Naturalistic action sampling as foraging in the option space

Alina Dracheva (Alina.Dracheva.Gr@Dartmouth.Edu)
Dartmouth College, Department of Psychological and Brain Sciences
Hanover, NH 03755 USA

Jonathan S. Phillips (Jonathan.S.Phillips@Dartmouth.Edu)
Dartmouth College, Program in Cognitive Science
Hanover, NH 03755 USA

Abstract

Human decision-making involves navigating unbounded spaces of possible goals, subgoals, and action sequences. Yet, computational models typically assume pre-defined option sets. This creates a critical gap between the algorithms developed in cognitive science research on decision-making and the open nature of real-world decisions. We propose that option generation in open-ended settings operates as a search through structured decision space. Drawing on foraging theory, we hypothesized that option generation follows Lévy flight distributions, a pattern observed in both spatial foraging and memory retrieval. We found that the inter-generation time between consecutive responses in open-ended option generation problems approximated a Lévy distribution, while semantic distances demonstrated properties of heavy-tailed distributions. These findings reveal novel connections between open-ended option generation, information search, and memory retrieval, highlighting shared computational mechanisms that underlie diverse cognitive processes.

Keywords: open-ended decision making; sampling; information search; memory retrieval; foraging; Lévy distribution; natural language processing

Introduction

Real-world decision making is a complex open-ended process in which cognitive sub-components interact dynamically. Consider finding yourself as the ring bearer at your sister's wedding, realizing ten minutes into the ceremony that you forgot the ring in your taxi. What would you do? Your response emerges from the interaction of multiple sampling processes: sampling available data (checking for the taxi company's contact information), which might prompt overt information search; sampling beliefs about the situation (assessing the importance of this specific ring and the likelihood of obtaining a replacement), which drives covert information search; and sampling actions that proved useful in similar past situations (e.g., immediately disclosing the situation). These sampling processes operate in parallel and influence each other bidirectionally.

It has been suggested that sampling could serve as an overarching theoretical framework unifying processes that have been studied independently but that together subserve decision making (Brockbank, Holdaway, Acosta-Kane, & Vul, 2023). As such, sampling data, beliefs, and actions could be jointly studied by examining the number of samples generated, their distribution patterns, and their generative process.

In this work, we propose that action sampling shares computational principles with semantic memory retrieval. Intuitively, when thinking about what we could do in an open-ended situation, we may rely on sequential retrieval of possible actions in much the same way that we do when we try to recall different kinds of animals or vegetables.

When resources are distributed in patches, efficient search strategies often involve a mixture of local exploitation and global exploration, resulting in distinctive statistical patterns of movement. Previous work on semantic memory retrieval has demonstrated that inter-item distances, operationalized as either inter-retrieval time intervals (IRIs) (Rhodes & Turvey, 2007; Hills, Jones, & Todd, 2012) or distances in the semantic space (Montez, Thompson, & Kello, 2015; Zhu, Sanborn, & Chater, 2018) approximate inverse power-law distributions associated with Lévy flights, characteristic of optimal foraging in the natural environment.

By definition, open-ended decision-making problems do not have a pre-defined set of options that could be retrieved. However, this does not mean that the space over which possible actions are sampled is unstructured; in fact, this space should reflect the structure of the real world as represented by the decision making agent, in the same way that the conceptual space reflects the patchy structure of the natural environment. If humans make optimal use of their structured representation of the real world when sampling possible actions in open-ended situations, and if this optimality requires auto-correlations between options, the distances between the options they generate should have the shape of heavy-tailed distributions characteristic of foraging in clustered environments.

In contrast, most of the research on generating possible actions in open-ended decision problems to date has treated it as a fundamentally distinct computational process. This approach has been inspired by the formal tools of reinforcement learning (RL) and models option generation as sampling from an empirical distribution of state-action pairs governed by their likelihood (Amir, Tyomkin, & Hart, 2022), or a combination of likelihood and value (Morris, Phillips, Huang, & Cushman, 2021), conditioned on the current context. Importantly, such accounts remain agnostic to potential semantic relationships between generated actions and do not address whether successive options exhibit systematic dependencies.

To test our predictions, we developed a novel experimental paradigm that elicits sequences of options in open-ended decision scenarios. By leveraging a state-of-the-art transformer-based language model (`instructor-xl`), we were able to represent each proposed action as a point in high-dimensional

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

semantic space, allowing us to compute cosine distances between consecutive responses. Our analysis focused on whether thinking times and semantic distances between consecutive options follow patterns characteristic of optimal foraging — i.e., whether they conform to heavy-tailed distributions and, more specifically, whether they exhibit properties of power-law distributions with the exponents in the range of what has been previously observed in both spatial foraging and memory retrieval.

Methods

Lévy distribution

In the real world, resources are distributed in patches, or clusters, and it has been argued that animals have evolved foraging strategies adapted to this. One way to formalize an optimal foraging strategy in a patchy environment is a Lévy distribution of traversed distances – a right-skewed heavy-tailed distribution, a special case of the broader class of power law distributions. In a Lévy distribution, the probability of a distance of length l given by:

$$P(l) \propto l^{-\mu} \tag{1}$$

where $1 < \mu < 3$. This distribution generates a series of short distances, interspersed with longer "jumps", which is advantageous in a patchy environment with sparsely and randomly distributed resources, as the probability of returning to an already visited site is smaller compared to a random walk.

It has been shown that both humans and non-human animals follow a Lévy distribution of distances when foraging for resources in a patchy natural environment (Viswanathan et al., 1996; González, Hidalgo, & Barabási, 2008; Reynolds, Ceccon, Baldauf, Karina Medeiros, & Miramontes, 2018). Beyond the search for physical resources, searching for resources in virtual environment (Garg & Kello, 2021) and visual search for cues in humans have also been shown to follow a Lévy distribution (Bella-Fernández, Suero Suñé, & Gil-Gómez De Liaño, 2022).

Similar to the natural environment, the conceptual space is patchy, and humans exploit this patchiness when retrieving category members from memory (e.g., in a semantic fluency task), generating kinds of animals in semantically related clusters. This pattern is also reflected in the time participants spend searching in memory: inter retrieval intervals (IRIs) have been found to follow this heavy-tailed distribution (Rhodes & Turvey, 2007).

Participants and Procedure

We recruited 134 native English speakers through Prolific to complete an option generation task. Each participant was presented with 3 open-ended decision-making scenarios describing everyday situations (see Table 1), drawn out of the total set of 6 such contexts, in randomized order. For each scenario, participants were given 3 minutes to generate as many different possible actions as they could. They indicated each new option by pressing the space bar, which revealed an input box for the next option. To incentivize option generation without restricting the total number of options, participants were told that they would receive additional compensation for each option generated beyond the initial seven responses.

Context	Question
1	Heinz's wife has recently fallen ill and needs an expensive medication that is not covered by her medical insurance. They don't have the money needed to purchase the expensive prescription, but they know that it's vital for her to have it if she is going to recover.
2	Josh is on the way to the airport to catch a flight to a hunting safari in Africa. He leaves with plenty of time to make it there, but his car breaks down on the highway. Now Josh is sitting in his car near a busy intersection, and knows he needs to get to the airport soon if he is going to catch his flight.

Table 1: Examples of 2 out of 6 total decision contexts.

For each context, participants generated a variety of potential actions. For example, for the scenario 2 from Table 1, one participant generated options including"Call Uber", "Call highway patrol to tow car", "Call airport to discuss options", "Hitch-hike", and "Try to walk to the airport", among others.

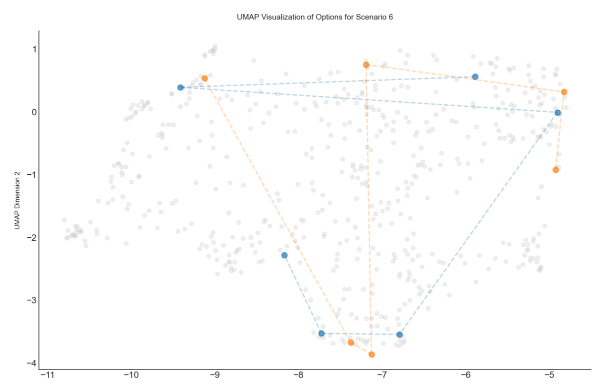

Figure 1: 2D semantic space of all generated options for scenario 4. Each dot denotes one generated option. Two participants' trajectories through the space are highlighted with orange and blue colors.

Distance measures

We used two measures of between-option distances: (1) time spent thinking of an option (in seconds), operationalized as the time between hitting a space bar, which indicated that an option came to mind, and (2) cosine distance between the embeddings of options projected into a high-dimensional semantic space. For (2), we first embedded each generated action into a high-dimensional semantic space using a transformer-based language model, fine-tuned for generating embeddings for any task (`instructor-xl`). We then reduced the dimensionality of the resulting 768-dimensional space using

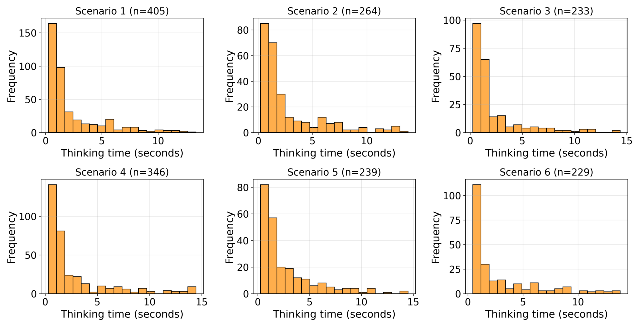

Figure 2: Distributions of time (in seconds) participants spent before reporting that an option came to mind, for each decision context.

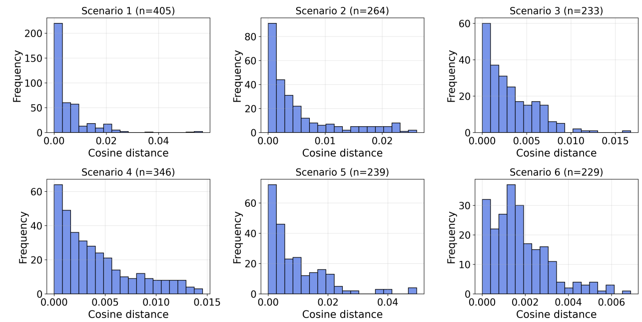

Figure 3: Distributions of cosine distances between the UMAP-compressed embeddings of consecutive options, for each scenario.

UMAP and computed cosine distances between consecutive responses for each participant in each scenario.

Data exclusion

We excluded 7 participants (5.2% of the sample) for potential AI use (detected via keystroke patterns and copy-paste events) and trimmed outliers (RTs $>$ 14.58s, 9.8% of trials) to minimize noise.

Analysis approach

Semantic distance validation We assume that a given decision context and the space of possible actions it affords lie within a subspace manifold of the full 768-dimensional representation space. Thus, to better capture the relevant conceptual differences between the options generated by participants, we applied UMAP to reduce the dimensionality of the semantic space while preserving manifold structure. UMAP parameters were set as follows: `n_neighbors` = 100, `min_dist` = 0.05, `n_components` = 20, which ensured the preservation of both global and local structure. We validated these parameters by running 10 iterations of UMAP algorithm and calculating the correlation between resulting distance matrices.

To analyze the distributions of the two types of distances and estimate power-law exponents, we used two techniques typically used for analyzing heavy-tailed distributions: logarithmic binning with normalization (Rhodes & Turvey, 2007; Zhu et al., 2018; Viswanathan et al., 1999) and direct fitting of distances using Maximum Likelihood Estimation (Clauset, Shalizi, & Newman, 2009). In LBN, distances are grouped into logarithmically-increasing bins, with geometric midpoints used for plotting the data. Power-law distributions produce a straight line on a log-log plot, and the slope of this line – the power law exponent μ – can be estimated using standard linear regression.

For the second technique (MLE), we used the powerlaw Python package (Alstott, Bullmore, & Plenz, 2014). For each scenario, the package calculates the complementary cumulative distribution functions (CCDFs) of both types of distances. Then, we fit power laws to these distributions using Maximum Likelihood Estimation (MLE), which automatically determines the scaling parameter α – our power-law exponent μ.

To validate whether these distributions followed a Lévy model, we compared the power law fits to two alternative distributions (lognormal and exponential) using likelihood ratio tests. The goodness-of-fit was assessed using Kolmogorov-Smirnov distance (D) between the empirical data and fitted distributions.

Results

Analysis of thinking times (see histograms of thinking times in Fig.2) between consecutive responses revealed strong evidence for heavy-tailed distributions characteristic of Lévy flights. The log-log plot of normalized frequencies against thinking times showed a clear linear relationship (Fig.3), indicating power law scaling. A linear relationship on a log-log plot is a defining characteristic of power law distributions because when both axes are logarithmically transformed, the power law equation $y = x^{\mu}$ becomes $\log(y) = \mu \log(x)$, which is the equation of a straight line with slope μ.

Distances as thinking times

For each scenario, we identified the optimal cutoff point for power-law fitting using a weighted maximum R^2 approach. This method systematically evaluates different starting points in the distribution and selects the one that produces the best weighted linear fit in log-log space, focusing on the tail of the distribution where power-law behavior is typically most pronounced (Viswanathan et al., 1999; Zhu et al., 2018). Using this method, we determined the optimal cutoff for each scenario, as indicated by the transition from gray to orange points in Fig. 4. The estimated power law exponents from log-binned normalization (LBN) ($\mu \in [0.98, 1.77]$) fall within the range typically associated with optimal foraging strategies ($1 < \mu < 3$), with high goodness-of-fit as indicated by R^2 values ranging from 0.971 to 1.000.

The power law fit using maximum likelihood estimation

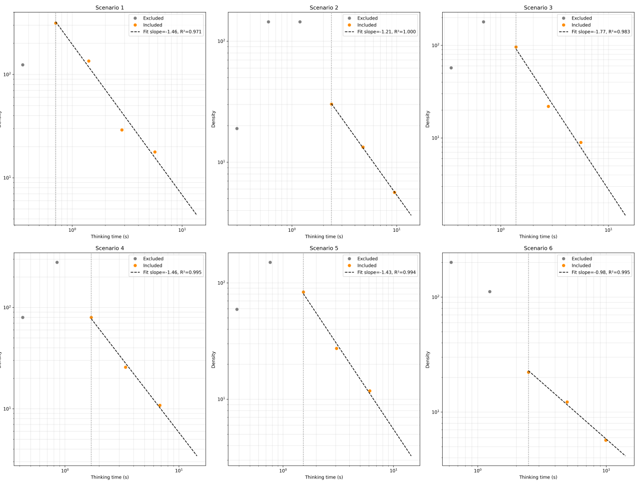

Figure 4: Log-log distributions of thinking times between consecutive options by scenario, across all participants. The best fit line is determined by the cutoff point that gives the highest R^2 value.

(MLE), with empirically determined x_{min} values for each scenario, revealed the results shown in Table 2.

Across scenarios, the exponents α generally lie within a range consistent with heavy-tailed behavior ($\alpha \in [1.82, 1.97]$). The goodness-of-fit tests yielded high p_{gof} values (0.842 to 0.954), indicating that power law distributions cannot be rejected as plausible models for our thinking time data.

Scenario	x_{min}	α	KS	p_{gof}	n_{tail}/n
1	0.652	1.932	0.068	0.950	406/482
2	0.668	1.928	0.065	0.954	300/323
3	0.822	1.957	0.081	0.842	225/300
6	0.717	1.965	0.059	0.932	386/428
11	0.725	1.859	0.096	0.858	287/321
18	0.555	1.816	0.069	0.948	299/309

Table 2: Power law fitting results for thinking times across different scenarios using maximum likelihood estimation. x_{min} is the cutoff used for the fit, α is the estimated power law exponent, KS is the Kolmogorov-Smirnov distance, p_{gof} is the p-value for goodness-of-fit from bootstrap tests, and n_{tail}/n shows the number of points in the tail distribution (where $x \geq x_{\text{min}}$) compared to the total sample size. High p_{gof} values indicate that power law distributions cannot be rejected as plausible models.

However, the negative R values and small p values ($p < 0.01$ for all scenarios except scenario 2) in the comparison with a lognormal distribution suggest that lognormal distributions provide a statistically better fit to the data. Given that distinguishing between power law and lognormal distributions through maximum likelihood estimation is notoriously difficult, often requiring extremely large sample sizes, the comparison results do not necessarily invalidate the power law interpretation, as both distributions describe heavy-tailed phenomena with similar behavioral implications. Taken together with the log-linear fits obtained through LBN (Fig. 4), these findings indicate that thinking times in our option generation task display statistical patterns consistent with heavy-tailed distributions.

Scenario	x_{min}	α	KS	p_{gof}	n_{tail}/n
1	0.0052	2.360	0.097	0.919	164/451
2	0.0034	2.098	0.115	0.880	139/301
3	0.0065	5.319	0.062	0.989	35/267
6	0.0118	11.502	0.138	0.937	23/394
11	0.0151	3.095	0.102	0.952	74/276
18	0.0024	3.637	0.082	0.963	79/265

Table 3: Power law fitting results for semantic distances across different scenarios using maximum likelihood estimation. x_{min} is the cutoff value for the fit, α is the estimated power law exponent, KS is the Kolmogorov-Smirnov distance, p_{gof} is the p-value for goodness-of-fit from bootstrap tests, and n_{tail}/n shows the number of points in the tail distribution compared to the total sample size. High p_{gof} values indicate that power law distributions cannot be rejected as plausible models for semantic distance patterns.

Distances in semantic space

Analysis of semantic distances between consecutive responses (see histograms of distances in Fig. 3) revealed heavy-tailed distributions, indicating that participants' exploration of the semantic space involved both small, local steps and occasional large jumps. While the distributions showed some characteristics reminiscent of power laws (approximately linear relationships in log-log plots at the tail end of the distribution, see Fig. 6), formal distribution fitting suggests a more complex pattern.

Maximum likelihood estimation applied to the cosine distances revealed power law exponents (α) ranging from 2.098 to 11.502 (see Table 3). These values exceed the range typically associated with Lévy flights ($1 < \alpha < 3$), suggesting steeper decay in the probability of long-distance semantic jumps compared to thinking times. Nevertheless, all scenarios demonstrated high goodness-of-fit ($p_{\text{gof}} > 0.8$), indicating that heavy-tailed distributions remain plausible models for semantic distance patterns.

Unlike the thinking time analysis, distribution fit comparisons for semantic distances yielded mostly non-significant differences ($p > 0.05$ for four out of six scenarios, with the exception of scenario 1, for which a lognormal distribution was favoured ($p < 0.01$) and scenario 2, for which an exponential distribution showed a better fit, $p < 0.01$).

Overall, the presence of heavy-tailed distributions in thinking times and semantic space exploration suggests that participants employed a mixture of local and global search strategies, alternating between exploring semantically nearby con-

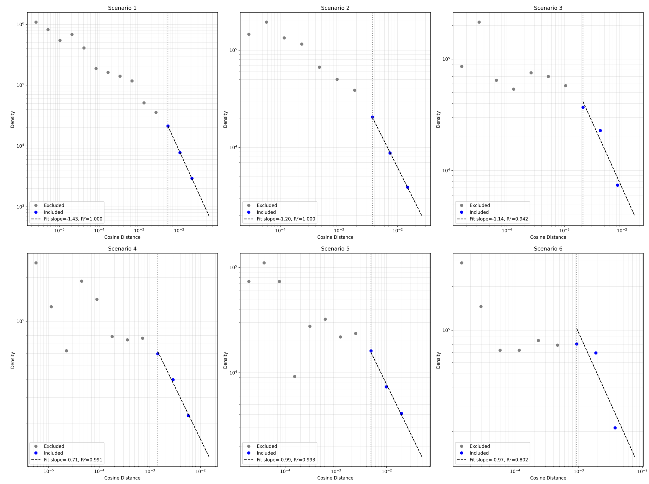

Figure 5: Log-log distributions of cosine distances between the UMAP-compressed embeddings of consecutive options by scenario, across all participants. The best fit line is determined by the cutoff point that gives the highest R^2 value.

cepts and making longer jumps to distant semantic areas. While our primary analyses rely on aggregated data, preliminary examination of individual participant data revealed qualitatively similar heavy-tailed patterns, suggesting consistency at both individual and group levels.

Relationship between semantic and temporal distances

To investigate whether increased semantic distances between consecutively generated options corresponded with longer thinking times, we initially attempted a direct log transformation of cosine distance values. However, given that these values were extremely small (mean < 0.005), a straightforward log transformation was ineffective. To address this, we conducted a systematic scaling factor analysis to determine an appropriate scaling that would allow for effective log transformation of these small distance values.

Linear mixed-effects models (LMEM) predicting log-transformed thinking times from log-transformed, scaled cosine distances were fitted across a range of scaling factors (1 to 100,000). At lower scaling factors (1 and 100), the relationship between semantic distance and thinking time remained undetectable (all adjusted $p > 0.17$). However, beginning with scaling factors of 1,000 and above, the model successfully detected a statistically significant relationship.

Specifically, a scaling factor of 10,000 provided optimal conditions for meaningful log transformation, yielding significant main effects for semantic distance ($\beta = 0.027$, $SE = 0.010$, $z = 2.64$, $p = 0.008$, adjusted $p = 0.021$) and generation number ($\beta = 0.030$, $SE = 0.004$, $z = 7.79$, $p < 0.001$). These results confirmed that longer thinking times corresponded with greater semantic distances, even after applying the Benjamini-Hochberg correction for multiple comparisons. The relationship remained significant at scale factor

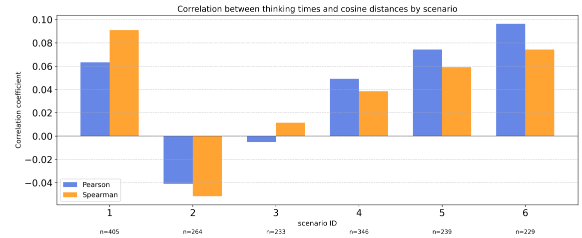

Figure 6: Correlation coefficients between cosine distances and thinking times for each context.

100,000, with adjusted p-value of 0.019. Importantly, the coefficient values stabilized at scale factor 10,000, indicating we had reached an appropriate scaling level for the statistical analysis.

To examine the relationship between thinking time and semantic space navigation at a more fine-grained level, we performed initial exploratory analyses identifying semantic clusters using the similarity drop model (Hills et al., 2012). Following this model, we identified a cluster switch as a transition between option B and C when the cosine distance between B and C is higher than the ones between A,B and C,D, given the options generated in sequence A,B,C,D. We then examined the thinking times that correspond to different positions with respect to the moment of "switch" (see Fig. 7).

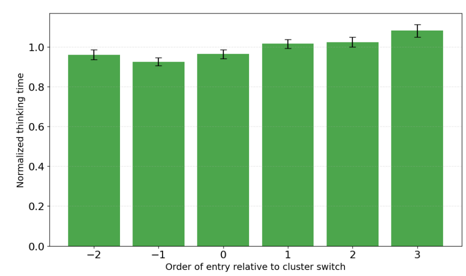

Figure 7: Thinking times around cluster transitions.

The analysis revealed that at position -1, participants showed significantly faster thinking times relative to their mean ($M = 0.856, SEM = 0.033, p < 0.001$), whereas at position 3, thinking times became significantly slower ($M = 1.158, SEM = 0.064, p = 0.0137$). Overall, this pattern suggests a gradual increase in thinking time as participants continued to sample from a new semantic cluster. One possible explanation is that thinking of new candidate actions within a given semantic region becomes increasingly hard given the relatively sparse option space, as compared to, e.g., the space of animals in semantic memory (with participants typically generating hundreds of animal names, as compared with 7-10 options per decision scenario).

Discussion

Our findings suggest that option generation in open-ended decisions may share computational principles with other forms of sampling and information search. The robust temporal patterns align with previous work on both spatial foraging and memory retrieval, suggesting that the mechanisms underlying action sampling may be optimized for exploring clustered conceptual spaces. The presence of heavy-tailed distributions in semantic distances indicates that participants likely employed a mixture of local and global search strategies, alternating between exploring semantically related options and making longer jumps to distant areas of the conceptual space.

The complex relationship between thinking times and semantic distances revealed by, on the one hand, our mixed effects models establishing correlations between thinking times and semantic distances, with bigger distances in semantic space associated with longer thinking times, and, on the other hand, a more fine-grained analysis demonstrating an increase in thinking time coupled with a deeper exploration of a semantic cluster, provides further insight into the nature of option generation. Unlike traditional semantic memory search where cluster switches correspond to longer retrieval times and within-cluster retrievals are relatively fast, our analysis of cluster transitions suggests that prolonged sampling from the same cluster in the action space might be associated with longer thinking time.

This distinction between semantic memory retrieval and action generation potentially reflects important differences between these cognitive processes. When retrieving animal names, individuals typically draw from well-established, frequently accessed categories with clear boundaries. In contrast, generating possible actions in novel situations might involve more constructive processes that integrate domain knowledge, causal reasoning, and evaluation of feasibility and consequences. The increased thinking time observed in deeper cluster exploration could reflect this constructive component of action generation, where each additional option within a semantic space requires more cognitive effort to formulate.

Our findings contribute to a growing body of evidence suggesting a domain-general cognitive search mechanism. Previous research has demonstrated that neural mechanisms initially evolved for spatial foraging have been repurposed for attention modulation and information search (Lundin et al., 2023). The present study extends this framework to action planning, suggesting that searching for possible actions in decision-making scenarios may rely on similar computational mechanisms as spatial navigation and memory retrieval.

Several limitations of our study warrant consideration. First, the small number of options generated per participant (≈ 10 per scenario) prevented robust statistical analysis at the individual level, requiring us to rely on aggregated distributions across participants. Although preliminary qualitative inspection of individual data suggested similar heavy-tailed patterns, future studies with longer generation periods or multiple sessions would be needed to confirm whether the observed group-level distributions truly reflect individual-level search processes rather than artifacts of aggregation. Second, our analysis relied on a single transformer model and dimensionality reduction technique, which may not fully capture the semantic relationships between generated options; future work should validate these findings using different embedding models and dimensionality reduction parameters. Third, the similarity drop model we adopted uses a simplistic mechanism of cluster-switching, which might be unable to account for complicated exploration behavior. Other work attempted extensions of this model, such as delta-similarity (Lundin et al., 2023), and future research should test these models using data from open-ended decision-making.

Furthermore, the relationship between option generation and option selection deserves further investigation. While our study focused on the process of generating possible actions, real-world decision-making also involves evaluating and selecting among these options. Future research could examine how the structure of the option generation process influences subsequent selection, and whether search optimality could be related to the general value of the generated options.

From a methodological perspective, our approach demonstrates how transformer-based language models can help quantify the dynamic nature of thought in open-ended decisions. More broadly, this work establishes important connections between action planning, information search, and semantic memory retrieval, suggesting that the mind may employ general-purpose sampling mechanisms for exploring complex spaces across different domains.

Conclusion

In this study, we investigated whether action sampling in open-ended decision-making exhibits patterns similar to those observed in spatial foraging and memory retrieval. Our results provide evidence that both thinking times and semantic distances between consecutively generated options follow heavy-tailed distributions characteristic of search in patchy environments. The temporal patterns, in particular, align with Lévy flight distributions associated with optimal foraging strategies, with power-law exponents falling within the theoretically predicted range. These findings suggest that the computational mechanisms underlying action sampling may share similarities with those employed in other forms of information search, supporting the idea of common cognitive search processes operating across different domains.

The observed distinctions between action generation and traditional semantic memory retrieval – particularly the increased thinking time associated with deeper exploration of semantic clusters – highlight the constructive nature of action planning in novel situations. Future research should investigate how these sampling patterns might vary across different populations and how they relate to the quality of decision outcomes.

References

Alstott, J., Bullmore, E., & Plenz, D. (2014). Powerlaw: A Python package for analysis of heavy-tailed distributions. *PLoS ONE*, *9*(1), e85777. doi: 10.1371/journal.pone.0085777

Amir, O., Tyomkin, L., & Hart, Y. (2022). Adaptive search space pruning in complex strategic problems. *PLOS Computational Biology*, *18*(8), e1010358. doi: 10.1371/journal.pcbi.1010358

Bella-Fernández, M., Suero Suñé, M., & Gil-Gómez De Liaño, B. (2022). Foraging behavior in visual search: A review of theoretical and mathematical models in humans and animals. *Psychological Research*, *86*(2), 331–349. doi: 10.1007/s00426-021-01499-1

Brockbank, E., Holdaway, C., Acosta-Kane, D., & Vul, E. (2023). Sampling data, beliefs, and actions. In J. Denrell, K. Fiedler, & P. Juslin (Eds.), *Sampling in judgment and decision making* (pp. 513–548). Cambridge: Cambridge University Press. doi: 10.1017/9781009002042.028

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661–703. doi: 10.1137/070710111

Garg, K., & Kello, C. T. (2021). Efficient lévy walks in virtual human foraging. *Scientific Reports*, *11*(1), 5242. doi: 10.1038/s41598-021-84542-w

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782. doi: 10.1038/nature06958

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

Lundin, N. B., Brown, J. W., Johns, B. T., Jones, M. N., Purcell, J. R., Hetrick, W. P., ... Todd, P. M. (2023). Neural evidence of switch processes during semantic and phonetic foraging in human memory. *Proceedings of the National Academy of Sciences*, *120*(42), e2312462120. doi: 10.1073/pnas.2312462120

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. doi: 10.1177/09567976211005702

Reynolds, A., Ceccon, E., Baldauf, C., Karina Medeiros, T., & Miramontes, O. (2018). Lévy foraging patterns of rural humans. *PLoS ONE*, *13*(6), e0199099. doi: 10.1371/journal.pone.0199099

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

Viswanathan, G. M., Afanasyev, V., Buldyrev, S. V., Murphy, E. J., Prince, P. A., & Stanley, H. E. (1996). Lévy flight search patterns of wandering albatrosses. *Nature*, *381*(6581), 413–415. doi: 10.1038/381413a0

Viswanathan, G. M., Buldyrev, S. V., Havlin, S., Da Luz, M. G. E., Raposo, E. P., & Stanley, H. E. (1999). Optimizing the success of random searches. *Nature*, *401*(6756), 911–914. doi: 10.1038/44831

Zhu, J., Sanborn, A., & Chater, N. (2018). Mental sampling in multimodal representations. In *Advances in neural information processing systems* (Vol. 31). Curran Associates, Inc.