To understand what dimensions our representations of category members most strongly encode, we next designed an experiment to test how naturally people can think about category members in terms of different features. For each category, we selected a range of features from the constructed feature space which varied in their predictiveness of coming to mind.

Participants were recruited from MTurk ($N = 300$, $M_{age} = 38.3$, $SD_{age} = 12.1$, 148 females) and again assigned to one of the 7 categories, and then asked to list, over 8 trials, members of that category that had one of the selected features. On each trial, participants were asked to list as many category members as possible in 30 seconds. For example, participants assigned to the zoo animals category were told that they would be asked to list zoo animals that fit certain descriptions. In one of the 8 trials, a participant may be repeatedly asked to 'list a zoo animal that has large feet relative to its body size.'[1]

For each trial, we can estimate the ease with which the participant was able to think about category members in terms of the relevant feature from a combination of (1) the number of responses given during the 30 second trial and (2) the speed of each response. We quantified the "ease of response" for a single trial by calculating the sum of each response divided by the amount of time it took the participant to generate that response. So for trial t in which n responses were given with respective response times $rt_1, rt_2, ...rt_n$, the trial ease of response $t_{eor} = 1/rt_1 + 1/rt_2 + ... + 1/rt_n$. The ease of response for a certain feature is then calculated by taking the average ease of response for each trial in which participants are asked to list items with that feature, and normalizing this value by dividing it by the maximum ease of response for any feature in the category's feature space. So if FS is the set of all features in a category's feature space, and T_F is the set of all trials in which participants are asked to list items with feature F, then the ease of response to F, F_{eor} is calculated as $mean(t_{eor} \forall t \in T_F)/max(f_{eor} \forall f \in FS)$. Thus for each feature, the more responses participants tended to give, and the more quickly they tended to give these responses, the greater the ease of response to that feature (F_{eor}). We take the ease of response to a feature to indicate the extent to which participants encoded category members along that dimension, and from this point on refer to the ease of response to a feature simply as "feature relevance."

Similar to the previous study, this analysis revealed two key findings: (1) for each category, there were clear features which had a high ease of response, and (2) there was significant variation among features in terms of their ease of response, see **Fig**. 4 for an illustration.

[1] According to the feedback provided, not all participants enjoyed this task.

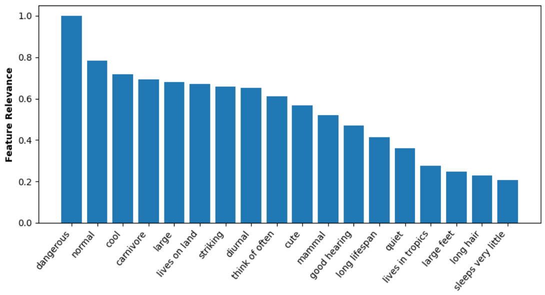

Figure 4: Feature relevance of each dimension of zoo animal feature space, based on the ease with which participants listed zoo animals that have that feature, according to our ease of response metric.

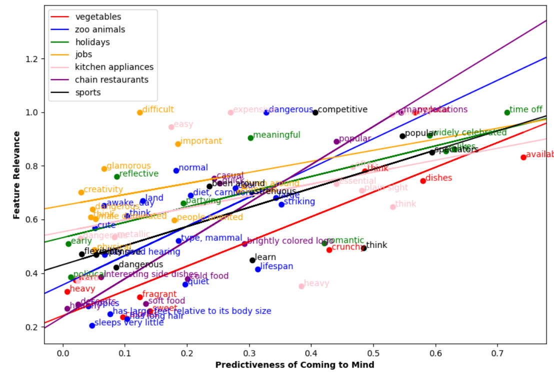

Figure 5: For each feature in all category feature spaces, predictiveness of coming to mind is plotted against feature relevance

The relationship between what comes to mind and feature relevance

Given the preceding set of results, we are now in a position to test the proposal we started with: that what comes to mind for a given category is a reflection of which members exist in the relevant part of the category-specific representational space. If this proposal is correct, it predicts that the feature relevance score for a given category-feature pair will predict that same feature's predictiveness for what comes to mind.

To as whether this was the case, we calculated the correlation between a feature's predictiveness of coming to mind and feature relevance over all 7 categories. We found a highly significant relationship overall, $r = 0.607, p < 0.001$, and clear positive relationships in each of the 7 categories: zoo animals ($r = 0.553$), vegetables ($r = 0.894$), holidays ($r = 0.748$), jobs ($r = 0.278$), kitchen appliances ($r = 0.390$), chain restaurants ($r = 0.911$), and sports ($r = 0.675$). See **Fig**. 5 for the overall relationship across categories, and **Fig**. 6 for an illustration in the case of zoo animals. The fact that this relationship exists across categories provides evidence that what comes to mind in general is a product of the features in terms of which we encode members of the category in question.

Prior work on concepts has found that category members that first come to mind are also judged to be the most typical of that category (Mervis, Catlin, & Rosch, 1976; Rosch, Simpson, & Miller, 1976; Janczura & Nelson, 1999; Barsalou, 1985). This suggests that members in the relevant portions of category feature space we identified are also more likely to be judged as typical. The idea that certain relevant features play an important role in determining a category member's typicality is consistent with existing literature (Rosch & Lloyd, 1978; Malt & Smith, 1984; Kellogg, 1981).

General Discussion

Across a large series of studies, we analyzed what came to mind for participants within 7 familiar categories (zoo animals, holidays, jobs, kitchen appliances, chain restaurants, sports, and vegetables) in terms of item locations in empirically constructed feature spaces. We find that within each category, certain features predict what comes to mind. In other words, an item's location along certain dimensions of a category's feature space predicts how likely it is to be called to mind. This finding encourages the conceptualization of the process of calling category members to mind as a search through the category's feature space, weighted towards certain dimensions. Alternatively, one can think of it as sampling from a relevance-based feature space. We find that across categories, these certain dimensions are also those that are relevant to people's representations of category members, with a given feature's predictiveness for coming to mind correlating positively with that feature's relevance within that category.

Our findings that features vary in their relevance to certain categories, and that the extent to which category members are well-described by more relevant features predicts their likelihood of coming to mind, can be used to interpret and enrich previous work on what comes to mind during decision making. While previous findings indicate that options generated during decision making tend to be generally valuable (Phillips et al., 2019; Johnson & Raab, 2003; Klein et al., 1995; Morris et al., 2021; Zhang et al., 2021; Bear et al., 2020), the notion of purely general value is vague or ill-defined across contexts. By contrast, our proposed framework–that what comes to mind is guided by the most relevant dimensions of feature space for the type of thing being generated–can explain option generation in contexts regardless of whether or not general value plays a role in option generation in that context. While a category's relevant features may sometimes approximate or be collinear with value (e.g., 'cool' in the case of zoo animals), it is also often the case that many of the features that are most relevant will be orthogonal or inversely related to a category member's value (e.g., 'diurnal' and 'dangerousness' respectively).

Just as this work seeks to build on and expand prior work on option generation in decision making, we have also sought

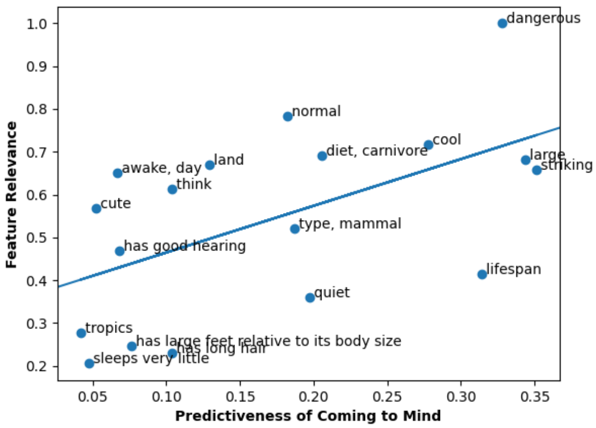

Figure 6: For each feature in zoo animal feature space, predictiveness of coming to mind is plotted against feature relevance

to go beyond prior work on instance generation in the case of established concepts or novel categories (Battig & Montague, 1969; Mervis & Rosch, 1981; De Dayne et al., 2019; Barsalou, 1985). In particular, we demonstrated a clear positive relationship between (i) the extent to which a feature is predictive of what comes to mind for a given category, and (ii) the extent to which people typically encode information about that feature for category members. We also sought to offer an explanation of this positive relationship: that the relevance of certain features to members of a category may be a product of the usefulness of having information about those features in past experiences with that category. For example, the feature ‘dangerousness’ may be especially relevant to a person’s representations of zoo animals because it has historically been useful for them to know how dangerous various zoo animals are (whether to avoid them, or just to tell their friends about them). This would explain why participants’ mental representations of various zoo animals may include some sort of cached ‘dangerousness’ metric. Since general value is presumably an extremely useful dimension across many different categories, this interpretation is consistent with some sort of ‘value’ feature indeed guiding what comes to mind in various contexts.

Our framework might also be used to explain the limited differences in what comes to mind within a category across different contexts. If what comes to mind is a product of a search through feature space, biased towards more relevant features, this search might also be biased towards features that are important in a certain context. For example, when thinking of a food you’d least like to have for dinner, what comes to mind would be biased towards generally relevant features (such as how much you like the food), but may also be biased towards contextually appropriate dimensions of feature space in terms of which you represent foods, such as ‘slimy’ or ‘gross.’ This proposition fits well with the findings of Morris et al. (2021), who found that what comes to mind in these low-value contexts tends to be a mix of generally ‘valuable’ responses but also foods that conform to the context-specific constraints (e.g., ‘not too moist”).

Future studies concerning what comes to mind might build off the work presented here in a number of ways. First, our proposition that the historical usefulness of having information about a feature determines how relevant that feature is to representations of category members might be tested by introducing people to a novel category and manipulating which features of category members are useful to learn. In such an experiment, we could then ask whether a feature’s usefulness in prior tasks affects what category members later come to mind in an open-ended task. Additionally, it would be worthwhile to further test our theory of what comes to mind by asking whether individual differences in what comes to mind are predicted by individual differences in feature relevance. Finally, the empirical tools we’ve developed in this paper might be applied more directly to decision making contexts in which the options being called to mind are possible actions.

In closing, we put forth the consideration of item locations in category-specific feature space as a useful framework for making progress on the question of what comes to mind. The method we have presented for constructing and locating items within such a feature space (as well as the relationships we have established between an item’s location in feature space and its likelihood of coming to mind) may be useful for future research on this question. We hope that the present work serves as an illustration of how research on how people build and navigate conceptual spaces can inform our understanding of decision making processes.

Acknowledgments

We would like to thank Nathan Schneider for his assistance with full stack web development.

References

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of experimental psychology: learning, memory, and cognition*, *11*(4), 629.

Battig, W., & Montague, W. (1969). Category norms of verbal items in 56 categories a replication and extension of the connecticut category norms. *Journal of Experimental Psychology*, *80*, 1–46.

Bear, A., Bensinger, S., Jara-Ettinger, J., Knobe, J., & Cushman, F. (2020). What comes to mind? *Cognition*, *194*, 104057.

De Dayne, S., Navarro, D., Perfors, A., Brysbaert, M., & Storms, G. (2019). The “small world of words” english word association norms for over 12,000 cue words. *Behavioral Research Methods*, *51*, 987–1006.

Janczura, G., & Nelson, D. (1999). Concept accessibility as the determinant of typicality judgments. *American Journal of Psychology*, *112*(1), 1–19.

Johnson, J. G., & Raab, M. (2003). Take the first: Option-generation and resulting choices. *Organizational behavior and human decision processes*, *91*(2), 215–229.

Kellogg, R. (1981). Feature frequency in concept learning: What is counted? *Memory and Cognition*, *9*, 157–163.

Klein, G., Wolf, S., Militello, L., & Zsambok, C. (1995). Characteristics of skilled option generation in chess. *Organizational behavior and human decision processes*, *62*(1), 63–69.

Malt, B., & Smith, E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 250–269.

McRae, K., Cree, G., Seidenberg, M., & Mcnorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*(37), 547–559.

Mervis, C., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, *7*, 283–284.

Mervis, C., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, *32*, 89–115.

Morris, A., Phillips, J., Huang, K., & Cushman, F. (2021). Generating options and choosing between them depend on distinct forms of value representation. *Psychological science*, *32*(11), 1731-1746.

Phillips, J., Morris, A., & Cushman, F. (2019). How we know what not to think. *Trends in cognitive sciences*, *23*(12), 1026–1040.

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

Rosch, E., & Lloyd, B. (1978). *Cognition and categorization*. Lawrence Elbaum Associates.

Rosch, E., Simpson, C., & Miller, S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 491–502.

Zhang, Z., Wang, S., Good, M., Hristova, S., Kayser, A. S., & Hsu, M. (2021). Retrieval-constrained valuation: Toward prediction of open-ended decisions. *Proceedings of the National Academy of Sciences*, *118*(20).