Optimal Foraging in Semantic Memory

Thomas T. Hills
University of Warwick

Michael N. Jones and Peter M. Todd
Indiana University

Do humans search in memory using dynamic local-to-global search strategies similar to those that animals use to forage between patches in space? If so, do their dynamic memory search policies correspond to optimal foraging strategies seen for spatial foraging? Results from a number of fields suggest these possibilities, including the shared structure of the search problems—searching in patchy environments—and recent evidence supporting a domain-general cognitive search process. To investigate these questions directly, we asked participants to recover from memory as many animal names as they could in 3 min. Memory search was modeled over a representation of the semantic search space generated from the BEAGLE memory model of Jones and Mewhort (2007), via a search process similar to models of associative memory search (e.g., Raaijmakers & Shiffrin, 1981). We found evidence for local structure (i.e., patches) in memory search and patch depletion preceding dynamic local-to-global transitions between patches. Dynamic models also significantly outperformed nondynamic models. The timing of dynamic local-to-global transitions was consistent with optimal search policies in space, specifically the marginal value theorem (Charnov, 1976), and participants who were more consistent with this policy recalled more items.

Keywords: optimal foraging, memory search, category fluency, semantic space models, marginal value theorem

Supplemental materials: http://dx.doi.org/10.1037/a0027373.supp

Animals often search for resources that occur in spatial patches, such as the berries on separate bushes or nuts beneath a cluster of trees. Humans also search for cognitive resources that can be seen as patchy with respect to some metric other than space, such as memory representations of words grouped by semantic categories (Bousfield & Sedgewick, 1944; Raaijmakers & Shiffrin, 1981; Romney, Brewer, & Batchelder, 1993), or sets of solutions that can be navigated by working memory processes in a problem-solving task (Hills, Todd, & Goldstone, 2008; Payne, Duggan, & Neth, 2007; Wilke, Hutchinson, Todd, & Czienskowski, 2009). In spatial environments, adaptive foraging involves making appropriate global transitions between locally exploited resource clusters: decisions that prevent animals from staying too long in overexploited patches and from giving up too early on patches full of resources yet to be found (Stephens & Krebs, 1987). A classic model of optimal foraging theory (Charnov, 1976) predicts that the overall rate of return is optimized if the forager leaves a patch when the rate of finding new targets within the patch falls below the long-term average rate achieved by following the optimal strategy. We explore whether a version of this principle applies to search for items in semantic memory. Specifically, do humans move between memory patches when global opportunities outweigh local benefits, just as bumblebees forage between flower patches in an open field?

To investigate the parallel between spatial and memory search, we built models of participants' search through semantic memory when engaged in a fluency task (e.g., "name all the animals you can think of"; Lezak, 1995; Thurstone, 1938), and compared model fit to a classic model of optimal foraging in space—the marginal value theorem (Charnov, 1976). Two sources are used to represent the semantic space searched in the fluency task: hand-coded categorizations from Troyer, Moscovitch, and Winocur (1997) and lexical semantic representations from BEAGLE (bound encoding of the aggregate language environment), a corpus-based semantic space model (Jones & Mewhort, 2007). This approach allowed us to address two questions: (a) Does search in semantic memory involve switching between local exploitation of specific memory patches and global exploration between patches, and, if so, (b) is the switching between local and global search in semantic memory consistent with optimal search policies defined for animals foraging in space? In what follows we first explore the structural and neural parallels between spatial search and memory search that motivate this study. Then we develop the two questions above, before describing our data collection and modeling efforts to assess semantic foraging.

Structural and Neural Parallels Between Search in Space and Memory

Structural similarities between spatial and nonspatial environments have motivated a number of studies on human search

behavior. The key assumption underlying these investigations is that when information is distributed in clusters, or *patches* (local high-density areas of resources separated by regions with little resource availability), the optimal foraging policy of humans searching for information should share features associated with animals foraging for food in space. Research has demonstrated these parallels across spatial and cognitive search in tasks involving finding fish in virtual ponds (where patches are ponds and items are fish caught in each pond; Hutchinson, Wilke, & Todd, 2008), search for words in multiword anagrams (where patches are sets of random letters and items are words created from subsets of those letters; Hills et al., 2008; Wilke et al., 2009), and search for information on the web and in other naturalistic environments (where patches can be sets of similar web pages; Pirolli, 2007; Pirolli & Card, 1999).

Critical to the success of the forager in all these cases is the appropriate modulation between local and global search behaviors—deciding when to continue exploiting the current resource patch versus when to leave that patch and explore to find a new one. One particularly common strategy for making these ongoing trade-offs, observed in a wide range of animal species, is called *area-restricted search* in the ecological literature (Grünbaum, 1998; Karieva & Odell, 1987). This strategy involves restricting one's search to the local neighborhood for as long as resources continue to be found there and then at some point moving away from that area (sometimes gradually, and typically after the rate of finding resources falls off).

A comparative analysis of the underlying neural and molecular architectures guiding area-restricted search (Hills, 2006) gives rise to the second reason for the proposed parallel between spatial search and memory search: evidence for a generalized cognitive search process. Research from a number of fields has demonstrated that molecular and neural mechanisms that appear to have evolved initially for the purpose of area-restricted search in external environments have subsequently been exapted in later species for the purpose of modulating attention and search in internal environments (Hills, 2006). This exaptation hypothesis is supported by the observation that, across species, neural processes similar to those generally devoted to area-restricted search in space now modulate goal-directed behaviors and attention in search for information (e.g., Dulawa, Grandy, Low, Paulus, & Geyer, 1999; Floresco, Seamans, & Phillips, 1996; Hills, Brockie, & Maricq, 2004; Sawaguchi & Goldman-Rakic, 1991; Schultz, 2004), including search in human memory (Berke & Hyman, 2000; Kischka et al., 1996; Newman, Weingartner, Smallberg, & Calne, 1984; Wittmann et al., 2005). Thus, both shared environmental structure and shared mechanisms suggest the possibility of shared adaptive foraging policies for search in space and memory.

Dynamic Search in Semantic Memory

Giving up on one patch to move to another assumes that the memory search space is distributed in a patchy way, analogous to the distribution of many resources in the spatial environment. The patchiness of memory is evident in a variety of contexts including lexical decision tasks and, more importantly for our purposes, studies of free recall from natural categories, with clustered recall of related items noted in the earliest such studies (Bousfield & Sedgewick, 1944; Johnson, Johnson, & Mark, 1951). More recent

work on free recall from natural categories and list learning has consistently found that groups of semantically similar words are produced together (Bousfield & Barclay, 1950; Gruenewald & Lockhead, 1980; Howard, Jing, Addis, & Kahana, 2007; Romney et al., 1993).

This grouping of semantically similar words in recall is consistent with a cognitive foraging process that modulates between local and global memory cues, with the former producing clusters and the latter producing transitions between clusters. This dynamic search strategy is common to several different models of long-term memory retrieval (Gronlund & Shiffrin, 1986; Metcalfe & Murdock, 1981). One of the best known is the search of associative memory (SAM) model (Raaijmakers & Shiffrin, 1980, 1981). In SAM, memory is probed with cues that lead to activation and retrieval of memory items. Sets of cues make up the memory probe, which can change over the course of the retrieval period in a fashion similar to that outlined for patch-based foraging policies like area-restricted search. Initially, the probe consists of a global retrieval cue, related to the context and the category cue (i.e., the superordinate category that defines the boundaries of the search space; e.g., "animals"). Following successful item recovery, the probe is modified to include the most recently recovered item as a cue (e.g., DOG), which is a form of local information. This increases activation for items that are semantically proximal to the most recent cue (e.g., CAT). Following failures to retrieve an item, the memory probe eventually loses its local cue, and returns to its global form. This is area-restricted search in memory, dynamically moving between local and global search efforts.

The cluster–switching hypothesis is a similar but less formal model that has been investigated in the clinical literature (Troyer et al., 1997). This process involves "clustering" (the production of words in a semantic subcategory) and "switching" (making the transition from one subcategory to another; Robert et al., 1998; Troyer et al., 1997; Troyer, Moscovitch, Winocur, Leach, & Freedman, 1998). The cluster–switching model defines patches based on shared category membership (provided by hand-coded categorizations from Troyer et al., 1998), and offers a preliminary means for evaluating patch structure in memory. As we show next, this allowed us to map hand-coded categorizations directly to semantic similarity.

When and how does memory search transition from local within-patch search to global between-patch search? Recent research has investigated the algorithm-level question of what cues can lead to this transition, by modifying the standard free-recall paradigm, allowing participants to determine when to terminate memory search for items from a learned list (Dougherty & Harbison, 2007; Harbison, Dougherty, Davelaar, & Fayyad, 2009). In particular, Harbison et al.'s (2009) results suggest that when participants begin primarily to recover items that have already been retrieved, they are more likely to terminate the search process and revert to a global cue. Similar processes may drive patch switching prior to search termination (as proposed by Raaijmakers & Shiffrin, 1981). However, optimal foraging theory also focuses on a different level of description—rather than just the cues used to decide when to make a transition (the mechanistic or algorithmic level), it emphasizes the costs and benefits of deciding when to abandon a patch (the computational level)—that is, what opportunity costs are associated with staying or abandoning a given patch in memory. Thus, we focus here on asking the question of how the

memory system should make local–global transitions and whether people's search patterns are consistent with optimal foraging theory.

Optimal Foraging in Semantic Memory

In the animal foraging literature, dynamic responses to the environment are often assessed with respect to an optimal model representing a hypothesis about the trade-offs that must be negotiated in a given behavior–environment relationship. One of the first and most successful models of optimal patch foraging at this level is the marginal value theorem (Charnov, 1976). The marginal value theorem assumes that resources are distributed in patches that are monotonically depleted during foraging. The animal seeks to maximize the gain per unit time of foraging defined as the average resource intake, R, over all patches:

$$R = \frac{g(t_W)}{t_W + t_B}, \tag{1}$$

where t_W is the time spent foraging within each resource patch, t_B is the average time spent traveling between patches, and $g(t_W)$ is the cumulative gain within a patch.

Equation 1 provides a measure of resources per time unit, as a function of an individual's control over their time t_W within a patch. This is subject to patch quality, reflected by $g(t_W)$, and travel time t_B between patches. The organism is predicted to spend the optimal amount of time in a patch (t^*) such that R is maximized:

$$R^* = g'(t^*). \tag{2}$$

To maximize this resource intake, the optimal foraging policy is to leave a patch at time t^* when the instantaneous rate (or marginal value) of resource gain, $g'(t^*)$, is equal to the long-term average resource intake over the entire environment (patches and time between), R^*. In other words, the organism will switch to between-patch search when the within-patch rate (which usually starts high in a new, undepleted patch) drops to R^*. With respect to memory, the corresponding prediction is that individuals should leave the current memory patch when the benefits associated with searching further locally within it fall to the level of the expected benefits of searching elsewhere in memory. Indeed, the evidence for stopping rules in SAM based on failed retrievals (i.e., Harbison et al., 2009) suggests that patch depletion does lead to departure from local memory patches, but it is unclear whether such patch departures are consistent with optimal foraging theory. In the rest of this article, we test this prediction.

The Present Study

To test more directly the applicability of the marginal value theorem to human memory search, we had participants produce items from the category of "animals." We analyzed the search paths taken through memory in terms of the sequences of items produced. Search paths were assessed with two semantic representations. We first evaluated patch boundaries with the hand-coded subcategorization of animal terms (into specific subsets like "pets" and "water animals") derived by Troyer et al. (1997). We then compared search paths to the results of dynamic search models applied to a representation of the semantic space built by

BEAGLE, the lexical semantic memory model of Jones and Mewhort (2007). BEAGLE provides measures of semantic proximity between words based on their distributional regularities in a natural language corpus, with a level of local structural detail not possible with the nominal category-based representations of Troyer et al. Having a formal model of semantic proximity among animal names offers a quantification of the semantic search space, which we can then use to predict the retrieval of specific animals from memory and compare this with the search data we collected from people; the same approach can also be directly extrapolated to other categories (which is not the case for hand-coded subcategorization schemes).

By using both types of semantic representation, we extend prior work in memory search by making item-specific predictions, rather than merely recording number of items produced or retrieval time. Furthermore, these representations solve many of the technical difficulties previously associated with characterizing item similarity in memory (Bousfield & Sedgewick, 1944; Romney et al., 1993; for a similar approach, see Howard et al., 2007) or with using a random memory structure (Raaijmakers & Shiffrin, 1981). Semantic representations based on human coding or statistical regularities in language offer considerably more constraint to a model compared with randomly generated structures, which often allow excessive freedom for an incorrect process model to fit the data when it would have been rejected if the correct representational structure were used (Johns & Jones, 2010).

To model the search over these representational spaces, we applied a generic model of memory retrieval common to the frameworks of SAM (Raaijmakers & Shiffrin, 1981) and ACT–R (adaptive control of thought–rational; Anderson, 1993; Anderson & Lebiere, 1998). We then used various versions of these models to evaluate retrieval patterns and assess the dynamics of memory search and their correspondence with the marginal value theorem.

Method

Participants

Participants were 141 undergraduates (46 men and 95 women) at Indiana University, Bloomington, who received partial course credit. Participants were seated at a computer and followed instructions on-screen.

Procedure

Participants were asked to produce items from each of seven categories (animals, foods, vehicles, occupations, sports, cities, and movie titles), which were presented one at a time in a random order. Participants typed as many items in a given category as they could in 3 min. Entries were later corrected for spelling. Here we focus solely on the category "animals," for which we have the predetermined subcategories from Troyer et al. (1997). Some entries for this category were nonanimal items (e.g., "paw"), and these were omitted from the analyses. Together, participants produced 5,187 valid animal entries, consisting of 369 unique animal names. The mean number of animals per participant was 36.8 ($SD = 8.5$). There was no correlation between order of category appearance and number of items produced ($p = .32$).

Modeling Search in Semantic Memory

To model search in semantic memory, a structural representation of the search space is required in addition to a model of the search process. To represent the structure of semantic memory, we use both hand-coded (Troyer) and statistically derived (BEAGLE) schemes. We describe these two structural models next, followed by a description of the process model that will be applied to these structural representations.

Representing the structure of semantic memory. The Troyer et al. (1997; see also Troyer, 2000) categorization scheme contains 22 nonexclusive animal categories (e.g., "African animals," "water animals," "beasts of burden"). Support for the Troyer et al. categories comes via their usefulness in detecting specific clinical conditions in individuals, such as Alzheimer's disease, depression, and Parkinson's disease (e.g., Fossati, Le Bastard, Ergis, & Allilaire, 2003; Murphy, Rich, & Troyer, 2006; Raoux et al., 2008; Troyer et al., 1998). The categorization scheme contains 155 unique animal names, which we supplemented with 214 additional names to cover the 369 animals reported by our participants. We classified the new animals according to the original 22 categories found in Troyer et al., based on the descriptions of the additional animals found on Wikipedia. Our additions thus did not change Troyer et al.'s categorization coding scheme, so that our new investigations remain fully compatible with previous results. Our extended categorization coding is available in the supplemental materials (in Appendix 1).

To compute more fine-grained semantic similarities between words, we used the lexical semantic representations from the BEAGLE model (Jones, Kintsch, & Mewhort, 2006; Jones & Mewhort, 2007). BEAGLE representations have seen success at accounting for a variety of human semantic data including semantic typicality, categorization, and sentence completion (Jones & Mewhort, 2007), as well as for a range of semantic priming data (Jones et al., 2006). In the simulations here, we specifically used the version of BEAGLE that learns from only contextual information, similar to other high-dimensional semantic space models (e.g., Landauer & Dumais, 1997; Lund & Burgess, 1996).

The model begins by assigning each word an *initial vector* with vector elements sampled randomly from a Gaussian distribution with $\mu = 0$ and $\sigma = 1/\sqrt{D}$, where D is the arbitrary vector dimensionality (set to 1,000 in these simulations). As the text corpus is processed, each time a particular word is encountered a second vector, its *memory vector*, is updated as the sum of the initial vectors for the other words appearing in context with it. When the entire corpus has been learned, a word's memory representation is then a vector pattern reflecting the word's history of co-occurrence with other words. By this method, words that frequently co-occur will develop similar vector patterns (e.g., *bee* and *honey*), as will words that commonly occur in similar contexts, even if they never directly co-occur (e.g., *bee* and *wasp*). For all our comparisons, the similarity metric used is the vector cosine (a normalized dot-product) between two word vectors.

BEAGLE was trained on a 400-million-word Wikipedia corpus (Willits, D'Mello, Duran, & Olney, 2007), and its memory representations were used to compute the pairwise cosine similarity matrix for a list of 765 animals. The additional 396 animals that were not produced by our participants were added to the list to generate a richer memory space representing the semantic organization of the entire category of animals. In addition, it is expected that items will affect search in semantic space even if they are not produced by participants, just as berries on a bush affect foragers' external search behavior even if not consumed (e.g., by attracting the foragers to search in particular rich-looking areas of the bush). Details of the corpus preprocessing are found in the supplemental materials (in Appendix 2, as well as BEAGLE code and the animal similarity matrix).

Modeling the search process. The model framework we used to simulate the process of search is common to both the SAM and ACT–R architectures (described in Anderson, 1993; Raaijmakers & Shiffrin, 1981). The foundational assumption of our model is that recall is achieved by probing retrieval structures in memory with a specific cue set, that is, the memory probe. With I representing a possible target item for recovery in the search space, the probability of retrieving I is computed as the product of the individual retrieval strengths for I across a probe set of M cues, with $S(Q, I)$ representing the semantic similarity between cue Q and item I. This is incorporated into an overall probability of retrieval for item I via the ratio rule:

$$P(I_i|Q_1, Q_2, \ldots, Q_M) = \frac{\prod_{j=1}^{M} S(Q_j, I_i)^{\beta_j}}{\sum_{k=1}^{N} \prod_{j=1}^{M} S(Q_j, I_k)^{\beta_j}}, \quad (3)$$

where N represents the total number of items available in the category for retrieval and β represents the saliency (or attention weight) assigned to a given cue.

We examined various static and dynamic models (defined next), using either one or both of two possible cues: frequency and/or the previous item recalled. Frequency represents a *global* search cue, which generates a retrieval strength $S(Q, I)$ for each item based on that item's frequency of occurrence in the Wikipedia corpus. The previous-item cue represents a *local* search cue, which generates a retrieval strength for a new item based on its semantic similarity with that item—here the $S(Q, I)$ value is the cosine similarity in BEAGLE between the previous item generated and item I. Using the maximum likelihood method, we fit β to each participant's data, for both cue types, using the participant's individually generated sequence of items. This produced a log-likelihood fit, which was penalized based on the number of free parameters via the Bayesian information criterion. Results are presented as the median improvement in the Bayesian information criterion relative to a random model specifying that all remaining items in the search space are equally likely to be retrieved. Specific details of parameter optimization and model comparison may be found in the supplemental materials (in Appendix 3).

In our terminology, the *static models* we tested use the same memory probe (i.e., set of cues) over the entire retrieval interval, effectively ignoring the patchy structure of the environment. In contrast, *dynamic models* exploit that patchy structure, switching from patch to patch by changing the contents of the memory probe where local-to-global transitions occur. Specifically, when leaving a patch, dynamic models switch from the use of the previous-item cue (similarity-based local search) to the frequency cue (context-based global search) to find a new appropriate patch, and then back

again to the previous-item cue as the new patch is entered. For example, a sequence of DOG–CAT–HAMSTER–HORSE may transition from a local cue to a frequency cue following HAMSTER, and thereby retrieve the high frequency HORSE, which is not semantically similar to HAMSTER. We first used the Troyer et al. (1997) categorization scheme to determine where local-to-global transitions occurred in our participants' item sequences. On the basis of these results, we introduced a second patch scheme that can be extended to other environments without the use of hand coding.

Results

In this section we first check that the assumptions necessary to test our models of patchy memory search hold, and then examine three hypotheses predicted by a relationship between search in space and search in semantic memory, culminating in an evaluation of the marginal value theorem in semantic memory as a test of optimal memory foraging.

As discussed in the introduction, the marginal value theorem assumes that there is local structure in semantic memory analogous to spatial allocation of resources. If search in semantic memory is similar to search in space, retrieving an item from a specific location in memory should increase the likelihood that nearby (i.e., semantically similar) items are retrieved on subsequent trials. A recalled item should share the highest semantic similarity with the item retrieved just prior to it and share lower similarity with items retrieved further back in the sequence. Figure 1 demonstrates that this assumption holds for our data. The figure shows the data

backwards—averaging over all words—and indicates that words retrieved immediately preceding any word ("−1") were significantly more likely to be semantically similar to that word than words further away (e.g., the word two items prior, "−2"). A one-way analysis of variance predicting the similarity of words as a function of their order relative to the most recent word reveals that words are more similar to the most recent word the closer they are produced to that word, $F(4, 532) = 19.54$, $p < .001$.

The importance of local structure is also evident in the static model fits shown in the upper portion of Table 1. All models are a significant improvement over the random model with equal weightings (all participants had positive-adjusted Bayesian information criteria), and both global (frequency) and local (previous item) cues are supported as being relevant to the retrieval process. However, the best static model combines the two retrieval structures—via the integrated cue framework proposed by previous memory models (Anderson, 1993; Raaijmakers & Shiffrin, 1981). Combining local and global cues fit 100% of the participants better than either cue fit alone. This strongly supports the assumption of local memory structure in our data.

Another assumption of a model of memory search through patches in semantic space is that similarity between successively produced items will be lowest at transition points between local search and global search. Because local-to-global transition points imply that a depleted patch is being left and a new patch is being entered—with the local similarity cue being temporarily dropped from the memory probe—it follows that the semantic similarity
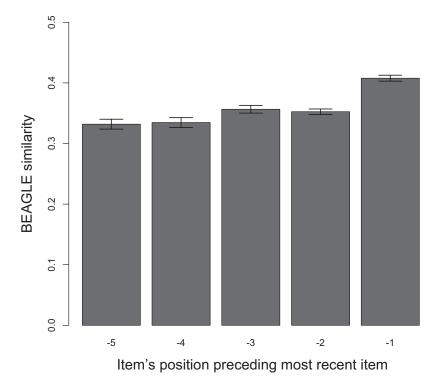

Figure 1. The BEAGLE similarity between a word and the words preceding it in the same categorical patch produced by participants. For all figures, patch transitions are computed with the Troyer et al. (1997) categorization. Words that are produced just prior to the most recent word in a patch (Position −1) are the most similar to it, with decreasing similarity for words produced earlier. Error bars are standard error of the mean.

Table 1

Bayesian Information Criterion (BIC) Comparisons (Median Improvement Relative to a Random Model) of Static and Dynamic Models Using a Combination of Frequency (Global) and Previous Item Similarity (Local) Cues, Fit to Participant Memory Retrieval Data

Model	β	BIC improvement
One cue static models		
Frequency (global)	8.47 (1.98)	75.5 (20.3)
Previous item (local)	4.34 (0.91)	70.1 (24.3)
Combined cue static model		98.6 (28.3)
Global cue	5.80 (2.09)	
Local cue	3.29 (1.10)	
Combined cue dynamic model		
Troyer et al. (1997) categories		100.12 (28.29)
Global cue	7.22 (2.18)	
Local cue	5.03 (1.67)	
Similarity drop model		104.82 (29.45)
Global cue	6.64 (2.15)	
Local cue	4.73 (1.36)	

Note. One parameter (β) was fit to each cue (global and/or local) for each participant. Standard deviations are shown in parentheses.

between two successively produced items should be lowest where local-to-global transitions occur. Figure 2 shows that this is the case. Based on the Troyer et al. (1997) norms to classify transition points, the semantic similarity between items that occur immediately before and after a transition point is substantially lower than the pairwise semantic similarities before or after this point.

This observation suggests an additional memory search model incorporating a new way of identifying local-to-global transitions, which we call the similarity drop model. This model identifies transitions by noting where similarities drop between words, in the following way: If S(A, B) represents the similarity between retrieved words A and B, then a switch following B is identified in a series of retrievals A, B, C, D if S(A, B) > S(B, C) and S(B, C) < S(C, D). The bottom of Table 1 shows that a dynamic model employing these transitions performs as well as the model with Troyer et al. category transitions. Moreover, approximately 65% of similarity drop switches were also patch switches using the Troyer et al. categories.

Hypothesis 1: A dynamic model that makes local–global transitions will outperform a static model.

Appropriate search through patchy structures implies modulating adaptively between local patch exploitation and global exploration in a dynamic fashion. The bottom portion of Table 1 shows that a dynamic memory search model based on the transition points defined by the Troyer et al. (1997) categorization—that is, a model that makes the local-to-global switches where we find Troyer-based subcategory switches in the data—accounts better for participant behavior than a static model that does not make any such transitions. This represents an improvement in 85 of the participants (results of a sign test, $p < .01$). Moreover, the similarity drop model fits 131 of the participants better than the static model (results of a sign test, $p < .001$). We show next that these transition points are appropriate and consistent with an underlying dynamic search process that shares important aspects with spatial

foraging, including patch depletion and optimal patch-leaving policies.

Hypothesis 2: Transition points occur when local semantic patches are depleted.

Semantic memory patch depletion occurs as words are retrieved, leaving fewer remaining similar words in the same patch left to be found. This implies that over time, retrieved items will have reduced similarity to all other remaining (still unretrieved) items in the semantic search space. At some point, a transition to a new patch will occur when the local patch is depleted. We call a word's semantic nearness to all other words its *residual proximity*—this is the word's retrieval strength calculated as the mean similarity (inverse distance) to all possible remaining (not yet produced) words in the overarching category (here "animals") in the BEAGLE semantic search space. Residual proximity is thus an indication of the richness of a word's remaining local neighborhood in semantic space, in terms of how distant the remaining unretrieved words are, and thus roughly how long it could take on average to continue to retrieve them.

Figure 3 displays the relationship between a word's residual proximity and its position relative to the beginning of a patch defined by the Troyer et al. (1997) categorization scheme. Residual proximity was averaged across all words that appear in a particular position with respect to any patch switch (e.g., over all words that immediately follow a patch switch, in Position 1, or are two positions before a patch switch, in Position −2). The figure clearly shows that words produced just prior to a patch switch have lower residual proximity to remaining items than items produced immediately after a patch switch. Moreover, the items produced
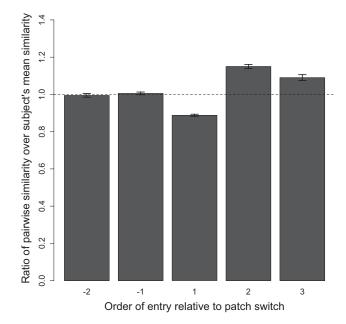

Figure 2. Mean ratio (and standard error of the mean) of pairwise similarity between successive items produced by a participant to that participant's mean pairwise similarity over all item pairs, by patch entry position. For example, the bar above "1" indicates the relative similarity between the first item in a patch and the last item in the preceding patch.
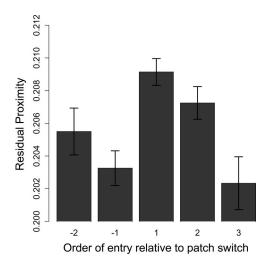
Figure 3. The residual proximity of an item in relation to an item's position before or after a patch transition. Only items not yet retrieved are included in the computation of an item's residual proximity value. Error bars are standard error of the mean.

immediately following a patch switch have the highest residual proximity, indicating that they mark the entry into a relatively undepleted patch in semantic memory. This is consistent with the prediction that transitions occur when local patches are depleted.

Hypothesis 3: Transitions occur at points predicted by the marginal value theorem.

The results above provide evidence for a dynamic memory search process that combines exploitation with exploration by transitioning between use of local and global cues. The marginal value theorem states that the optimal time for these switches should be when the current intake rate in the patch falls to the mean global intake rate for all patches (Equation 2). Does this optimal foraging policy hold for search in semantic memory? Here we take intake rate to be proportional to the inverse of the time between producing word items—that is, the interitem response times (IRTs)—and so evaluate this hypothesis in terms of the IRTs at patch switches, relative to the mean IRT across all items.

Figure 4 shows mean ratios (across participants) of item IRTs to each participant's long-term average IRT, at different retrieval positions defined relative to Troyer et al. (1997) category boundaries. The word immediately following a local-to-global transition (i.e., patch switch) takes significantly longer to produce on average than the mean IRT over the entire 3-min production interval (results of a within-participant paired *t* test), $t(140) = 13.14$, $p < .001$. The second word in a patch, however, takes significantly less time than the mean IRT, $t(140) = 11.92$, $p < .001$. Furthermore, as apparent in Figure 4 and in line with the marginal value theorem, word IRTs increase toward the patch transition point but do not exceed the long-term average IRT until the first word after the transition point (representing the first item in a new patch). The idea here is that as soon as the IRT following some word exceeds the overall mean IRT, search switches from local to global cues. The IRT from last item in the previous patch to first item in the new patch includes a longer-than-average within-patch search fol-

lowed by a global between-patch search. The same pattern of results was found for similarity drop switches. With similarity drop, the item immediately following a patch switch takes significantly longer to produce on average than the participant's mean IRT over the entire production interval ($M = 1.47$ s longer), $t(140) = -14.86$, $p < .001$, and the second item in a patch takes significantly less time, $t(140) = 12.97$, $p < .001$. Moreover, as with the Troyer et al. defined patches, as more items were produced within a given patch, the IRTs to produce those items grew longer.

To examine the optimal foraging model further, we tested the prediction from the marginal value theorem that each participant's preswitch IRTs should be at or below his or her long-term average IRT. On a per-participant basis, we assessed whether the distribution of IRTs for the single word immediately preceding a switch (Column −1 in Figure 4) was significantly different (using a one-sample *t* test) from that participant's own long-term average IRT (the IRTs for the earlier words were shorter, so were not checked individually). With the Troyer et al. (1997) defined patches, for most participants (132 of 141) the two distributions were not significantly different, and for the nine with a significant difference, all their preswitch IRT distributions were less than their long-term averages, again supporting optimal foraging in memory. With similarity drop defined patches, 24 participants had preswitch IRTs that were significantly different from their long-term average IRT, and all these preswitch IRTs were less than their long-term averages.

Finally, if the marginal value theorem is a plausible description for optimal foraging in memory—as it is in space—then individ-
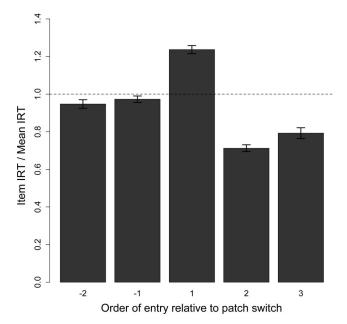

Figure 4. The mean ratio between the interitem retrieval time (IRT) for an item and the participant's mean IRT over the entire task, relative to the order of entry for the item. For example, the bar above "1" indicates the relative IRT between the first word in a patch (defined by Troyer et al., 1997, subcategories) and the last word in the preceding patch. The dotted line shows where item IRTs would be the same as the participant's mean IRT for the entire task (i.e., the inverse of the long-term average resource intake over all patches). Error bars are standard error of the mean.

uals who were more consistent with its patch departure policies should have recovered more items from memory. One indication of consistency with the theorem's patch-leaving policy is whether the last IRT in a patch (leading up to the last item) is close to the mean global IRT over all items. If that last IRT is much smaller, then it would indicate leaving the patch too soon, and if it is much greater, then it means the individual stayed in the patch too long. Using Troyer et al. (1997) defined patch boundaries, we computed the absolute difference between the mean last-item IRT across patches and the mean IRT over the entire task, and used this in a linear regression model to predict the number of items produced by each participant (see Figure 5). We found a significant negative relationship between the two, $B = -5.35$, $t(139) = -5.77$, $p < .001$. The same analysis with similarity drop defined patch boundaries found the same relationship, $B = -3.50$, $t(139) = -6.04$, $p < .001$. Greater deviation from the optimal departure time led to fewer items produced. This supports the idea that individuals who leave memory patches too early or too late will retrieve fewer items from memory than those who follow a policy more consistent with the marginal value theorem.

Discussion

Semantic memory search appears to be similar to search in physical space, involving a dynamic process of mediating between local exploitation and global exploration of clusters of information in much the same way that animals forage among patches of food in their environment. A dynamic process has been postulated before for semantic memory search (Raaijmakers & Shiffrin, 1981; Troyer et al., 1997), but had only been tested by predicting interitem retrieval times via a synthetic "random" search environment. By predicting retrieval patterns of memory items searched for over a structured representation of semantic space in patches, the work presented here shows how the dynamic local-to-global search process extends to patchy semantic space in a way that parallels optimal foraging search in physical space. In particular, we found evidence for local (i.e., patchy) structure in memory, patch depletion preceding patch departures, and optimal timing of patch departures—with participants who more closely adhered to optimal foraging theory (i.e., the marginal value theorem) producing more items.

Two common underlying factors that motivated our comparison between foraging in space and in memory were the shared patchy structure in both types of search environments and the shared mechanisms that may underlie search processes across environments. Shared patchy structure is seen in the clustering of animal- and plant-based food resources (Bell, 1991; Taylor, Woiwod, & Perry, 1978), in the clustering of items recalled from memory (e.g., Bousfield & Barclay, 1950) and in the small-world network structure of word co-occurrences in text (e.g., Ferrer i Cancho & Solé, 2001). Shared underlying processes are supported by shared neural correlates of search (Hills, 2006) and the ability to prime search from spatial to semantic domains (Hills et al., 2008), which has given rise to the theoretical notion that executive cognition is a domain-general search process (Hills, Todd, & Goldstone, 2010; see also Rhodes & Turvey, 2007). Our results here, demonstrating that memory search functions similarly to spatial search with regard to the marginal value theorem, adds an important component to this argument.
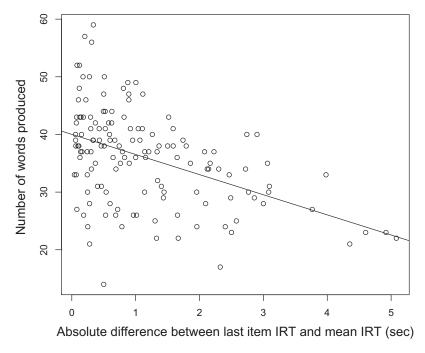

Figure 5. The relationship between a participant's deviation from the marginal value theorem policy for patch departures (*x*-axis) and his or her total number of words produced, showing lower performance with less consistency with the optimal foraging rule. Each circle corresponds to one participant; line is the best fitting linear regression. IRT = interitem retrieval time.

Together, this evidence supports a theory of semantic memory search in which individuals search locally through memory along a meandering similarity-based path until the difficulty of finding a new item nearby (as measured by the time it takes to retrieve it) rises to the average difficulty of finding items over the entire search domain, at which point local search is abandoned and a global search is made for a new patch. Our results indicate that the last item retrieved from a patch is relatively distant (as measured by residual proximity) from what remains to be recovered in long-term memory, and the subsequent transition to a global cue removes this local constraint. Doing this at the optimal time appears to improve memory production and may help explain why individuals from different clinical populations produce different numbers of items (e.g., Raoux et al., 2008).

Search environments run a wide gamut. They include visual search (e.g., Najemnik & Geisler, 2005), finding optimal paths on a map (Fu & Gray, 2006), searching for mathematical solutions (Hills, 2010), searching for web pages (Fu & Pirolli, 2007), seeking and recalling contacts in social networks (Adamic & Adar, 2005; Hills & Pachur, 2012), finding mates spread out over time (Todd & Miller, 1999), and searching in literal space. Given the generality of the search control problem—that is, mediating between exploration and exploitation of resources in patchy environments—the computational approaches invoked in these various domains are likely to provide many future cross-disciplinary insights into the nature of underlying search policies and mechanisms (see, e.g., Todd, Hills, & Robbins, in press).

References

Adamic, L., & Adar, E. (2005). How to search a social network. *Social Networks, 27,* 187–203. doi:10.1016/j.socnet.2005.01.007

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Lebiere, C. (1998). *Atomic components of thought.* Hillsdale, NJ: Erlbaum.

Bell, W. J. (1991). *Searching behaviour: The behavioural ecology of finding resources.* New York, NY: Chapman and Hall.

Berke, J. D., & Hyman, S. E. (2000). Addiction, dopamine, and the molecular mechanisms of memory. *Neuron, 25,* 515–532. doi:10.1016/S0896-6273(00)81056-9

Bousfield, W. A., & Barclay, W. D. (1950). The relationship between order and frequency of occurrence of restricted associative responses. *Journal of Experimental Psychology, 40,* 643–647. doi:10.1037/h0059019

Bousfield, W. A., & Sedgewick, C. H. W. (1944). An analysis of sequences of restricted associative responses. *Journal of General Psychology, 30,* 149–165.

Charnov, E. L. (1976). Optimal foraging: The marginal value theorem. *Theoretical Population Biology, 9,* 129–136. doi:10.1016/0040-5809(76)90040-X

Dougherty, M. R., & Harbison, J. I. (2007). Motivated to retrieve: How often are you willing to go back to the well when the well is dry? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 1108–1117. doi:10.1037/0278-7393.33.6.1108

Dulawa, S. C., Grandy, D. K., Low, M. J., Paulus, M. P., & Geyer, M. A. (1999). Dopamine D4 receptor-knock-out mice exhibit reduced exploration of novel stimuli. *Journal of Neuroscience, 19,* 9550–9556.

Ferrer-i-Cancho, R., & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society B: Biological Sciences, 268,* 2261–2265. doi:10.1098/rspb.2001.1800

Floresco, S. B., Seamans, J. K., & Phillips, A. G. (1996). A selective role of dopamine in the nucleus accumbens of the rat in random foraging but not delayed spatial win-shift-based foraging. *Behavioural Brain Research, 80,* 161–168. doi:10.1016/0166-4328(96)00031-9

Fossati, P., Le Bastard, G., Ergis, A.-M., & Allilaire, J.-F. (2003). Qualitative analysis of verbal fluency in depression. *Psychiatry Research, 117,* 17–24. doi:10.1016/S0165-1781(02)00300-1

Fu, W.-T., & Gray, W. D. (2006). Suboptimal tradeoffs in information seeking. *Cognitive Psychology, 52,* 195–242. doi:10.1016/j.cogpsych.2005.08.002

Fu, W.-T., & Pirolli, P. (2007). SNIF-ACT: A cognitive model of user navigation on the world wide web. *Human–Computer Interaction, 22,* 355–412.

Gronlund, S. D., & Shiffrin, R. M. (1986). Retrieval strategies in recall of natural categories and categorized lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 550–561. doi:10.1037/0278-7393.12.4.550

Gruenewald, P. J., & Lockhead, G. R. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 225–240. doi:10.1037/0278-7393.6.3.225

Grünbaum, D. (1998). Using spatially explicit models to characterize foraging performance in heterogeneous landscapes. *American Naturalist, 151,* 97–113. doi:10.1086/286105

Harbison, J. I., Dougherty, M. R., Davelaar, E. J., & Fayyad, B. (2009). On the lawfulness of the decision to terminate memory search. *Cognition, 111,* 397–402. doi:10.1016/j.cognition.2009.03.002

Hills, T. T. (2006). Animal foraging and the evolution of goal-directed cognition. *Cognitive Science, 30,* 3–41. doi:10.1207/s15516709cog0000_50

Hills, T. (2010). Investigating mathematical search behavior using network analysis. In R. Lesh, P. L. Galbraith, C. R. Haines, & A. Hurford (Eds.), *Modeling students' mathematical competencies* (pp. 571–581). New York, NY: Springer.

Hills, T., Brockie, P. J., & Maricq, A. V. (2004). Dopamine and glutamate control area-restricted search behavior in *Caenorhabditis elegans. Journal of Neuroscience, 24,* 1217–1225. doi:10.1523/JNEUROSCI.1569-03.2004

Hills, T. T., & Pachur, T. (2012). Dynamic search and working memory in social recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 218–228. doi:10.1037/a0025161

Hills, T. T., Todd, P. M., & Goldstone, R. L. (2008). Search in external and internal spaces: Evidence for generalized cognitive search processes. *Psychological Science, 19,* 802–808. doi:10.1111/j.1467-9280.2008.02160.x

Hills, T. T., Todd, P. M., & Goldstone, R. G. (2010). The central executive as a search process: Priming exploration and exploitation across domains. *Journal of Experimental Psychology: General, 139,* 590–609. doi:10.1037/a0020666

Howard, M. W., Jing, B., Addis, K. M., & Kahana, M. J. (2007). Semantic structure and episodic memory. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 121–141). Mahwah, NJ: Erlbaum.

Hutchinson, J. M. C., Wilke, A., & Todd, P. M. (2008). Patch leaving in humans: Can a generalist adapt its rules to dispersal of items across patches? *Animal Behavior, 75,* 1331–1349. doi:10.1016/j.anbehav.2007.09.006

Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review, 17,* 662–672. doi:10.3758/PBR.17.5.662

Johnson, D. M., Johnson, R. C., & Mark, A. L. (1951). A mathematical analysis of verbal fluency. *Journal of General Psychology, 44,* 121–128. doi:10.1080/00221309.1951.9711240

Jones, M. N., Kintsch, W., & Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language, 55,* 534–552. doi:10.1016/j.jml.2006.07.003

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and

order information in a composite holographic lexicon. *Psychological Review, 114,* 1–37. doi:10.1037/0033-295X.114.1.1

Karieva, P., & Odell, G. (1987). Swarms of predators exhibit "preytaxis" if individual predators use area-restricted search. *American Naturalist, 130,* 233–270. doi:10.1086/284707

Kischka, U., Kammer, T. H., Maier, S., Weisbrod, M., Thimm, M., & Spitzer, M. (1996). Dopaminergic modulation of semantic network activation. *Neuropsychologia, 34,* 1107–1113. doi:10.1016/0028-3932(96)00024-3

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240. doi:10.1037/0033-295X.104.2.211

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York, NY: Oxford University Press.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavioral Research Methods, Instruments, & Computers, 28,* 203–208. doi:10.3758/BF03204766

Metcalfe, J., & Murdock, B. B. (1981). An encoding and retrieval model of single-trial free recall. *Journal of Verbal Learning and Verbal Behavior, 20,* 161–189. doi:10.1016/S0022-5371(81)90365-0

Murphy, K. J., Rich, J. B., & Troyer, A. K. (2006). Verbal fluency patterns in amnestic mild cognitive impairment are characteristic of Alzheimer's type dementia. *Journal of the International Neuropsychological Society, 12,* 570–574. doi:10.1017/S1355617706060590

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434,* 387–391. doi:10.1038/nature03390

Newman, R. P., Weingartner, H., Smallberg, S. A., & Calne, D. B. (1984). Effortful and automatic memory: Effects of dopamine. *Neurology, 34,* 805–807.

Payne, S. J., Duggan, G. B., & Neth, H. (2007). Discretionary task interleaving: Heuristics for time allocation in cognitive foraging. *Journal of Experimental Psychology: General, 136,* 370–388. doi:10.1037/0096-3445.136.3.370

Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information.* New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195173321.001.0001

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review, 106,* 643–675. doi:10.1037/0033-295X.106.4.643

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 207–262). New York, NY: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134. doi:10.1037/0033-295X.88.2.93

Raoux, N., Amieva, H., Le Goff, M., Auriacombe, S., Carcaillon, L., Letenneur, L., & Dartigues, J.-F. (2008). Clustering and switching processes in semantic verbal fluency in the course of Alzheimer's disease participants: Results from the PAQUID longitudinal study. *Cortex, 44,* 1188–1196. doi:10.1016/j.cortex.2007.08.019

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

Robert, P. H., Lafont, V., Medecin, I., Berthet, L., Thauby, S., Baudu, C., & Darcourt, G. (1998). Clustering and switching strategies in verbal fluency tasks: Comparison between schizophrenics and healthy adults. *Journal of International Neuropsychological Society, 4,* 539–546. doi:10.1017/S1355617798466025

Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science, 4,* 28–34. doi:10.1111/j.1467-9280.1993.tb00552.x

Sawaguchi, T., & Goldman-Rakic, P. S. (1991). D1 dopamine receptors in prefrontal cortex: Involvement in working memory. *Science, 251,* 947–950. doi:10.1126/science.1825731

Schultz, W. (2004). Neural coding of basic reward terms of animal learning, game theory, microeconomics and behavioural ecology. *Current Opinion in Neurobiology, 14,* 139–147. doi:10.1016/j.conb.2004.03.017

Stephens, D. W., & Krebs, J. R. (1987). *Foraging theory.* Princeton, NJ: Princeton University Press.

Taylor, L. R., Woiwod, I. P., & Perry, J. N. (1978). The density-dependence of spatial behaviour and the rarity of randomness. *Journal of Animal Ecology, 47,* 383–406. doi:10.2307/3790

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago, IL: University of Chicago Press.

Todd, P. M., Hills, T. T., & Robbins, T. W. (Eds.). (in press). *Cognitive search: Evolution, algorithms, and the brain.* Cambridge, MA: MIT Press.

Todd, P. M., & Miller, G. F. (1999). From pride and prejudice to persuasion: Satisficing in mate search. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 287–308). New York, NY: Oxford University Press.

Troyer, A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychology, 22,* 370–378. doi:10.1076/1380-3395(200006)22:3;1-V;FT370

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,* 138–146. doi:10.1037/0894-4105.11.1.138

Troyer, A. K., Moscovitch, M., Winocur, G., Leach, L., & Freedman, M. (1998). Clustering and switching on verbal fluency tests in Alzheimer's and Parkinson's disease. *Journal of the International Neuropsychological Society, 4,* 137–143. doi:10.1017/S1355617798001374

Wilke, A., Hutchinson, J. M. C., Todd, P. M., & Czienskowski, U. (2009). Fishing for the right words: Decision rules for human foraging behavior in internal search tasks. *Cognitive Science, 33,* 497–529. doi:10.1111/j.1551-6709.2009.01020.x

Willits, J. A., D'Mello, S. K., Duran, N. D., & Olney, A. (2007). Distributional statistics and thematic role relationships. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 707–712). Austin, TX: Cognitive Science Society.

Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H.-J., & Düzel, E. (2005). Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron, 45,* 459–467. doi:10.1016/j.neuron.2005.01.010