UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
How We Search in Our Minds: Modeling Task-Specific Retrieval Dynamics

Permalink
https://escholarship.org/uc/item/40n815vw

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Kedrick, Kara
Liang, Jonas
Golman, Russell

Publication Date
2026

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

How We Search in Our Minds: Modeling Task-Specific Retrieval Dynamics

Kara Kedrick (kkedrick@andrew.cmu.edu)[1], Jonas Liang[2] & Russell Golman[3]
[1]Institute for Complex Social Dynamics, Carnegie Mellon University
[2]Department of Mathematics, University of Illinois Urbana-Champaign
[3]Social and Decision Sciences, Carnegie Mellon University

Abstract

How do people search their memory when answering a question? To examine how this process relates to the degree of constraint imposed by a question, we asked 329 participants to generate sequences of responses while answering either more constrained trivia questions or broader category-based questions. Across both tasks, answer generation reflected common retrieval dynamics: participants were more likely to produce generally accessible answers, and transitions were strongly shaped by similarity to previously generated responses. However, answers generated in response to trivia questions were more strongly aligned with the semantic content of the question itself. Analyses of response times also revealed a dissociation in retrieval dynamics while answering specific questions: answers semantically similar to prior responses were generated more quickly, whereas answers more closely aligned with the question were generated more slowly. This pattern suggests that more specific questions engage an additional control process that intermittently redirects memory search toward the question. The pattern does not arise from a random walk through semantic space with constant transition probabilities, but instead supports a view of memory search as a foraging process, in which search alternates between associative small steps (local exploitation) and question-directed big jumps (wider exploration) as the ease of finding a local transition varies.

Keywords: semantic search; question answering; foraging; random walk

Introduction

How do people retrieve semantic memories? Research on answer generation in response to broad questions—such as category exemplar retrieval, in which participants list items from a given category—suggests that this process is dynamic rather than static (Gronlund & Shiffrin, 1988; Hills et al., 2012; Troyer et al., 1997). People tend to transition from one retrieved example to another that is semantically related. Classic theories of semantic memory formalized this process as spreading activation through a structured network (Anderson, 1983; Collins & Loftus, 1975), and show that responses cluster by semantic similarity and exhibit systematic transition patterns (Hills et al., 2012; Troyer et al., 1997). In cases of category exemplar retrieval, question constraints are loose and activation spreads freely. Here, we ask how memory retrieval is altered when the task requires answering a specific question. Akin to category exemplar retrieval, people engage in an active search process, generating a sequence of candidate answers while moving through semantic space. As part of this process, the specific question may preferentially activate particular regions of the semantic space. Understanding how people navigate this search—what guides which answers come to mind and when they abandon a line of thought to explore alternatives—remains an open question.

In this study, we contrast answering specific trivia questions with answering broad, category-exemplar retrieval questions to examine how memory search differs across contexts. Relatively more specific questions should impose semantic constraints that bias retrieval toward answers more closely related to the question. For example, when asked *"What musical instrument was invented to sound like a human singing?"*, answers that align with the question—such as instruments commonly associated with voice-like qualities—should be especially likely to be retrieved. Consistent with this idea, computational models of associative retrieval show that incorporating contextual cues systematically improves the predictability of generated responses (Richie et al., 2023).

It is also possible that answer generation in response to specific questions is influenced by the same features that shape category exemplar retrieval. In particular, answer generation in this context may be biased by general accessibility, a well-documented feature of semantic fluency questions that involve category retrieval. Prior work on category retrieval shows that items encountered more frequently in everyday life are more likely to be generated (Zhang et al., 2021). If similar dynamics operate while answering specific questions, highly familiar concepts may come to mind even when less common but more appropriate concepts are available. For example, when asked *"What musical instrument was invented to sound like a human singing?"*, highly familiar instruments (e.g., *piano*) may be generated despite better-aligned alternatives.

Answer generation in response to specific questions may also be shaped by previously generated answers, a pattern well documented with semantic fluency questions. This process has been modeled as a random walk through semantic space, such that the probability of generating a given answer is determined by its similarity to the preceding answer (Abbott et al., 2015). Under this view, answer generation is governed primarily by associative similarity: producing one answer increases the likelihood of generating a semantically related answer next. For example, when answering the trivia question *"What musical instrument was invented to sound like a human singing?"*, a random walk model produces a tendency toward local continuity, such that generating an initial answer

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* ©2026 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

like *violin* increases the probability of generating other similar instruments (e.g., *viola*). To the extent that people are also drawn toward answers that they associate with the question, the transitions to these answers in the random walk may be more likely. Although a random walk can eventually drift into other regions of semantic space, such transitions arise passively through the structure of the network rather than through goal-directed jumps.

We hypothesized that answer generation in response to specific questions reflects the joint influence of three factors. First, the semantic content of the question should selectively cue relevant answers, preferentially activating concepts aligned with those in the question. Second, general accessibility should bias retrieval toward answers that are more commonly encountered, even when they are not optimal. Third, retrieval should preserve local associative structure, such that successive answers are more likely to be semantically related. Consistent with these predictions, we find that question relevance and general accessibility systematically shape retrieval while preserving local transition dynamics.

An additional prediction of the random-walk model is that answers with higher retrieval probability should also be retrieved more quickly. We tested this prediction and found the opposite pattern for specific questions: answers more closely related to the question were retrieved more slowly, whereas answers more closely related to the previous response were retrieved more quickly. This pattern suggests that the retrieval dynamics rely on a foraging process, in which memory search is organized around alternating modes of exploration and exploitation (Hills et al., 2012, 2015). Rather than continuously traversing semantic space through local transitions, individuals are thought to exploit local "patches" of related information before deliberately disengaging and exploring a different region of memory. In addressing the same trivia question, a foraging account predicts that individuals may initially start by generating an answer that is similar to the question (e.g., violin). They would proceed to generate several closely related answers from the region of the semantic space (or patch) in which the initial answer belongs (e.g., string instruments). When it becomes difficult to generate answers in that region of the semantic space, they then reorient to the question to consider whether a different answer (e.g., flute) better fits the question, and they subsequently consider a different group (or patch) of related instruments (e.g., wind instruments). While both the random walk and foraging models predict the fast generation of semantically related answers, the foraging model uniquely involves structured, question-driven reorientation as a mechanism for transitioning between regions of semantic space.

A key distinction between these accounts concerns how multiple influences on retrieval are coordinated. Random walk models typically rely on a constant transition mechanism, in which the probability of generating a response depends on its associative proximity to the current answer. In contrast, foraging models allow different factors to guide search at different moments: local associations may dominate within a patch, whereas question relevance may guide the initiation of search in a new patch. This distinction becomes especially important in tasks where answers must be evaluated against an explicit prompt, rather than produced freely.

In this paper, we asked participants to generate sequences of responses either while answering specific trivia questions or while answering broad, semantic fluency questions (i.e., What are items from a given category?). Across both tasks, answer generation reflected common retrieval dynamics: people were more likely to produce answers that were generally accessible, and transitions were strongly shaped by similarity to previously generated responses. Additionally, the answers generated in the specific questions condition were more strongly aligned with the semantic content of that question than those generated in the semantic fluency questions condition. We also found that memory retrieval when answering a specific question is better captured by a foraging model than by a random-walk model of memory search. In the specific questions condition, answers semantically similar to prior responses were generated more quickly, whereas answers more closely aligned with the question were generated more slowly. This dissociation is difficult to reconcile with a single ongoing random walk, but follows naturally from a foraging framework in which memory search alternates between rapid local exploitation and slower, goal-directed exploration. Together, these findings suggest that while local associative processes operate similarly across tasks, specific questions engage an additional control process that intermittently redirects search toward the question itself.

Methods

Participants

A total of 329 participants (207 female; Mean age = 39.31 years) were recruited through Prolific. Participants were required to be at least 18 years old and fluent in English. They were compensated $1.20, equivalent to $12 per hour.

Procedure

Participants produced responses under one of two conditions: *specific questions condition* or *semantic fluency questions condition*. In the specific questions condition, participants generated possible answers to a series of trivia questions. In the semantic fluency questions condition, participants generated items belonging to a given category. Each question in the specific questions condition referred to a category to which the answer belonged, and these same categories were given in the semantic fluency questions condition. For example, one question was "*What musical instrument was invented to sound like a human singing?*" and the corresponding category was "*Musical instruments*." In the specific questions condition, participants were asked to generate possible answers to the question and were instructed to think about musical instruments. In the semantic fluency questions condition, participants were asked to list all musical instruments that came to mind. Each par-

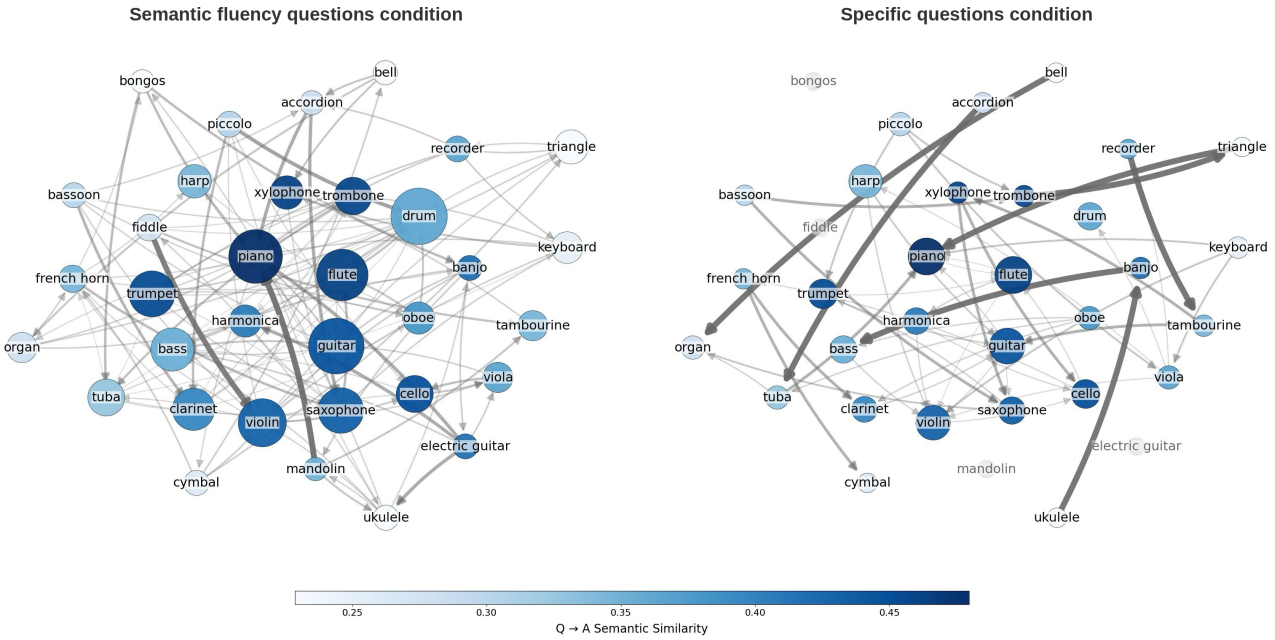

Figure 1: Answers generated in each condition are represented as networks, with nodes corresponding to specific answers and edges indicating transitions between successive answers. Node size indicates the frequency with which an answer was generated, edge width indicates the strength of transition probability, and node color indicates semantic similarity to the question. We positioned the nodes using a force-directed layout based on how semantically similar the answers were. Answers with more similar meanings were placed closer together. The networks represent the responses from all participants in the semantic fluency questions condition, in which participants were asked to list items from the category "*Musical instruments*," and in the specific questions condition, in which participants were asked to answer the question "*What musical instrument was invented to sound like a human singing?*."

ticipant completed six trials, randomly selected from a pool of 16 prompts and presented in a random order. Participants were given one minute to generate responses on each trial; the task took approximately six minutes to complete.

For each trial, participants were instructed to list all responses that came to mind while deliberating about the specific question or category. In the specific questions condition, participants were explicitly told that it was fine to generate items that might not ultimately be correct answers and that they should list any potential answers from the category to which the answer belonged (e.g., Musical Instruments). Responses were entered one at a time into a text box. Participants were free to stop generating responses at any point if they could no longer think of additional items.

Semantic Similarity Measure

To test our hypotheses, we needed measures of how closely participants' answers related to (1) the question and (2) other possible answers. We quantified these relationships using measures of semantic similarity, which capture the degree to which two texts are similar to each other in content. For example, consider the question, *"What musical instrument was invented to sound like a human singing?"*, and two answers provided by participants:

Answer 1: *Violin*

Answer 2: *Drum*

Although both responses are musical instruments, they differ in how they relate to the question. The answer *violin* is more semantically aligned with the question, as it is commonly associated with a voice-like timbre. In contrast, *drum* is generally less related to human singing, as a drum produces percussive rather than vocal-like sounds. The two answers also differ substantially from each other, in terms of how they produce sound (bowed strings versus being hit) and in their acoustic properties (sustained pitch versus rhythmic percussion). In comparison, *viola* is much more similar to *violin* because both are bowed string instruments with similar acoustic properties, whereas *snare drum* is more similar to *drum* because both are percussion instruments that produce sound by being struck. Measures of semantic similarity allow us to quantify both how strongly an answer aligns with the question and how strongly it relates to other answers in the sequence.

We computed semantic similarity using Sentence Transformers, which generate semantically meaningful vector embeddings that enable similarity comparisons between text (Reimers & Gurevych, 2019). We used the pretrained `all-MiniLM-L6-v2` model, which produces 384-dimensional embeddings and offers a balance between computational efficiency and semantic expressiveness. For each participant response and each comparison text (either the question

or another answer), we generated embeddings and computed cosine similarity between the embeddings. We also calculated semantic similarity with word embeddings and GPT-based scores. In this paper, we report results from the sentence transformers, which in a separate analysis showed the strongest inter-rater reliability with human judgments (Kedrick & Golman, 2026).

Popularity Measure

We used Google Trends data using the `pytrends` API to estimate the general accessibility or popularity of each answer. For each answer, we retrieved its search interest scores over the six months prior to the study and averaged these values to estimate how frequently the term occurs in everyday life. To validate search popularity as a proxy for how often answers are expressed in everyday life, we correlated it with word frequency from American English subtitles (Brysbaert & New, 2009), calculated as occurrences per million words. This measure captures the frequency with which these answers generally occur in naturalistic, conversational settings. We observed a positive correlation between the two measures ($\rho = .28$, $p < .001$). However, we did not adopt this alternative measure because the subtitle database covered only 1,047 of the 3,246 participant answers, whereas Google Trends covered 3,173 answers.

Results

We begin by examining whether answering specific or broad questions produce systematically different response patterns, both in the content of answers generated and in the number of responses produced. The distribution of specific answers differed strongly between the specific questions and semantic fluency questions conditions ($\chi^2(843) = 1366.2, p < .001$), indicating that the overall set of answers participants generated depended on the kind of question prompt that triggered them. We also found a robust difference in the number of items generated across conditions. Participants produced significantly more responses in the semantic fluency questions condition than in the specific questions condition ($t(1864.7) = 20.09, p < .001$). On average, participants generated more items when answering semantic fluency questions (M = 9.91) than when answering specific questions (M = 5.57), with a 95% confidence interval for the mean difference of [3.92, 4.77]. This difference is not surprising, as the set of items relevant to a broad category is substantially larger than the set of answers that appropriately satisfy the semantic constraints imposed by a specific question. See Figure 1 for a visualization of the answers generated in response to a pair of questions using semantic networks.

Answer generation predicted by question similarity, general accessibility, and similarity to prior answers.

We first examined what predicts *which* answers are generated, independent of their position in the sequence of responses. To do so, we fit conditional logit models predicting whether a given answer was mentioned, as a function of its semantic similarity to the specific question (Q–A Similarity), its overall popularity, task condition (specific questions vs. semantic fluency questions), and their interactions, with random intercepts for subject and stimulus (see Table 1).

Consistent with our hypotheses, answers that were more semantically aligned with the specific question were significantly more likely to be generated ($\beta = 0.59$, $p < .001$). This effect was moderated by task condition: question–answer similarity played a substantially stronger role in the specific questions condition than in the semantic fluency questions condition, as reflected by a significant negative interaction between similarity and the semantic fluency questions condition ($\beta = -0.22$, $p < .001$). Thus, when participants were tasked with answering specific questions, the generated answers depended on their relevance to the question. At the same time, answer generation was also influenced by general accessibility. Answers that were more popular, or more likely to occur in everyday life, were more likely to be mentioned across both tasks ($\beta = 0.24$, $p < .001$), and this effect did not differ reliably between conditions ($\beta = 0.01$, $p = .68$). This indicates that popularity influences responses in both specific- and semantic fluency questions conditions. However, specific questions also involve additional cues, such that responses are more strongly shaped by their semantic similarity with the question.

Next, we examined what predicts *how* answers unfold over time; specifically, how the generation of an answer is influenced by the prior answer. To do so, we modeled answer generation as a random walk and fit a conditional logit model, predicting whether a given answer transition occurred (transition vs. no transition) as a function of semantic similarity between the current answer and the preceding answer (A–A similarity), semantic similarity to the specific question (Q–A Similarity), task condition (specific questions vs. semantic fluency questions), and their interaction. The model included random intercepts for subject and stimulus. Table 2 reports fixed-effect estimates (log-odds scale) for the model.

Consistent with our hypothesis, local similarity between answers predicted whether an answer was generated. Across all similarity measures, higher answer–answer similarity was associated with a greater likelihood of generating an answer ($\beta = 0.73$, $p < .001$), indicating that participants were more likely to transition to related answers. This effect was slightly stronger in the semantic fluency questions condition, though the difference was small ($\beta = 0.07$, $p = .036$).

Response-time evidence supports a foraging account of question answering

To further test whether the random walk or the foraging model captures the dynamics of memory search when answering specific questions, we examined response times associated with each generated answer. Both models predict that we should observe shorter response times between semantically related

Table 1: Conditional logit model predicting probability of answer generation

Predictor	Model
Q–A similarity (scaled)	0.591^{***}
Popularity (scaled)	0.241^{***}
Q–A sim × condition	-0.220^{***}
Popularity × condition	0.007
N (choice sets)	1,910

Note. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

Table 2: Conditional logit model predicting answer transitions

Predictor	Model
A–A Similarity (scaled)	0.729^{***}
Q–A Similarity (scaled)	0.254^{***}
A–A Sim × condition	0.070^{*}
Q–A Sim × condition	-0.209^{***}
N (choice sets)	13,751

Note. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

answers. However, the random walk model predicts that answers that are semantically related to the specific question should also have shorter response times, because they too are easier to find. In contrast, the foraging model predicts that people revisit the question only when they struggle to find another answer related to their previous answer (i.e., in their current patch), so generating an answer that is more strongly aligned with the specific question should be correlated with longer response times. To test these predictions, we analyzed response times in the specific questions condition. We predicted log response time ($\log(RT)$) from answer–answer (A–A) similarity, question–answer (Q–A) similarity, and answer popularity, restricting analyses to responses beyond the first (see Table 3). All predictors were z-scored, and models included random intercepts for subject and stimulus.

Across models, A–A similarity was a negative predictor of response time: answers more similar to the preceding response were retrieved more quickly ($\beta = -0.217$, $p < .001$). In contrast, Q–A similarity was associated with slower retrieval, such that answers more closely aligned with the question took longer to generate ($\beta = 0.099$, $p < .001$). Thus, local associative transitions were fast, whereas question-directed retrieval took additional time.

The response-time results suggest that retrieval (for specific questions) alternates between fast local exploitation and slower, question-directed exploration. To test whether these shifts are influenced by retrieval difficulty, we next examined whether reliance on local versus question-directed retrieval varied as a function of the difficulty of the preceding retrieval step. If retrieval alternates between local exploitation and question-directed exploration, then moments of increased difficulty should be associated with shifts in

Table 3: Regression model predicting log response time

Predictor	Model
Intercept	1.719^{***}
Q–A sim (scaled)	0.099^{***}
A–A sim (scaled)	-0.217^{***}
Popularity (scaled)	-0.055^{***}
AIC	6,846
N (observations)	3,245
Var: subject (int.)	0.086
Var: stimulus (int.)	0.022

Note. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

Table 4: Conditional logit model of local and question-directed transitions as a function of retrieval difficulty

Predictor	Model
A–A similarity	0.722^{***}
Q–A similarity	0.332^{***}
Prev. RT (log)	-0.174
A–A sim × prev. RT	-0.076^{**}
Q–A sim × prev. RT	0.063^{**}
N (choice sets)	3,261

Note. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$.

how answers are generated. In particular, slower retrieval of one answer should reduce reliance on local associations and increase the likelihood that search is reoriented toward the question on the next step. We fit a conditional logit model predicting whether a given response reflected local continuation (measured by answer–answer similarity) or question-directed retrieval (measured by question–answer similarity), using the log-transformed response time of the previous answer as a proxy for retrieval difficulty (see Table 4). Higher answer–answer (A–A) and question–answer (Q–A) similarity each independently increased the likelihood of a transition, indicating contributions from both local associations and question alignment. These influences shifted systematically with retrieval difficulty. As prior response times increased, the effect of A–A similarity decreased ($\beta = -0.076$, $p < .01$), whereas the effect of Q–A similarity increased ($\beta = 0.063$, $p < .01$). Thus, when retrieval slowed, participants were less likely to continue exploiting locally related answers and more likely to reorient search toward the question, consistent with a foraging-style process in which local exploitation gives way to goal-directed exploration.

Together, these results show that retrieval alternates between fast local associative exploitation and slower, question-directed reorientation. This pattern is consistent with a foraging account of memory search, in which small steps (local exploitation) give way to question-directed jumps (wider exploration) when retrieval becomes more difficult (Hills et al., 2012, 2015).

Discussion

How do people search their memories to come up with answers to questions? By directly comparing memory retrieval while answering highly specific or broad category retrieval questions, we found clear evidence that these two tasks share core retrieval properties while also differing in task-specific ways. Across both conditions, answer generation was strongly shaped by general accessibility and local associative structure: more commonly encountered answers and answers semantically related to previously generated responses were more likely to be produced. These patterns are consistent with classic accounts of semantic memory retrieval that emphasize the influence of frequency and connectivity in semantic networks (Abbott et al., 2015; Anderson, 1983; Collins & Loftus, 1975). Transitions between successive responses were reliably predicted by local answer–answer similarity in both tasks, replicating prior findings from semantic fluency research (Hills et al., 2012; Troyer et al., 1997). Together, these results indicate that local associative dynamics operate robustly across both specific question answering and category exemplar retrieval.

At the same time, prompting participants with specific questions imposed an additional constraint on retrieval. Answers generated in the specific questions condition were more strongly aligned with the semantic content of the question than those generated in the semantic fluency questions condition. Importantly, this increased alignment did not replace local transition dynamics but coexisted with them, suggesting that answering specific questions integrates question relevance into an otherwise similar search process.

This integration was most clearly revealed in the temporal dynamics of retrieval. In the specific questions condition, answers semantically similar to prior responses were generated more quickly, whereas answers more closely aligned with the question were generated more slowly. This dissociation bears directly on competing accounts of memory search: the random-walk model predicts a tight coupling between retrieval probability and speed driven by local associative structure, whereas the foraging model allows retrieval timing to reflect shifts between rapid local exploitation and slower, goal-directed reorientation.

Further evidence for such reorientation comes from the finding that retrieval difficulty systematically altered transition patterns. As response times increased, reliance on local associative structure decreased, while reliance on question–answer similarity increased. When experiencing moments of increased difficulty, participants were less likely to continue exploiting locally related answers and more likely to redirect search toward the question. This shift is difficult to reconcile with a random-walk account, in which transitions are governed by a single associative mechanism. Instead, this pattern is indicative of a foraging account that allows search to alternate between local exploitation and goal-directed exploration. That is, memory search during question answering shifts between fast local associations and slower, question-directed retrieval.

One limitation of our study is that participants' written responses serve as our proxy for the items generated during the underlying search process. It is possible that not all items that came to mind were ultimately reported. For instance, a participant might briefly think of several related items in quick succession, but only record a subset of them. As a concrete example, suppose "flute" and "clarinet" both come to mind early on. By the time the participant begins typing one of these responses, they may already be considering a different item such as "oboe," and choose to report that instead because it seems like a better or more accurate answer. This raises the possibility that our task could influence which items are selected for reporting. Future work could more directly address this issue using think-aloud protocols or by asking participants to indicate whether any generated items were not reported, which would allow us to estimate the prevalence of such reporting biases.

In sum, our findings indicate that answer generation for specific questions preserves the local associative structure observed in semantic fluency question contexts, while engaging an additional control process that intermittently redirects search toward task-relevant goals. More broadly, these results contribute to the long-standing debate between random-walk and foraging accounts of memory search by showing that the distinction is not whether local associative processes operate but how and when they are overridden. The combination of fast local exploitation and slower, question-directed exploration is difficult to reconcile with a random walk through semantic space. Rather, memory search is best understood as a foraging process in which local associative dynamics provide a default mode of retrieval, while goal-directed control mechanisms guide search when it is difficult to think of a relevant association (Hart et al., 2017; Hills et al., 2012, 2015).

References

Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. *Psychological Review*, *122*(3), 558–569.

Anderson, J. R. (1983). *The architecture of cognition*. Harvard University Press.

Brysbaert, M., & New, B. (2009). Subtlexus: Word frequencies for american english subtitles [Accessed: 2025-01-13]. https://www.ugent.be/pp/experimentele-psychologie/en/research/documents/subtlexus

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Gronlund, S. D., & Shiffrin, R. M. (1988). Search for information in long-term memory: Evidence from semantic memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 523–536.

Hart, Y., Mayo, A. E., Mayo, R., Rozenkrantz, L., Tendler, A., Alon, U., & Noy, L. (2017). Creative foraging: An exper-

imental paradigm for studying exploration and discovery. *PLOS ONE*, *12*(8), e0182133. https://doi.org/10.1371/journal.pone.0182133

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

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., & Couzin, I. D. (2015). Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences*, *19*(1), 46–54.

Kedrick, K., & Golman, R. (2026). Explanation generation: Questions direct exploration in semantic space. https://osf.io/preprints/psyarxiv/shnkf_v3

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, 3982–3992. https://doi.org/10.18653/v1/D19-1410

Richie, R., Aka, A., & Bhatia, S. (2023). Free association in a neural network. *Psychological Review*, *130*(5), 1360–1382. https://doi.org/10.1037/rev0000396

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.

Zhang, Z., Wang, S., Good, M., Hristova, S., Kayser, A., & Hsu, M. (2021). Retrieval-constrained valuation: Toward prediction of open-ended decisions. *Proceedings of the National Academy of Sciences*, *118*(23), e2022685118. https://doi.org/10.1073/pnas.2022685118