How Do We Choose What to Think? Semantic Space Navigation and Individual Variation

Doris Yu (dorisyu@uw.edu)
Foster School of Business, University of Washington
Jonathan Phillips (Jonathan.S.Phillips@dartmouth.edu)
Department of Cognitive Science, Dartmouth College
Max Kleiman-Weiner (maxkw@uw.edu)
Foster School of Business, University of Washington

Abstract

How do people generate options in everyday purchase decisions? Building on prior work, we studied how people search semantic space when generating options for daily-use products. Study 1 (N=200) showed that items coming to mind first were both more preferred and more frequently used, demonstrating that accessibility and personal preference predict retrieval order. Studies 2-3 (N=220) constructed empirically-derived feature spaces for product categories. Within these spaces, semantic distance weakly predicted generation speed, showing Lévy-like foraging but with substantial deviation. This raised the possibility that individual differences fundamentally shape semantic search in option generation. Study 4 (N=100) directly tested this interpretation, revealing that mental search is recalcitrant: broad category activation intrudes into subsequent feature-focused search, while feature-focused search first facilitates later global exploration. These findings demonstrate that option generation reflects population-level foraging dynamics modulated by individual differences in how people weight features and navigate personalized semantic neighborhoods.

Keywords: consideration sets, decision making

Introduction

Imagine you are deciding where to go for lunch among the dozens of fast food restaurants that are within walking distance. When you are seeking something light and fast, you may think of going to places like Sweetgreen, Chipotle, or Subway. However, as soon as your friend says they want fried chicken, you cannot get the thought of Chick-fil-A or KFC out of your mind.

This seemingly effortless process happens every day, but it masks a fundamental challenge in decision-making: how do people generate options from vast option spaces? Real-world decisions often begin with an open-ended search through memory. While prior work has identified that familiar, accessible, and high value options tend to be considered, the cognitive mechanisms underlying how people search through semantic and episodic memory to construct consideration sets, the small subset of alternatives that will be evaluated for decision-making, have received less systematic attention. Recent work has begun to address this gap by examining how option generation follows structured sampling principles (Morris et al., 2021; Mills & Phillips, 2023; Dracheva & Phillips, 2025), yet fundamental questions remain about how the mind navigates representational spaces during naturalistic, goal-directed decision-making.

Phillips et al. (2019) show that people consistently sample actions that are both probable and generally valuable across diverse tasks. Morris et al. (2021) reveal that option generation relies on cached, context-free values while final choice depends on context-specific evaluation. Mills & Phillips (2023) demonstrate that what comes to mind can be predicted by items' locations in empirically-derived feature spaces, where certain category-relevant dimensions guide retrieval. This structured sampling resembles optimal foraging behavior. People navigate semantic spaces to maximize options retrieved per unit cognitive effort. Previous work shows that inter-item distances in semantic memory retrieval approximate Lévy flight distributions characteristic of optimal foraging (Rhodes & Turvey, 2007; Hills et al., 2012; Montez et al., 2015). Extending this to decision-making, Dracheva & Phillips (2025) found that both inter-generation times and semantic distances between consecutively generated actions follow Lévy distributions.

This search process is, at least in part, automatic and resistant to control. Morris et al. (2021) showed people involuntarily think of high-value options even when explicitly asked to generate low-value ones, with such intrusions occurring six times more frequently than the reverse. Mills & Phillips (2023) found that "large" animals came to mind even when participants were asked to think of small animals. This reveals an important constraint: while people can ultimately select task-appropriate options, they cannot entirely prevent their default search mechanisms from initially activating strongly-weighted semantic regions.

To investigate real-world decision-making processes and predict the sequence of items that come to mind, we conducted four experiments adapted from Mills & Phillips (2023) (illustrated in Figure 1A). We sought to determine how generating choices from specific versus general feature spaces influences the formation of consideration sets in everyday product categories. Furthermore, we examined how external cues shift search patterns within structured representational spaces. We tested three central hypotheses: **(1)** Items retrieved first reflect both general accessibility and relevance to daily usage; **(2)** Semantic search follows a foraging pattern where individuals exploit local "neighborhoods" before making cognitive jumps to distant regions; **(3)** Activating specific feature dimensions (e.g., "restaurants that sell fried chicken") systematically biases subsequent search, either facilitating or interfering with

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

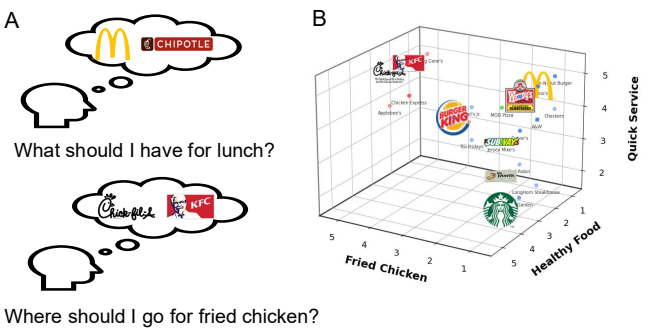

Figure 1: Illustration of an everyday decision making problem. (**A**) Sample of questions we ask participants. (**B**) Embedding of frequently selected chain restaurants. The 3D feature space represents the most common participant responses along three dimensions: healthiness, availability of fried chicken, and service speed.

broader category exploration depending on the search order.

Method

To investigate the systematic nature of consumer search across diverse contexts, we conducted four experiments adapted from Mills & Phillips (2023) focusing on highly familiar product categories including fast-food restaurants, breakfast cereals, sneakers, and soft drinks.

Participants

We recruited participants from Prolific who were current US residents. We collected demographic information including zipcode, age, gender, education level, and household income. All studies received IRB approval.

Procedure

We conducted four experiments to investigate how people search through semantic space during everyday decision-making:

- Study 1: Collected baseline consideration sets representing natural, spontaneous retrieval (N=200) in each category through timed free generation, usage frequency ratings, and preference rankings.
- Study 2: Elicited features (N=100) of the items from participants through pairwise comparisons of items retained from Study 1. These elicited features were clustered into 8–10 features per category using GPT-4o-mini.
- Study 3: Located items in feature space (N=120) by collecting ratings for 30 item-feature pairs per participant.
- Study 4: Examined search dynamics (N=100) by comparing general category searches versus feature-specific searches (e.g., "all restaurants" vs. "restaurants with fried chicken", see 1A) in counterbalanced order.

Study 1: Baseline Consideration Sets. Participants were presented with a realistic purchase context (e.g., "Imagine you are ordering a quick meal from a fast food restaurant") and asked to list all brands that came to mind within 120 seconds, with bonus compensation for generating more than seven brands. Participants were instructed to enter items as quickly as possible without worrying about correct spelling or names. They typed each item and pressed Enter to proceed to the next entry. After the generation phase, participants evaluated each brand on three dimensions: purchase likelihood ("How likely are you to buy this item next?"), familiarity (1–7 scale: "How familiar are you with this brand/product?"), and recency of use (six-point scale from "today" to "never"). Finally, all generated brands were presented in shuffled order, and participants ranked them by choice preference.

Study 2: Feature Elicitation. To ensure reliable feature extraction, we retained only items mentioned by at least two participants in Study 1, which included 49 items in breakfast cereals, 57 in fast food/chain restaurants, 46 in soft drinks, and 36 in sneakers. Participants first indicated which items they recognized to ensure familiarity. Then, completed 10 comparison trials. On each trial participants listed one similarity or one difference between randomly paired items. This pairwise comparison approach allowed participants to naturally generate the features they use to distinguish between category members. We used GPT-4o-mini to systematically extract and categorize features from participants' free-text responses. GPT-4o-mini was used only to consolidate semantically redundant participant-generated descriptors into higher-level feature dimensions, and resulting feature clusters were manually reviewed for interpretability. To identify the most diagnostic features for each category, we retained only features that explained more than 20% of comparisons within that category.

Study 3: Feature Ratings. To locate each item within the feature space identified in Study 2, participants rated how well features described specific items. Each participant was assigned to one category and rated 30 randomly selected item-feature pairs on a 5-point scale ranging from 1 (feature does not describe the item at all) to 5 (feature describes the item very well), with 3 as a neutral midpoint. Participants could indicate "not known" if they were unfamiliar with the item or "not sure" if they were uncertain about the feature. This distinction allowed us to identify items with low general familiarity and avoid treating missing knowledge as low feature ratings, reducing noise in the data.

Study 4: Search in Global and Local Feature Space. To examine how the prior activation of a specific feature dimension influences subsequent search, we manipulated the task order between general and feature-specific generation. Participants were randomly assigned to one of two conditions: in the *Feature→General* condition, they first completed a feature-specific trial (e.g., listing restaurants that sell fried chicken) followed by a 30-second general recall trial (e.g., listing any restaurants); in the *General→Feature* condition, this order was reversed.

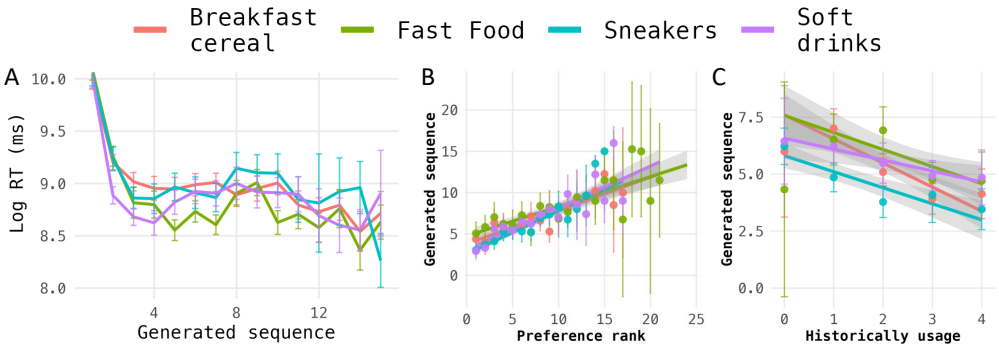

Figure 2: Results from Study 1 (colored by category). (**A**) Log-transformed inter-response intervals (RT) by item generation order (1st to 15th items). (**B**) Correlation between preference rankings and item generation order. (**C**) Correlation between daily usage frequency and item generation order.

Results

Unless otherwise noted, analyses used linear mixed-effects models with participant random intercepts and category included as a fixed effect.

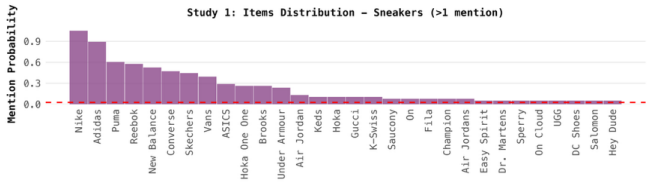

Figure 3: Distribution of sneaker brand accessibility across participants. Bars show the proportion of participants who mentioned each brand. Items mentioned by fewer than two participants were excluded.

Preference and Accessibility in Consideration Sets

In daily decision-making, consideration sets are not exhaustive. Instead, people sample from a latent representational space where specific dimensions are weighted by their historical usefulness or relevance for that category (Tversky & Kahneman, 1973).

We examined which items participants spontaneously generated in Study 1. Participants freely generated items within 120 seconds across four product categories, producing an average of 11.3 items for fast food/chain restaurants (SD = 0.99), 10.0 for soft drinks (SD = 0.72), 9.7 for breakfast cereals (SD = 0.72), and 8.2 for sneakers (SD = 0.60). Despite generating only a small subset of available brands in each category, participants showed striking consensus in which items came to mind. Dominant brands showed near-universal accessibility. For example, in the sneaker category, Nike was mentioned by 98% of participants, with 30.2% listing it as their first item, followed by Adidas which 89% of participants listed, with 24.8% listing it first (Figure 3). This high consensus, despite the open-ended nature of the task and the large number of available items within each category, suggests that participants sample from a relevance-biased manifold where big brands act as high-probability anchors in semantic space.

To understand what determines which items first come to mind during everyday decision-making, we examined how usage recency of the items and preference relate to the order in which options are generated. Analysis of inter-response intervals revealed a distinctive temporal pattern (Figure 2A). The first item showed the longest generation time (M = 27.56 ± 1.53 s), likely inflated by the time required to process the task instructions. Response times decreased sharply for the second item (M = 11.34 ± 0.76 s) and third item (M = 8.71 ± 0.68 s), then remained relatively stable (M = 8.77 s, range: 7.38–9.95 s) from positions 4 through 12.

This pattern contrasts with previous findings suggesting increasing generation times as participants exhaust more accessible options and must traverse greater semantic distances to locate additional exemplars (Troyer et al., 1997; Hills et al., 2012). The relatively flat response time profile from items 2 to 12 suggests that participants may employ efficient search strategies within structured semantic neighborhoods, rather than making large jumps across semantic space. One alternative explanation for the flat IRI profile is a motor bottleneck, where typing speed limits the rate of output rather than cognitive search. However, typical inter-key intervals for non-professional typists range between 60-80 words per minute (Gentner, 1983; Salthouse, 1984). Given that the average IRI in our study was 8.71 ± 0.68 s—significantly longer than the time required to physically type a single brand name—it is unlikely that motor execution was the rate-limiting factor. Instead, this stability suggests a consistent cognitive foraging strategy.

Based on Mills & Phillips (2023), we investigated whether generation order reflects underlying preferences. After the free recall task, participants rank-ordered the brands they had generated, with lower ranks indicating higher preference (rank 1 = most preferred). We found a significant positive correlation between preference rank and recall sequence (r = 0.443, R^2 = 0.196, β = 0.44, $p < .001$; Figure 2B), demonstrating

that preferred items (those with lower rank numbers) came to mind earlier during free generation (lower generation positions). In other words, the items participants preferred were also the items that came to mind first.

Recency Usage and Recall Order

Following Thomas et al. (2008), who demonstrated that ease of generation influences frequency judgments, likelihood estimates, and preference ratings, we examined the relationship between historical usage patterns and generation dynamics. We asked participants when they had last used each item, from today to never used. To test whether more frequent usage predicts earlier generation, we computed the correlation between the inverse of usage frequency (where higher values represent more recent/frequent use) and generation position. We observed a significant positive correlation ($r = -0.332$, $R^2 = 0.110$, $\beta = -1.05$, $p < .001$; Figure 2C). This indicates that items with more recent usage were generated earlier in the sequence (lower generation position values). This finding supports the accessibility principle: items with more recent or more frequent usage come to mind more readily.

Search in Semantic Space

To understand how participants navigate semantic space during option generation, we examined the relationship between inter-response intervals and semantic distance between consecutively generated items. Prior work by Dracheva & Phillips (2025) showed mental search follows Lévy flight principles: long periods of local exploitation within a semantic "patch" (short steps, low thinking time) followed by rare, effortful "jumps" to new patches. Under this framework, semantic distance between the consecutive generated items should predict the thinking time, with larger jumps requiring high cognitive cost. Therefore, in our study, we used the features identified in Study 2 and item locations constructed in Study 3 to calculate the cosine distance between consecutive brands in the high-dimensional feature space collected previously and correlated it with the thinking time required to generate each item.

Contrary to this prediction, we found remarkably weak correlations between cosine distance and log response time across all categories (Figure 4). Overall, $R^2 = 0.017$ ($r = 0.130$, $\beta = 4.55$, $p < 0.001$), with category-specific effects being similarly weak: breakfast cereal ($R^2 = 0.005$, $r = 0.071$, $\beta = 1.87$, $p = 0.205$), chain restaurants ($R^2 = 0.019$, $r = 0.138$, $\beta = 5.87$, $p = 0.011$), sneakers ($R^2 = 0.026$, $r = 0.161$, $\beta = 7.82$, $p = 0.007$), and soft drinks ($R^2 = 0.001$, $r = 0.033$, $\beta = 1.50$, $p = 0.559$).

The weak correlation between semantic distance and inter-response intervals reveals a departure from traditional Lévy flight dynamics. Participants in this study primarily exploit densely clustered local semantic neighborhoods, leading to minimal semantic distance and consistent retrieval times. Most consecutive items originate from the same semantic "patch," effectively decoupling semantic distance from the variation in thinking time. We propose that when participants hold strong preferences for daily goods, they immediately enter a preferred semantic neighborhood and efficiently exhaust items within that local region before considering distant alternatives.

To directly test whether people can flexibly navigate between general category space and specific feature subspaces, we designed Study 4 to manipulate search targets explicitly. By comparing performance when participants search the global category space versus feature-constrained subspaces, we can examine how activation of different semantic regions influence subsequent search and whether feature-based search leaves lasting traces that facilitate or interfere with later retrieval.

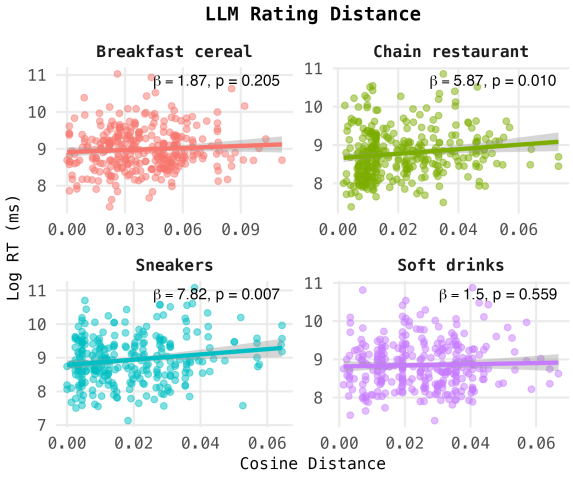

Figure 4: Log-transformed inter-response intervals (RT) as a function of cosine distance between consecutive items in human semantic space for each product category.

Foraging in Global vs Local Space

In Study 4, participants were randomly assigned to one of two task orders: either a general trial (e.g. listing all fast food restaurant come to mind) followed by two feature-specific trials (e.g. list fast-food restaurants that serve fried chicken), or the reverse sequence (feature trials first, then general trial). This design allowed us to test whether searching within feature-constrained semantic subspaces is more cognitively demanding than searching the full category space.

Searching in Feature Subspace is Harder than General Conceptual Space We quantified retrieval fluency using ease of response, calculated as:

$$\text{Ease of Response (EOR)} = \sum_{i=1}^{n} \frac{1}{\text{RT}_i} \tag{1}$$

Supporting this hypothesis, we used linear mixed effects models with participants as random effects to account for within-participant repeated measures across all analyses, where RT_i is the response time (in seconds) for the i-th brand in Equation 1. This metric captures both the number of items

retrieved and the speed of retrieval, with higher values indicating more fluent recall. To enable comparison across trial types, we standardized these values by dividing by the number of brands available in the relevant semantic space: for feature trials, we used brands rated ≥ 4 on that feature in Study 3; for general trials, we used all brands in the category.

First, feature trials showed significantly longer first-item response times ($M = 7.73 \pm 0.24$ s) compared to general trials ($M = 5.91 \pm 0.35$ s), regardless of task order ($\beta = -2.18$, $t(320.5) = -3.83$, $p < .001$; interaction with task order: $F(1, 319.5) = 0.70$, $p = .403$). This suggests that initiating retrieval from a feature-constrained subspace requires additional cognitive effort to locate and enter the appropriate semantic neighborhood. Second, ease of response was significantly lower for feature trials compared to general trials ($\beta = 0.004$, $t(314.8) = 6.65$, $p < .001$, see Figure 5A). This large effect indicates that participants experienced substantially less fluent retrieval when searching within feature subspaces. Third, participants generated fewer brands in feature trials ($M = 4.35 \pm 1.95$) than in general trials ($M = 6.13 \pm 2.00$)), significant with participants as random effect ($\beta = 1.84$, $t(314.9) = 10.33$, $p < .001$ (Figure 5B). When standardized by the number of available brands in each semantic space, participants retrieved approximately 11.2% of feature-relevant brands versus 13.4% of all category brands within same amount of time, indicating that feature constraints limit the accessible portion of semantic memory.

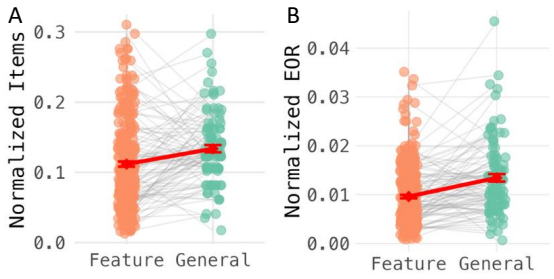

Figure 5: Search in feature subspace versus general space. (**A**) Standardized number of items generated in each condition (standardized by the number of items available in the space). (**B**) Ease of response of search in feature versus general space.

Activation and Inhibition

Representational weights are recalcitrant, which means they guide search even when we actively try to override them. If the general category space is indeed the default mode of semantic access, then searching the global space first should limit participants' ability to subsequently enter specific feature subspaces. We hypothesized that the order of general versus feature trials would produce distinct patterns: (1) When general trials precede feature trials, we expected increased intrusions of general-space items into feature trials, reflecting difficulty inhibiting the activated global representation, and reduced ease of accessing feature subspaces, as participants must overcome the default global activation. (2) When feature trials precede general trials, we expected increased overlap between feature and general trials, as activating feature subspaces facilitates deeper search within the global space, and increased ease of search in the global space due to stronger activation. Those patterns would demonstrate a conflict between automatic generation (what comes to mind based on general relevance) and goal-directed selection (the specific feature-based goal).

We used linear mixed effects models with participant random intercepts across all analyses to account for the nested structure of multiple trials within participants.

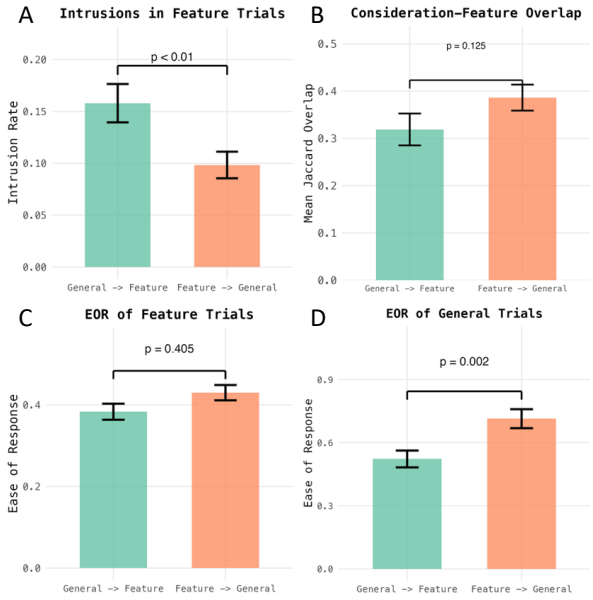

Figure 6: Comparison of feature and general trials by condition order. Green indicates general trials before feature trials; orange indicates feature trials before general trials. (**A**) Intrusion rate (probability of a non-feature item intruding in feature trials) in feature trials by condition . (**B**) Jaccard similarity between feature and general trials by conditions. (**C**) Ease of response in feature trials by condition. (**D**) Ease of response in general trials by condition.

Feature Intrusions We calculated intrusion rates as the probability of generating brands that were rated <3 on the target feature in Study 3. These rates differed significantly based on task order, with more frequent intrusions occurring when the general trial preceded the feature trial ($M = 0.158 \pm 0.211$) than when the feature trial occurred first ($M = 0.098 \pm 0.171$), $F(1, 80.9) = 7.20$, $p = .009$ (Figure 6A). This result suggests that when the general category space is activated first, it continues to "intrude" into subsequent feature-based search, causing participants to retrieve brands that are generally relevant to the category but do not possess the target feature.

Overlaps in Consideration Sets We examined whether performing feature trials first might facilitate subsequent general search by pre-activating relevant semantic networks. We computed Jaccard overlap coefficients between each partici-

pant's feature trials and general trial to quantify the degree of shared content. Participants in the Feature→General condition showed higher overlap ($M = 0.39 \pm 0.25$) compared to those in the General→Feature condition ($M = 0.32 \pm 0.23$), though this difference was not statistically significant (Figure 6B). These order effects may reflect both automatic semantic activation and strategic reuse of previously generated items across trials, though additional evidence would be needed to support this interpretation.

EOR by Trial Order. We examine task order effects on EOR in feature and in general trials. We found minor evidence that prior general search impairs subsequent feature search (General→Feature: $M = 0.38 \pm 0.23$; Feature→General: $M = 0.43 \pm 0.25$, 5C), though this did not reach statistical significance ($\beta = 0.033$, $t(99.3) = 0.837$, $p = .405$; Figure 6A). EOR in general trials showed participants who performed feature trials first exhibited significantly higher retrieval fluency in the subsequent general trial (Feature→General: $M = 0.71 \pm 0.33$) compared to those who performed the general trial first (General→Feature: $M = 0.52 \pm 0.27$; $t(96.8) = -3.18$, $p = .002$; Figure 6D). This large effect demonstrates that prior feature-based search systematically activated the accessibility of that concept for subsequent global category exploration.

Consideration Set Composition Predict Feature-Specific Foraging. If representational weights truly guide automatic retrieval, then the composition of participants' spontaneously generated consideration sets should predict their feature-based foraging efficiency. Specifically, if a participant's general consideration set is dominated by items high in a particular feature (e.g., healthy options), they should demonstrate greater fluency when explicitly searching for items within that feature subspace. For each participant, we correlated their EOR for each feature in feature trials with the average rating of that feature across brands they generated in their general trial.

We observed weak positive correlations between feature-specific EOR and feature density in general trials across most categories (Breakfast cereal: $r = 0.31$, $R^2 = 0.10$, $\beta = 0.09$, $p = 0.007$; Soft drinks: $r = 0.33$, $R^2 = 0.11$, $\beta = 0.09$, $p = 0.001$), with one exception showing a weak negative trend (Fast Food: $r = -0.21$, $R^2 = 0.04$, $\beta = -0.06$, $p = 0.099$) and one showing minimal relationship (Sneakers: $r = 0.07$, $R^2 = 0.00$, $\beta = 0.02$, $p = 0.619$; Figure 7). These results likely reflect substantial individual differences rather than the absence of a relationship. Different individuals appear to organize the same product categories around different feature dimensions based on personal experience and preferences. This heterogeneity means that individuals access different semantic neighborhoods with varying efficiency, shaping which items spontaneously come to mind in ways that vary systematically across people but average out at the population level.

Discussion

Our research extends prior work from (Mills & Phillips, 2023) to investigate how people search in semantic space when generating options for daily purchases. Across four studies, we found that consideration sets reflect structured sampling from a relevance-weighted semantic space rather than exhaustive search (Vul et al., 2014; Simon, 1955). Items generated first were both more preferred and more frequently used in daily life, confirming that accessibility is a powerful predictor of eventual choice.

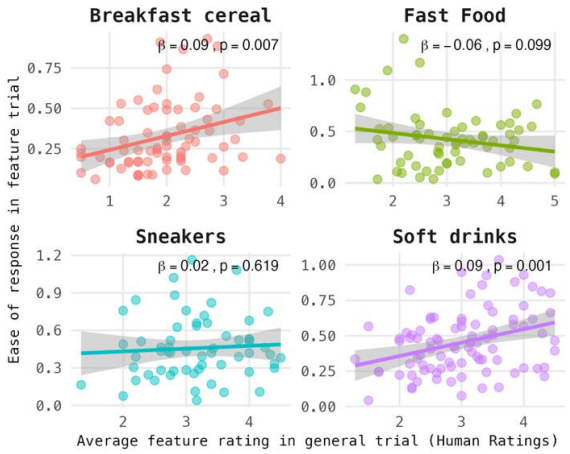

Figure 7: Ease of response in feature trials predicted by mean feature ratings of items from general trials across product categories.

However, the weak correlation between semantic distance and inter-response intervals suggests a weaker relationship between semantic distance and retrieval latency than predicted by traditional Lévy flight accounts. Our results specifically demonstrate that representational weights are recalcitrant. When the general category space is activated first, it intrudes into subsequent feature-based search, making it difficult for participants to inhibit globally relevant but feature-inappropriate items. Conversely, prior feature-based search facilitated subsequent general exploration, revealing an asymmetric relationship between local and global semantic search. These findings extend prior work on option generation (Morris et al., 2021; Mills & Phillips, 2023; Dracheva & Phillips, 2025) by demonstrating that, in real-world contexts, retrieval reflects a dynamic interaction between automatic activation patterns and goal-directed constraints.

The weak correlation regarding search efficiency across specific feature subspaces may reflect substantial individual differences in how people construct and navigate their mental representations. Individuals appear to weight features differently and access semantic neighborhoods with varying degrees of efficiency, ultimately shaping the composition of their consideration sets. Future work will compare consumer-domain retrieval against classical semantic fluency tasks to determine whether the present effects are specific to highly familiar product categories.

References

Dracheva, A., & Phillips, J. S. (2025). Naturalistic action

sampling as foraging in the option space. In D. Barner, N. R. Bramley, A. Ruggeri, & C. M. Walker (Eds.), *Proceedings of the 47th annual conference of the cognitive science society* (pp. 5706–5712). Cognitive Science Society.

Gentner, D. R. (1983). Keystroke timing in transcription typing. In *Cognitive aspects of skilled typewriting* (pp. 95–120). Springer.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*(2), 431–440. doi: 10.1037/a0027373

Mills, C., & Phillips, J. (2023). Locating what comes to mind in empirically derived representational spaces. *Cognition*, *240*, 105549.

Montez, P., Thompson, G., & Kello, C. T. (2015). The role of semantic clustering in optimal memory foraging. *Cognitive Science*, *39*(8), 1925–1939. doi: 10.1111/cogs.12249

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.

Rhodes, T., & Turvey, M. T. (2007). Human memory retrieval as Lévy foraging. *Physica A: Statistical Mechanics and its Applications*, *385*(1), 255–260. doi: 10.1016/j.physa.2007.07.001

Salthouse, T. A. (1984). Effects of age and skill in typing. *Journal of Experimental Psychology: General*, *113*(3), 345.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*(1), 99–118.

Thomas, R. P., Dougherty, M. R., Sprenger, A. M., & Harbison, J. I. (2008). Diagnostic hypothesis generation and human judgment. *Psychological Review*, *115*(1), 155–185.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138–146. doi: 10.1037/0894-4105.11.1.138

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599–637.