

## **UC Merced**

### **Proceedings of the Annual Meeting of the Cognitive Science Society**

#### **Title**

Strategic search in semantic memory

#### **Permalink**

<https://escholarship.org/uc/item/1fw9h4fq>

#### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

#### **Authors**

Nikitin, Evgenii

Hills, Thomas

#### **Publication Date**

2016

Peer reviewed

# Strategic search in semantic memory

Evgenii Nikitin (E.Nikitin@warwick.ac.uk)

University of Warwick

Thomas Hills (T.T.Hills@warwick.ac.uk)

Department of Psychology, University of Warwick  
Coventry, United Kingdom, CV4 7AL

## Abstract

We search for various things every day – food, information on the Internet or someone’s name in memory. Despite the different nature of these tasks, they all have a common feature – a final goal with an unknown location in a complex environment. This property of the search raises a problem of trade-off between exploration of new opportunities and exploitation of the known information. We used the data from the semantic fluency task experiment to investigate how humans switch between exploration and exploitation strategies when they search in memory and whether they do it optimally. On comparing four different search models, the one that assumes that humans switch search strategies according to the semantic quality of the current neighbourhood best fits the data. Moreover, participants who set higher thresholds for the words with better quality of the neighbourhood tend to retrieve more words from memory. We also used regression analysis to find out which factors affect efficiency of both search strategies.

**Keywords:** Semantic memory; Memory search; Exploration and exploitation

## Introduction

Search tasks can be classified into three broad categories (Hills & Dukas, 2012): (a) external physical search (for example, animals which are foraging for food), (b) external information search (e.g., visual search), and (c) internal information search (such as search in memory). At first glance, these domains do not have much in common. However, it turns out that characteristics of these environments and strategies that individuals use are quite similar. One of the crucial properties that these strategies should have is the ability to successfully deal with the exploitation-exploration trade-off. Too much exploration can prevent a searching subject from maximizing the short-term reward, whereas redundant exploitation can lead to decreasing long-term reward.

The marginal value theorem (MVT) is formal description of optimal patch leaving rules for animals foraging in environments with well-defined patches (Charnov, 1976). It states that in a patchy system an animal should leave the patch when its food intake drops below the average long-term level expectation. The elements of this theory have been applied generally across a number of domains with patchy environmental structure, such as the search for digital information (Pirolli & Card, 1999) and, what we focus on here, search in semantic memory (Hills, Jones, & Todd, 2012).

In one test of semantic memory, the semantic fluency task (SFT), participants are given a fixed amount of time to produce as many words from a category (e.g. animals) as they can. Hills et al. (2012) suggested that semantic memory is organized as a spatial environment, where similar words are

grouped in patches. To measure the distance between words in this space, different measures of semantic similarity have been implemented. Hills et al. (2012) used the BEAGLE model (Jones & Mewhort, 2007): Here a word is represented by two vectors – a random initial vector and a memory vector. As the text corpus is being processed, every time a particular word is encountered in a corpus, its memory vector is updated by adding the sum of the random vectors of the other words that appear with it. Final similarity between two words is calculated as the dot product between their resulting memory vectors. One of the advantages of this model is that two words might be recognized as semantically similar even if they rarely occur in context with one another, but nonetheless share similar contexts (e.g., bee and wasp).

The SFT is often analysed by looking at the time between successful retrievals, namely interresponse time or IRT (Davelaar & Raaijmakers, 2012). Using BEAGLE memory representation and hand-coded categories of animals, Hills et al. (2012) showed that IRTs increase up to the long-term average IRT, when participants then leave patches. This suggests that cognition is responding strategically to the structure of semantic memory.

Another approach to search in semantic memory finds that the same retrieval patterns can be generated by a random walk in memory (Abbott, Austerweil, & Griffiths, 2015). However, this study potentially confounds the search process with the representation (Jones, Hills, & Todd, 2015), and it remains unclear whether or not people search in memory strategically or how they might respond strategically to local memory structure. People may adapt their memory search strategies according to the properties of the environment (Hills, Kalff, & Wiener, 2013). The aim of the current study is to expand research in this area by creating a comprehensive model of search that incorporates the potential influence of local memory structure.

This study does this by incorporating the concept of *proximity*, which measures how semantically similar words are to the most recent retrieval. This measure roughly describes the quality of the local neighbourhood of a word. In particular, we find support for the cognitive influence of a maximally local measure of proximity, which we call *maximum proximity*; that is, how similarity of the nearest word.

The paper is organized as follows. In Section 2, we give a full description of several competing computational models of semantic memory search. In Section 3, we describe the details of the experimental study used for analysis. Section

4 includes the results of the model fitting. In Section 5, we discuss these results and give our conclusion in Section 6.

### Model description

We begin by defining a formal model of the retrieval process. We assume that there are  $n$  words in semantic space that form the set  $W = \{w_1, w_2, \dots, w_n\}$ , and each transition from the word  $w_a$  to the word  $w_b$  can be written as  $w_a \rightarrow w_b$ . There are also  $1, \dots, i, \dots, M$  participants, and each of them has retrieved  $N_i$  words. The retrieval sequence of each participant can be described by vector  $R_i = (r_1^i, \dots, r_j^i, \dots, r_{N_i}^i)$ , where each element is an element of  $W$ . For example, vector  $R_i = (w_5, w_4, w_8)$  would mean that participant  $i$  retrieved three words – word 5, then word 4, and word 8. Finally, there are vectors  $T_i = (t_1^i, \dots, t_j^i, \dots, t_{N_i}^i)$  for each participant that contain retrieval times of each word.

### Dynamic retrieval model

To describe a stochastic process of retrieving items from memory, we use a dynamic Luce choice rule (Luce, 1959), similar to the SAM model (Raaijmakers & Shiffrin, 1981). Given a set of cues, this model allows us to calculate the probability of retrieving a particular word from the set of the possible alternatives. We used the dynamic model that implies switching between exploitation (local search, denoted by  $L$ ) and exploration (global search,  $G$ ). In this context, exploitation is an attempt to find semantically similar words in the local neighbourhood, whereas exploration is associated with "jumps" across memory that do not rely on similarity between words but on individual word frequency. Probability of retrieval of item  $j + 1$  after item  $j$  of vector  $R_i$  is calculated as:

$$P^i(r_j^i \rightarrow r_{j+1}^i | L) = \frac{F_{r_{j+1}^i}^{\alpha_i} S_{r_j^i, r_{j+1}^i}^{\beta_i}}{\sum_{w_k \notin \{r_1^i, \dots, r_j^i\}} F_{w_k}^{\alpha_i} S_{r_j^i, w_k}^{\beta_i}}, \quad w_k \notin \{r_1^i, \dots, r_j^i\} \quad (1)$$

$$P^i(r_j^i \rightarrow r_{j+1}^i | G) = \frac{F_{r_{j+1}^i}^{\gamma_i}}{\sum_{w_k \notin \{r_1^i, \dots, r_j^i\}} F_{w_k}^{\gamma_i}}, \quad w_k \notin \{r_1^i, \dots, r_j^i\} \quad (2)$$

for the local and global search cases respectively. The numerator represents retrieval strength of the word  $r_{j+1}^i$ , and cue intensities of all the not yet retrieved words are added up in the denominator.  $\alpha_i$ ,  $\beta_i$ , and  $\gamma_i$  represent the attention weights assigned to the given cue;  $F_{r_{j+1}^i}$  stands for log-frequency of the word  $r_{j+1}^i$ ;  $S_{r_j^i, r_{j+1}^i}$  is semantic similarity between words  $r_j^i$  and  $r_{j+1}^i$ .

The likelihood function is as follows:

$$Lik_i(\alpha_i, \beta_i, \gamma_i) = \prod_{j=1}^{N_i} [P_j^i(L)P(r_j^i \rightarrow r_{j+1}^i | L, \alpha_i, \beta_i) + P_j^i(G)P(r_j^i \rightarrow r_{j+1}^i | G, \gamma_i)] \quad (3)$$

Unfortunately, we are not able to directly observe which type of strategy (local or global) was used for retrieval of each

word. However, we can use the idea of the search thresholds, which is described in the next subsection, to obtain  $P_j^i(L)$  and  $P_j^i(G)$ .

### Search threshold modelling

Humans do not have to use only one of the strategies (local or global) within one retrieval attempt and do not have to make the decision about which strategy to use just after retrieval of the new word. In this study, we assume that the subject always starts by browsing in the local neighbourhood in an attempt to find a similar word in semantic space (local search). After a certain amount of time, if this attempt fails, the subject switches to the global strategy. Analogously, foraging animals do not leave a patch immediately after finding a new piece of food, but rather do it when the rate of resource intake from this patch falls below a long-term average level (or at another chosen moment).

The maximum amount of time the subject is willing to spend on the local search we call the *search threshold*. According to Marginal Value Theorem, this threshold should be set to the moment when intake from the current patch becomes equal to the long-term average retrieval rate. However, some assumptions of MVT are often not satisfied in real-world environments (McNamara & Houston, 1987). For example, MVT assumes that individuals experience a smooth, continuous flow of rewards, whereas in reality rewards are often discrete. In addition, recent work showed that switching between local and global search increases over time and this may indicate that patch quality reduces as individuals retrieve from successive patches (Hills, Todd, & Jones, 2015). Moreover, MVT ignores the idea that a forager might be able to estimate the quality of the current patch and use this information to improve his strategy. With this in mind, we compare several competing threshold models:

1. *Static (infinite threshold) model*. This model assumes that there is no switching between strategies, and both cues (similarity and frequency) are always used. Therefore, the threshold is

$$X_j^i = \infty \quad (4)$$

2. *Fixed threshold model*. According to this model, participants use constant thresholds. Participant  $i$  will switch to the global search strategy if he fails to find a similar word within  $a_i$  seconds.

$$X_j^i = a_i \quad (5)$$

3. *MVT threshold model*. Marginal Value Theorem predicts that the subject leaves the patch when the retrieval rate in the current patch drops to the average long-term rate of retrieval in the memory space as a whole. The instantaneous retrieval rate from the patch is calculated as the average cumulative reward obtained in the patch (McNamara & Houston, 1987). We can then calculate the threshold by

equalling these two rates:

$$\frac{N_i}{\sum_{k=1}^{N_i} t_k^i} = \frac{NP}{TP + X_j^i} \quad (6)$$

$$X_j^i = \frac{NP \sum_{k=1}^{N_i} t_k^i}{N_i} - TP, \quad (7)$$

where NP is the number of words retrieved from the current patch, TP is the time spent in the current patch,  $X_j^i$  is the search threshold,  $\frac{N_i}{\sum_{k=1}^{N_i} t_k^i}$  is the final long-term retrieval rate<sup>1</sup>.

4. *Maximum Proximity threshold model.* This model is based on the hypothesis that people will switch to the global strategy faster if they sense that there are no semantically similar words nearby. For instance, the words *orangutan* and *octopus* have roughly the same frequency in language. However, humans might be able to almost instantaneously understand that *orangutan* has very similar words in the semantic space (for example, other monkeys), whereas *octopus* does not. Hence, they may switch to global search faster for *octopus*. To model this hypothesis, we find the not yet retrieved word that has the maximum similarity with the current word, and use this measure (*maximum proximity*) as a factor in the model. Furthermore, we use an exponential transformation of the similarity space (Shepard, 1987).

$$X_j^i = e^{a_i + b_i \max_k S_{r_j^i, w_k}, w_k \notin \{r_1^i, \dots, r_j^i\}} \quad (8)$$

Where  $a$  and  $b$  are parameters fitting the exponential and  $\max_k S$  represents the maximum proximity. We also checked how the fit of the model changed if we used other measures of the quality of the local neighbourhood. In particular, we fit the models where we use the total similarity of 10, 20, 30, etc. nearest words as a factor in the threshold model.

Using the above models (Equations 4-8), we can then calculate  $P_j^i(L)$  and  $P_j^i(G)$  using the following rule:

$$P_j^i(L) = \begin{cases} 1, & \text{if } X_j^i \geq t_j^i \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

$$P_j^i(G) = \begin{cases} 1, & \text{if } X_j^i < t_j^i \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

That is, if the subject did not manage to retrieve a word within the threshold time  $X_j^i$ , he or she switches to the global strategy ( $G$ ). We can now rewrite our likelihood function (Equation 3). For example, for the Maximum Proximity

<sup>1</sup>We also fit models using constantly updating long-term retrieval rates, but this produced inferior models to those presented here.

threshold model, it will look like this:

$$Lik_i(\alpha_i, \beta_i, \gamma_i, a_i, b_i) = \prod_{j=1}^{N_i} [P_j^i(L|a_i, b_i)P(r_j^i \rightarrow r_{j+1}^i | L, \alpha_i, \beta_i) + P_j^i(G|a_i, b_i)P(r_j^i \rightarrow r_{j+1}^i | G, \gamma_i)] \quad (11)$$

We can now find the values of parameters that maximize this likelihood function and compare the fit of the different models.

### Search times analysis

To investigate which factors affect efficiency of local and global strategies, we used regression modelling. Each threshold model was used to classify all retrievals into one of two groups,  $L$  (local) or  $G$  (global). After categorising all retrievals, we calculated the time of local retrieval ( $TL_j^i$ ) and the time of global retrieval ( $TG_j^i$ ) for each observation:

$$TL_j^i = \begin{cases} t_j^i, & \text{if } X_j^i \geq t_j^i \\ X_j^i, & \text{otherwise} \end{cases} \quad (12)$$

$$TG_j^i = \begin{cases} t_j^i - X_j^i, & \text{if } X_j^i < t_j^i \\ NA, & \text{otherwise} \end{cases} \quad (13)$$

We then used regression modelling to find the relationships between  $TL_j^i$  and  $TG_j^i$  and different predictors, such as maximum proximity of the current word, time elapsed from the beginning of the task, etc. However, we need to remember that observations where  $X_j^i < t_j^i$ , are censored in the local case. We cannot calculate the actual time of retrieval because the strategy switch occurred before the local search was successful. For the further analysis we removed these observations from the dataset, although other approaches are also possible<sup>2</sup>. The dependent variable is positive and has a skewed distribution. Gamma GLM is often used for modelling such cases.

The following set of potential predictors was used:

- Maximum proximity – it should be easier to retrieve the word if MP of the word is higher.

$$MP_j^i = \max_k S_{r_j^i, w_k}, w_k \notin \{r_1^i, \dots, r_j^i\} \quad (14)$$

- Current elapsed time – this variable was used to control for the different time effects, such as fatigue or time pressure. We also used a squared term of this variable as its effect might be non-monotonic.

$$TT_j^i = \sum_{k=1}^j t_k^i \quad (15)$$

<sup>2</sup>For example, Cox proportional hazards model (Cox & Oakes, 1984) could be used as it can account for the censored observations. However, this model requires proportional hazard assumption to hold, and it is in general less robust than gamma GLM (Basu, Manning, & Mullahy, 2004).

- Total remaining frequency – retrievals should become slower after more frequent words are retrieved from memory

$$RF_j^i = \sum_{k=1}^n F_{w_k}, w_k \notin \{r_1^i, \dots, r_j^i\} \quad (16)$$

- Average interresponse time (IRT) – we use this measure as a proxy of each participant’s overall retrieval speed instead of adding a dummy variable for each participant to the model.

$$IRT_i = \frac{\sum_{k=1}^{N_i} t_k^i}{N_i} \quad (17)$$

- Global – a binary variable that indicates whether this retrieval was a result of the local (0) or global (1) search. We also added interactions of this variable with other predictors to the model.

Estimation results are described in the Results section.

## Method

The dataset used in this study was initially collected for Hills et al. (2012) and is available online. Participants were 141 undergraduate students at Indiana University, who participated for course credit. Participants used computers for completing the task. They were asked to type in as many animals as they could remember in three minutes. One participant was eliminated from the current analysis because he produced a word that was not encountered in the Wikipedia text corpus used for training the BEAGLE model. Moreover, participants were allowed to type the last word after three minutes had passed, and we have removed the words retrieved after the deadline from consideration. This leaves us with  $n = 354$  unique animals and a  $354 \times 354$  semantic similarity matrix. This matrix gives us a spatial representation of the semantic memory. More detailed description of the experiment can be found in Hills et al. (2012).

## Analysis

### Evaluation of models

First, using the maximum likelihood method, we estimated parameters for the four threshold models (Equations 4-8). We then calculated individual AICs (Akiake Information Criterion) for each participant. Table 1 includes AIC summed across the subjects, the number of participants for whom each model achieves the minimum AIC, and average number of retrieved words of these participants. As we can see, the maximum proximity model is the best fit, and the MVT model is the worst. Furthermore, performance of participants who are better described by the MP models is on average higher than performance of the other participants. Additionally, we included Bayesian Information Criterion (BIC) values in this table, and they support our findings.

We now consider only the Maximum Proximity model that best fit the data. This model is not completely identifiable, as

Table 1: Comparison of the threshold models

Model	AIC	$\Delta_i$	Number of people with min AIC	Number of retrieved words	BIC
Exp. maximum proximity threshold	46453.96	0.00	105	35.25	47499.21
Fixed threshold	46965.22	511.26	26	33.81	47801.42
Static model	47739.86	1285.90	9	34.22	48157.96
Marginal Value Theorem threshold	48894.71	2440.75	0	–	49521.86

for given values  $\alpha_i, \beta_i, \gamma_i$  there are many combinations of parameters  $a_i$  and  $b_i$  that lead to the same value of the likelihood function. For the analysis of the effect of the coefficients, we used median value of  $b_i$  for each participant as a point estimator. When calculating thresholds below, we used an average value of the thresholds predicted by all equally likely models.

84 participants have positive median values of the coefficient  $b_i$ , which means that they to a global strategy more quickly when the maximum proximity of the word is lower. These participants produced on average 36.18 words, whereas participants with negative median  $b_i$  retrieved on average 33.20 words, and a Wilcoxon rank sum test indicated that this difference was statistically significant ( $z = 2.22, p < .05$ ). After that, we split participants in five roughly equal groups based on the value of  $b_i$ . Figure 1 demonstrates that average performance grows as a function of  $b_i$ , reaches its maximum in  $b_i = [3; 5]$  interval and then drops again. This suggests that setting higher thresholds for words with high MP is beneficial for task performance, but only up to a certain point. This is consistent with U-shaped performance functions proposed for many tasks that rely on attentional control (Hills & Hertwig, 2011).

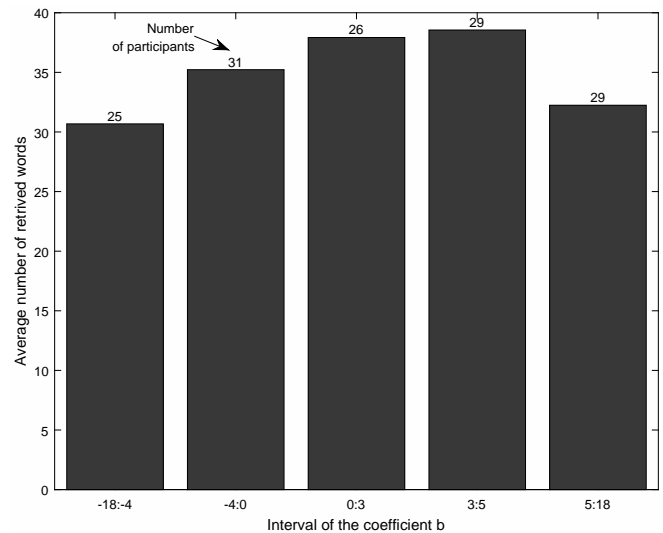


Figure 1: Relationship between number of retrieved words and the coefficient  $b_i$

Next, we analysed how the fit of this threshold model depends on the measure that is used to estimate the quality of the local neighbourhood. We varied the number of nearest items that we included in the local neighbourhood and refit the model. As Figure 2 demonstrates, the goodness of fit is

decreasing as a function of the number of the words, and the maximum proximity model that used only the nearest word fits the data the best. Sensitivity of the model to the choice of this quality measure also indirectly demonstrates that the model outperforms its competitors not just due to the greater number of the free parameters.

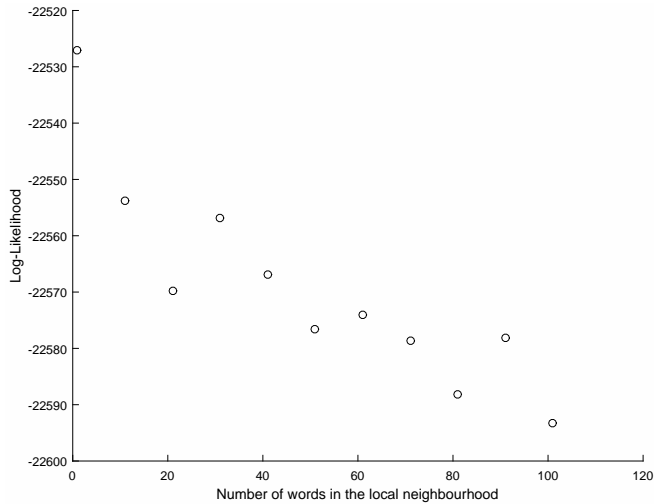


Figure 2: Connection between the quality of fit and the number of words in the local neighbourhood

Results of the GLM model estimation<sup>3</sup> are presented in Table 2. The results suggest that higher maximum proximity on average leads to faster retrievals. However, the effect of remaining frequency is important only for the global search, indicating a potential separation of frequency between local and global search. Interaction of the global search and remaining frequency variables is significant, and the magnitude of this effect is much larger than for the other variables. The effect of elapsed time is non-monotonic; we suppose that this form of the relationship might be caused by various time effects, such as fatigue and time pressure (e.g., participants might start making more effort closer to the end of the task).

## Discussion

Overall, the results of analysis support the idea that search in memory is not purely random, but rather sensitive to the structure of environment. Humans seem to be adapting their search strategy to the current state of the semantic space. We provided the following evidence to support this claim.

First, the retrieval process was best described by models with a finite threshold, which implies switching between local and global strategies. The infinite threshold model had the lowest AIC for only 6% of participants. Furthermore, the best model assumes that the time of this switching depends on the quality of the local neighbourhood, which can be approximated by a maximally local measure of proximity, MP. The MP model has the lowest AIC for 75% of participants. This

Table 2: Regression modelling

	<i>Standardized coefficients:</i>
Intercept	1.30*** (0.02)
Maximum proximity	-0.05** (0.02)
Time elapsed	0.18*** (0.05)
(Time elapsed) <sup>2</sup>	-0.13*** (0.01)
Remaining frequency	-0.04 (0.05)
Average IRT	0.18*** (0.02)
Global search	0.005 (0.03)
Global search × Maximum proximity	-0.02 (0.02)
Global search × Remaining frequency	-0.36*** (0.08)
Global search × Average IRT	0.23*** (0.03)
Global search × Time elapsed	-0.18* (0.07)
Observations	4,758
Nagelkerke R <sup>2</sup>	0.18
Log Likelihood	-10,181.250

*Standard errors are in parentheses* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

<sup>3</sup>Variables were standardized prior to the estimation.

finding supports the idea that humans use additional information about the current state of the environment to improve their retrieval process.

The poor fit of the MVT suggests that memory foraging strategies people use may involve more (or different) information than that described for the MVT. Going back to animal foraging, Iwasa, Higashi, and Yamamura (1981) noted that if additional information about the quality of the current patch (neighbourhood) is available, then simple strategies, such as MVT or constant threshold, will no longer be optimal. In our case, people might be able to estimate remaining richness of the neighbourhood in order to improve their search strategy. Similar to animals, they would be more prone to exploration in poor neighbourhoods and to exploitation in rich environments.

Next, participants who set higher thresholds for the words with greater MP tend to retrieve more words from memory. However, setting thresholds that are too high seems to hurt the performance. These results suggest that at least some participants were capable of quickly estimating quality of the semantic neighbourhood without directly accessing the words in it. The mechanism and efficiency of such estimation are both of great interest. One of the possible explanations is that people store "pre-harvest information" about quality of the different neighbourhoods and are capable of quickly accessing this information during the retrieval process (Valone, 1991).

Finally, regression analysis showed that remaining frequency of semantic space has strong effect only on the speed of the global search, whereas local search is affected only by the quality of the local neighbourhood. This finding supports the idea that these are two distinctive search strategies, and that they have the different nature.

## Conclusion

Undoubtedly, search in memory involves a fair amount of randomness. However, calling this process purely random seems to be an overstatement. In this study, we attempted to show that humans might use certain methods to improve their performance in situations when they are required to retrieve words from a semantic category, such as names, countries or animals. In particular, we might be capable of saying that some words have more similar items in memory than other words, and of using this information for optimal strategy switching. Further research is necessary to understand how accurate these estimations are and what they are based on.

## References

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

Basu, A., Manning, W. G., & Mullahy, J. (2004). Comparing alternative models: log vs cox proportional hazard? *Health Economics*, *13*, 749–65.

Charnov, E. (1976). Optimal foraging, the marginal value theorem. *Theoretical population biology*, *9*(2), 129–136.

Cox, D., & Oakes, D. (1984). *Analysis of Survival Data*. New York: Chapman & Hall.

Davelaar, E., & Raaijmakers, J. (2012). Human Memory Search. In P. Todd, T. T. Hills, & T. Robbins (Eds.), *Cognitive Search: Evolution, Algorithms, and the Brain* (chap. 11). Cambridge, MA: MIT Press.

Hills, T. T., & Dukas, R. (2012). The Evolution of Cognitive Search. In P. Todd, T. T. Hills, & T. Robbins (Eds.), *Cognitive Search: Evolution, Algorithms, and the Brain* (pp. 11–24). Cambridge, MA: MIT Press.

Hills, T. T., & Hertwig, R. (2011). Why Aren't We Smarter Already: Evolutionary Trade-Offs and Cognitive Enhancements. *Current Directions in Psychological Science*, *20*(6), 373–377.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*(2), 431–40.

Hills, T. T., Kalff, C., & Wiener, J. M. (2013). Adaptive Levy Processes and Area-Restricted Search in Human Foraging. *PLoS ONE*, *8*(4).

Hills, T. T., Todd, P. M., & Jones, M. N. (2015). Foraging in semantic fields: how we search through memory. *Topics in Cognitive Science*, *7*(3), 513–534.

Iwasa, Y., Higashi, M., & Yamamura, N. (1981). Prey Distribution as a Factor Determining the Choice of Optimal Foraging Strategy. *The American Naturalist*, *117*(5), 710–723.

Jones, M. N., Hills, T. T., & Todd, P. M. (2015). Hidden Processes in Structural Representations: A Reply to Abbott, Austerweil, and Griffiths (2015). *Psychological Review*, *122*, 570–574.

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1–37. doi: 10.1037/0033-295X.114.1.1

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York: John Wiley and Sons.

McNamara, J., & Houston, A. (1987). Foraging in patches: there's more to life than the Marginal Value Theorem. In S. S. Commons M.L. Kacelnik A. (Ed.), *Quantitative analysis of Behavior, Volume IV* (pp. 23–39). Lawrence Erlbaum Associates, Inc.

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*(4).

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.

Shepard, R. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Valone, T. J. (1991). Bayesian and prescient assessment: foraging with pre-harvest information. *Animal Behaviour*, *41*, 569–77.