Cognition 128 (2013) 64–75

Contents lists available at SciVerse ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

# Multiply-constrained semantic search in the Remote Associates Test

Kevin A. Smith *, David E. Huber, Edward Vul

*Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0109, United States*

ARTICLE INFO

*Article history:*
Received 30 May 2012
Revised 10 January 2013
Accepted 7 March 2013

*Keywords:*
Multiply-constrained problems
Semantic retrieval
Semantic search
Problem solving
Remote Associates Test

ABSTRACT

Many important problems require consideration of multiple constraints, such as choosing a job based on salary, location, and responsibilities. We used the Remote Associates Test to study how people solve such multiply-constrained problems by asking participants to make guesses as they came to mind. We evaluated how people generated these guesses by using Latent Semantic Analysis to measure the similarity between the guesses, cues, and answers. We found that people use two systematic strategies to solve multiply-constrained problems: (a) people produce guesses primarily on the basis of just one of the three cues at a time; and (b) people adopt a local search strategy—they make new guesses based in part on their previous guesses. These results inform how people combine constraints to search through and retrieve semantic information from memory.



## 1. Introduction

### 1.1. Multiply-constrained problems

Imagine you are planning a vacation with three finicky friends. Sam wants to relax on a beach. Pat lost her passport and must stay in the United States. Alex, an amateur volcanologist, wants to visit volcanoes. What destination would satisfy everyone? People figure out that Hawaii is good choice, and regularly solve similar problems with relative ease. They combine disparate constraints to plan the best route home based on road, weather, and traffic conditions; or to prioritize work based on demands of bosses, available resources, and dependencies from other projects. These problems are all 'multiply-constrained': many alternatives satisfy one constraint in isolation, but the small number of acceptable solutions can only be found via all constraints.

Multiply-constrained problems have two key features: first, each of the constraints defines qualitatively different and mutually uninformative objectives, and second, there is no common currency by which to make a principled tradeoff between criteria. The first feature differentiates multiply-constrained problems from probabilistic cue combination (Ernst & Banks, 2002; Ernst & Bulthoff, 2004; Hillis, Watt, Landy, & Banks, 2004). In probabilistic cue combination each datum provides uncertain information about the same latent variable and combining the data increases certainty; for example, obtaining a more accurate estimate of the height of a ridge by combining tactile and stereoscopic percepts (Ernst & Banks, 2002). In contrast, the constraints in multiply-constrained problems provide different types of information: in the prior example, a location's distance to the beach has no bearing on its proximity to a volcano. The second criteria captures the fact that there is no information within the problem about how to weight the constraints: one cannot judge whether a location closer to the beach but further from a volcano is preferable to one with the opposite tradeoff. Thus it is possible to have multiple acceptable answers depending on how individuals decide to weight the constraints.

* Corresponding author. Address: Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0109, United States. Tel.: +1 858 534 4468.

*E-mail address:* k2smith@ucsd.edu (K.A. Smith).

0010-0277/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.03.001

Although people often solve these problems effortlessly, this apparent ease hides the computational difficulty of the task. The space of possible answers to such a problem is usually enormous (e.g., all possible vacation spots on Earth), and an exhaustive search of all possible answers is impractical. Instead, people direct their search to promising alternatives; but how? Many theories of multiply-constrained problem solving propose a two-stage process: first people *search* for a potential answer, then they *test* this candidate against all of the constraints to rate its acceptability. If the answer is considered acceptable, people will use it as a solution; otherwise, they will search for and test another potential answer. This search-test process has been proposed as a mechanism for many cognitive tasks such as hypothesis generation (Thomas, Dougherty, Sprenger, & Harbison, 2008), analogy (Forbus, Gentner, & Law, 1995), or solving word problems (Gupta, Jang, Mednick, & Huber, 2012).

In this paper we focus on the search process – how do people come up with candidate answers. Although this test process is required to identify when the search process outputs a solution, the search process can be studied separately under the assumption that, in general, people are able to recognize a good answer when it is provided (i.e., the test process does not vary greatly across different problems). We studied the search process by obtaining a sequence of guesses as people attempted to solve a multiply-constrained problem. Prior studies have typically not studied this process as it unfolds; instead they have fit models based on a single (final) answer for each problem. We hope to gain further traction on the issue by examining the search process in an 'online' fashion, under the assumption that a sequence of guesses is a subset of proposals from the true underlying search process.

We partitioned the space of human search strategies in multiply-constrained problems along two dimensions. First, how do people use the constraints to limit the pool of candidate answers? Second, how do people search through these potential answers? Here we address these questions in a novel Remote Associates Test (RAT; Mednick, 1962) paradigm by collecting sequences of responses and quantitatively evaluating the search strategies people use to explore candidate answers.

### 1.2. Search in the Remote Associates Test (RAT)

The goal in RAT problems is to find one word that is associated with three cues (e.g., cues: 'moon' 'dew' 'comb'; answer: 'honey'). This task illustrates key features of multiply-constrained problems: each cue indicates a different aspect of the target word ('honeymoon' relies on a different meaning of 'honey' than 'honeycomb'), and there is no principled way to trade off association to each of the three cues. Moreover, RAT problems provide a controlled environment for studying how people solve multiply-constrained problems: all constraints are of the same type (word-word relationships), and unlike many naturalistic multiply-constrained problems, RAT problems are designed to have a unique best solution.

Not only is the Remote Associates Test a controlled multiply-constrained problem, but it is also correlated with real-world problem solving ability and creativity (Mednick, 1962), so elucidating human search strategies in the RAT can inform what drives these individual differences. Moreover, RAT performance is used to measure manipulations related to creativity, such as incubation (Vul & Pashler, 2007), affect (Isen, Daubman, & Nowicki, 1987), sleep (Cai, Mednick, Harrison, Kanady, & Mednick, 2009), and performance assessment (Harkins, 2006). Although these manipulations affect RAT solution rates, the mechanisms they impact remain unknown, so characterizing search strategies in the RAT might inform how these interventions improve creativity and problem-solving.

We next review previous attempts to specify the search process employed while taking the RAT; however, we note that these studies only considered a single final answer, rather than collecting intermediate responses during the search process. Spreading activation accounts (Collins & Loftus, 1975) of the RAT proposed that the cues activate their close associates and thus jointly activate the answer, making it more likely to be produced (Bolte, Goschke, & Kuhl, 2003; Topolinski & Strack, 2008). However, these accounts did not specify the weighting scheme for the cues, the quantitative definition of 'close associate', or the process for choosing amongst equally activated words. Gupta et al. (2012) provided evidence that the search process is affected by the frequency of candidate answers, although their model assumed an equal weighting of the cues rather than testing whether this was the case. Supporting the claim that the cues are not equally weighted, Harkins (2006) found that if the answer to a RAT problem comes to mind easily when prompted by just one of the three cue words, that problem is easier to solve. However, it is possible that these easily answered RAT problems were different in other ways—for instance, the answer to these problems may have been more strongly associated with the other cues as well. Although these studies yield promising clues about how people search for an answer in the RAT, they do not fully specify the weighting scheme for the cues, and, more importantly, do not investigate dynamic changes in the weighting scheme as the search process unfolds.

In this study, we investigated how the cues act as constraints on the words produced by the search process. The number of words related to at least one of the cues is a truly vast set of words, and an unordered exhaustive search of this set would take considerable time. Instead, we suggest that the search process samples words probabilistically, such that the constraints impact the probability that a given word is considered as a potential answer. Thus we want to know how the cues combine to impact this probability: is it the case that cues act multiplicatively, meaning that candidate answers are likely to be considered only if they are related to all three of the cues, or do the cues act additively, such that a word need only be strongly related to a single cue to be considered? To explore these questions within a probabilistic sampling framework, we considered a range of stochastic search algorithms that people could be using (Russell & Norvig, 2003). *Global search* algorithms explore the search space with no sequential dependencies, such that each word is randomly and independently selected from the same set of possibilities. In contrast, *local search* algorithms explore the space via a sequentially dependent chain, where each word is selected from the neighborhood of the previously considered word.

Using tasks as varied as free recall (Howard & Kahana, 2002), production of category members (Gruenewald & Lockhead, 1980), solving anagrams (Hills, Todd, & Goldstone, 2010), and naming personal social relations (Hills & Pachur, 2012), prior research finds close similarity relationships between responses that are adjacent within the sequence of responses. But these tasks all require retrieval based on a single constraint (e.g., naming animals) for which close associates of a response will also tend to satisfy the constraints of the problem; close associates of 'dog' might include 'cat' and 'wolf', which also satisfy the single constraint of naming animals. However, in the multiply-constrained vacation example, locations similar to beach resorts may satisfy the 'beach' requirement, but will not necessarily satisfy the 'volcano' constraint. Thus, it is not clear whether local search is an effective search strategy for multiply-constrained problems. As such, it is important to test whether the finding of local search for singly-constrained problems generalizes to multiply-constrained problems.

There are different search strategies that can give rise to local search behavior (i.e., sequential dependencies) and, beyond determining if local search is used in multiply-constrained problems, we seek to differentiate between these alternatives. The simplest form of local search involves the use of each response as an additional constraint that affects the choice of the next response, resulting in a relatively smooth progression. However, some studies have found evidence of a different kind of local dependency based on 'clustering': the tendency in free association for people to rapidly and sequentially name instances of a subcategory from the task (e.g., naming a series of farm animals as a subcategory of all animals) before discretely switching to another subcategory after a pause (Bousfield, 1953; Graesser & Mandler, 1978). This has led some to suggest that local dependence between responses is due to a two-stage process, in which people select a subcategory and then list items that are part of that group (Gruenewald & Lockhead, 1980). We therefore investigated whether there was evidence of direct dependence or clustering as people search for an answer to RAT problems.

Using classic RAT tasks that only collect a single response, it is difficult to distinguish between global search and local search strategies, or to investigate the process by which local search might arise. To gain data capable of differentiating between search algorithms, we used a new experimental method that elicits sequences of response from participants and analyzed these responses with a technique for measuring the semantic relationships between responses. We asked participants to respond with each word they considered as it came to mind while searching for a RAT solution, and we measured the semantic similarity between each of these responses and the cues, answer, and other responses using Latent Semantic Analysis (LSA; Landauer & Dumais, 1997). In doing so, we do not assume that these responses faithfully represent every single candidate answer considered by our participants. However, if these responses reflect a subset of the candidate answers, and if these responses are ordered in the same manner as the full set of candidate answers, then our analyses can be used to address these questions.

## 2. Methods

### *2.1. Participants and design*

Seventy-one native English speaking undergraduate students at the University of California San Diego participated for course credit. Each participant was given 25 RAT problems in a randomized order. For each problem they saw three cues (e.g., 'moon', 'dew', and 'comb'), and attempted to find a word related to all three (e.g., 'honey'). At the start of the experiment, participants saw one solved RAT problem and were then given three easy practice problems.

Participants had 2 min to solve each problem, and were instructed to enter every word they considered while searching for an answer, regardless of whether that word was correct. If they were sure they had produced the right answer, they were instructed to press F5. After pressing F5, the problem ceased and a point was given if the last entered word was correct. If it was not, a penalty period ensued before moving onto the next problem. The first two times this occurred, the penalty period was 10 s; thereafter it was the remaining time of the problem. If participants never pressed F5, the problem ceased after 2 min and participants received a point if the correct answer was entered as any one of their responses. Participants were informed of their score every five trials.

During each problem, the three cue words were presented in a randomly determined order and a timer displayed the remaining time. Participants entered responses in a space below the cues and pressed the "Enter" key or the F5 key when done typing. While typing, the backspace key was enabled.

### *2.2. Materials*

We selected RAT problems from three sources (Bowden & Jung-Beeman, 2003; Bowers, Regher, & Balthazard, 1990; Mednick, 1962) with the constraint that all of the cues and answers were unique (see Supplementary Table S1 for a list of these RAT problems). Three additional (easy) RAT problems were used as practice trials to familiarize participants with the task.

### *2.3. Metrics of word similarity*

Word-word similarities were calculated via a Latent Semantic Analysis (LSA) of the TASA corpus.[1] LSA summarizes the relationship between words in a corpus by placing them in a multi-dimensional space (300 dimensions in our analysis) such that similar words are close to each other;

[1] This corpus was developed by Touchtone Applied Science Associates (Zeno, Ivens, Millard, & Duvvuri, 1995), and consists of a collection of texts appropriate for students between third grade and the first year of college.

thus word-word similarity is defined as the distance between words in the LSA space (Landauer, 2007). As is typical for LSA, distance was measured as the cosine angle between each pair of words (Martin & Berry, 2007). Although this metric of similarity is only a crude approximation of human semantics, LSA based on the TASA corpus approximates word-sorting and relatedness judgments (Landauer, Foltz, & Laham, 1998), judgments of sentence cohesion (McNamara, Cai, & Louwerse, 2007), and memory intrusions in free-recall tasks (Steyvers, Shiffrin, & Nelson, 2005), to name some examples. To check whether the reported results generalize to other semantic spaces, we performed all of the same analyses using the Word Association Space (WAS; Steyvers et al., 2005), reaching nearly identical conclusions. For details of the LSA space see Supplementary Materials A and for the WAS results see Supplementary Materials B.

### 2.4. Data pre-processing

Fifteen participants were excluded from all analyses because they produced fewer than three responses per RAT problem on average, which may indicate that they were performing the task in a different manner.[2]

For each response to a particular RAT problem, we calculated the similarity between that response and: (a) each cue, (b) the answer, and (c) all prior responses from the same problem. 500 (4.6%) of the 10,862 responses were not found in the TASA corpus and were excluded from all analyses (these were typically misspellings).

LSA calculates the similarity between a word and itself as 1, which is well outside the normal range of word-word similarities. Thus, additional steps were taken to eliminate identity relations from the analyses. First, the answer words were not analyzed, as this would skew tests for a relationship between responses and answers. In 34 instances, a response was given that was identical to one of the cue words, and these responses were dropped from the analyses. In 147 instances, participants repeated the same word twice in a row. We treated these as if the participants had entered that word only once. In 532 instances, participants entered the same word at different non-adjacent places within the same problem. When these word pairs were compared to each other, they were excluded from analyses; otherwise they were retained. Elimination of repeats was a conservative assumption – words tended to be repeated nearby in the sequence of responses, so including identical word-word pairs would have increased the chance of finding local dependencies within the response chain.

## 3. Results

### 3.1. Performance on RAT problems

On average, participants solved 42% of problems within the 2-min period. Although the word-word similarity values did not differ greatly across participants, there were large accuracy differences across both problems and subjects, consistent with the RAT literature. The number of participants solving each problem varied from 5 (9%) to 48 (86%) participants (see Supplementary Table S1 for solution rates by problem). Similarly, participants varied in their RAT proficiency, solving between 4 (16%) and 17 (68%) problems.

### 3.2. How do people constrain potential answers?

Prior research has shown that if people give a single response to RAT problems, that response tends to be related to all three cue words (Gupta et al., 2012). But these final answers reflect the end result of the search/test processes; even if participants primarily search using one cue at a time, they may not give a final answer until the search process produces a word that happens to be similar to all three cues. Here we focus exclusively on the search process rather than the final answer – do they combine the cues to constrain potential answers to words related to all three, or do they pick a single cue at a time to use as a constraint? If potential answers are constrained by all three cues simultaneously, then all of the candidate answers considered prior to a final answer should also be related to all three cues. However, if participants consider candidate answers mainly based on a single cue, then intermediate responses prior to a final response should 'bunch' around that cue in semantic space (we call this the *primary cue*). By bunching, we are not implying that response distribution will have multiple modes (one around each cue), but only that responses will tend to be more related to one of the three cues than would be expected if all cues were weighted equally. If responses bunch in this manner, then responses with the same primary cue (within-cue word pairs) should be more similar to each other than responses with different primary cues (across-cue word pairs).

We tested for this 'bunching' by assigning each response a primary cue, defined to be the cue most similar to the response. We then sorted all adjacent response pairs into within-cue pairs and across-cue pairs. We found that the similarity between the within-cue pairs was higher than across-cue pairs (0.250 vs. 0.142, $t(5572) = 27.7$, $p < 0.001$). This demonstrates that subjects' responses tended to bunch around particular cues, suggesting that participants based their responses largely on a single cue. This test may suffer from selection bias, because words that tend to be closest to the same cue will also tend to be similar to each other. However, the effect remains even when we controlled for this confound by using permutation tests (see Appendix A for details).

This result does not preclude all three cues from constraining the search for potential answers; instead it indicates that one cue is used primarily, but not necessarily exclusively. As reported in Supplementary Materials C, we ran a statistical test demonstrating that responses are more similar to the other two cues than chance when controlling for similarity to the primary cue. However, this statistical test can only be interpreted as providing evidence that participants jointly use all 3 cues under the assumption that our analyses have accurately identified the primary cue for each response. If instead the primary cue is

[2] We ran all tests with these subjects as well and found no qualitative difference in any of the reported results

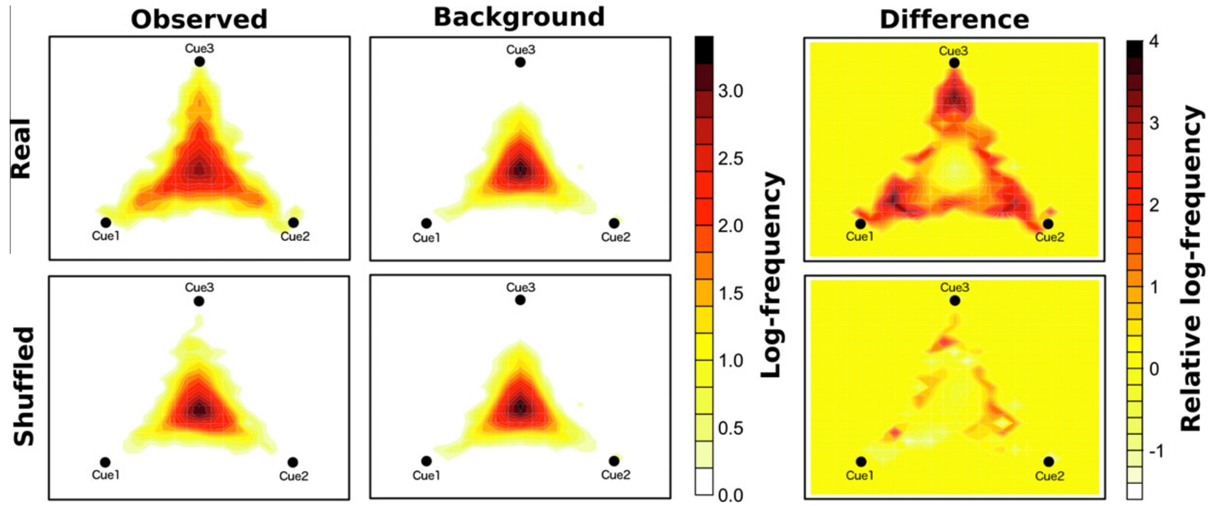


**Fig. 1.** Projection of responses onto the simplex defined by the three cues in semantic space. The top row represents actual responses, while the bottom row represents a single random shuffling of responses such that each response is paired with non-related cues. The left column ('observed') is a heatmap of the log-frequency of finding a response in a given area on that plane. The middle column ('background') is the matched log-frequency of finding any word on the same plane. The right column ('difference') is the log-ratio of the observed and background frequencies.

misidentified for some responses (e.g., the participant used cue1 and sampled a word that was by chance more similar to cue2, resulting in our misidentification of cue2 as the primary cue) this analysis will falsely appear to indicate an influence of more than one cue. Therefore, we cannot make strong claims based on this test.

In lieu of a specific statistical test, we graphically quantified the extent to which responses where similar to each of the cues in Fig. 1 by projecting subjects' responses in LSA semantic space onto the plane defined by the three cues, normalized such that the cues form a standard 2-D simplex.[3] On this simplex, responses equally related to all three cues will appear in the middle, whereas responses more related to one of the three cues will be shifted towards the vertex representing that cue. We find that responses tend to lie in the middle of all three cues (the upper-left, 'real observed' plot); however, so do unrelated responses from other problems (the 'real background' plot)[4], indicating that the preference for the middle is simply a feature of our semantic space. Thus, we must compare how the real responses *differ* from the background distribution: how much more likely are we to find a response in a spot on the simplex compared to all other words (the 'real difference' plot). We find that compared to the background distribution, responses are much more likely to arise near one of the three cues. We also 'shuffled' responses by assigning each to a different RAT problem, calculating where they would be positioned relative to those cues. When we do, the proximity of the responses to the cues disappears. The fact that responses show these relationships to the cues in semantic space suggests that people primarily use a single cue to constrain the search for potential answers. Based on these data and analyses, we cannot rule out whether the other two cues have any effect, although if they do, that effect is very small.

### 3.3. Does search exhibit sequential dependence?

The above analyses demonstrated how the cues were used to constrain the set of potential answers, but did not investigate the impact of prior responses. Here we ask whether the search process has sequential dependencies between responses, which indicates a local search, or whether each response is independent of the other responses, which indicates a global search. Many semantic search tasks use a single semantic constraint, and for these problems use of prior responses may be an effective search strategy (i.e., because prior responses were generated from the original constraint, they may be as similar to valid answers as the original constraint). Indeed, for these single constraint tasks, there is evidence that people use a local search rather than a global search. However, in a multiply-constrained problem such as the RAT, the three cues are completely unrelated to each other. Thus, a prior response generated from a single cue does not necessary bring the semantic search any closer to the goal of finding a word that is related to all three cues. Therefore, it is of interest to test for sequential dependencies in the chain of responses to these RAT problems.

To test whether search is local or global, we investigated whether response pairs from a single problem that are closer together are more similar to one another than pairs from that same problem that are further apart in the response chain. Supporting the local search algorithm, there was a significantly negative linear relationship between the number of intervening responses and average response-response similarity (See Fig. 2; $F(1, 53{,}701) = 482$, $p < 0.001$).

[3] Because projection requires Euclidean distance metrics rather than cosine-distance measures of similarity, this depiction is an approximation of the clustering analysis.

[4] Background words were chosen such that they were responses given to other RAT problems but never as responses to the cues they were paired with. We selected 9361 words in this manner to match the number of real responses analyzed.

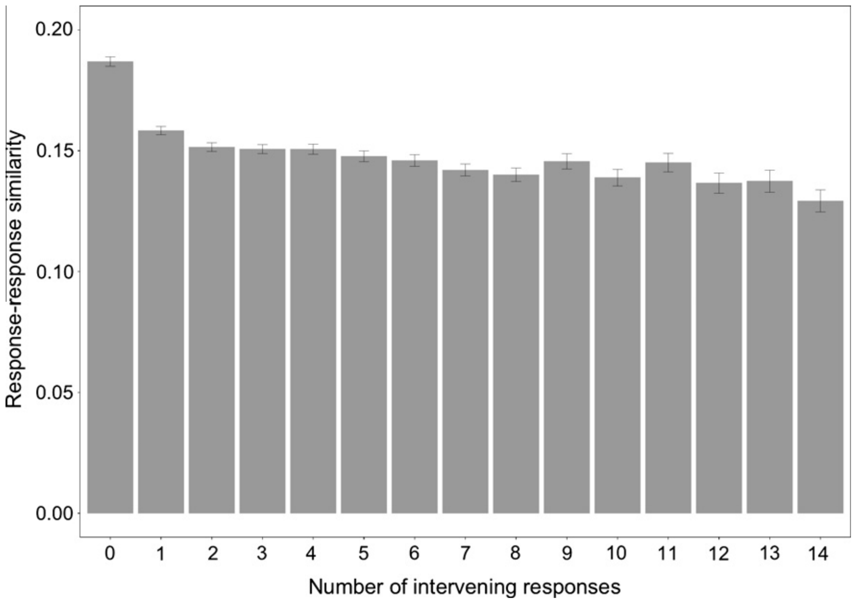


**Fig. 2.** Average similarity between response pairs from the same trial as a function of number of intervening responses (±1 s.e.m.). Responses that were farther apart were more dissimilar than nearby responses, suggesting sequential dependence.

If prior responses are influencing the search process, we might also expect that close associates of the prior response would be produced more quickly. More specifically, if there is a variable degree of semantic drift in the search process between each response, either due to a diffusion process that has run for a variable duration, or due to a variable number of latent samples of candidate answers between each overt response, then subsequent responses that are close associates (i.e., less semantic drift) should correspond with shorter response times. Confirming this prediction, there was a linear relationship between similarity and response time for responses that were adjacent in the sequence ($F(1,8071) = 108$, $p < 0.001$).[5]

### 3.4. How does sequential dependence arise?

The sequential dependence we find between adjacent responses could arise from two distinct mechanisms. We have suggested that it reflects a direct dependence on prior responses or mental states (e.g., direct use of a prior response in the sampling process). On the other hand, apparent sequential dependence could arise through a cluster search process – people may search through separate 'clusters' of words in memory and produce responses from that group, moving to another cluster when the remaining words in the local cluster are few and far between (Gruenewald & Lockhead, 1980; Hills, Jones, & Todd, 2012). In this case, responses from within a given cluster will be more similar to one another than chance, while responses across clusters would not. Because adjacent responses are more likely to be from the same cluster, they are also likely to be more similar to each other as compared to non-adjacent responses. And since the above results suggest that people primarily use a single cue for each response, this clustering hypothesis amounts to longer than chance runs of responses that use the same primary cue (in effect treating this run of responses from a primary cue as a cluster; see Fig. 3).

We can test whether our definition of clusters using the primary cues is informative of the search process. Adjacent within-cluster responses are more similar to one another than adjacent across-cluster responses (0.2471 vs. 0.1389; $t(5761) = 28.7$, $p < 0.001$). Furthermore the adjacent responses from within the same cluster as the prior response are generated more quickly than adjacent responses from different clusters (8.5s vs. 9.4s; $t(7394) = 4.14$, $p < 0.001$). Thus if people are using a cluster search strategy, defining each cluster as a string of responses related to one of the three cues is informative of the response relationships.

A simple test of whether people select from cue 'clusters' is to check whether participants continued to use the same primary cue for multiple responses in a row, which would suggest that they had picked one primary cue and produced a number of words from it. However, the chance level of repetition is not 1-in-3; some cues are assigned as primary cues more often than others. After correcting for the base rate of cue use assuming independent transitions, the chance probability of staying with the same primary cue would be 40.6%. Of the 8342 response pairs tested, 3573 had the same primary cue (42.8%), which is statistically greater than chance (*Binomial Test (Exact)*, $p < 0.001$), but is a small numerical difference.

Qualitatively, these results appear to favor the cluster hypothesis. However, they are not diagnostic of this process: even with direct influence, some response pairs will

[5] Because we do not make claims about the linearity of this relationship, we also tested for a relationship between rank similarity and rank response time, and still found a significant effect ($F(1,8071) = 185$, $p < 0.001$).

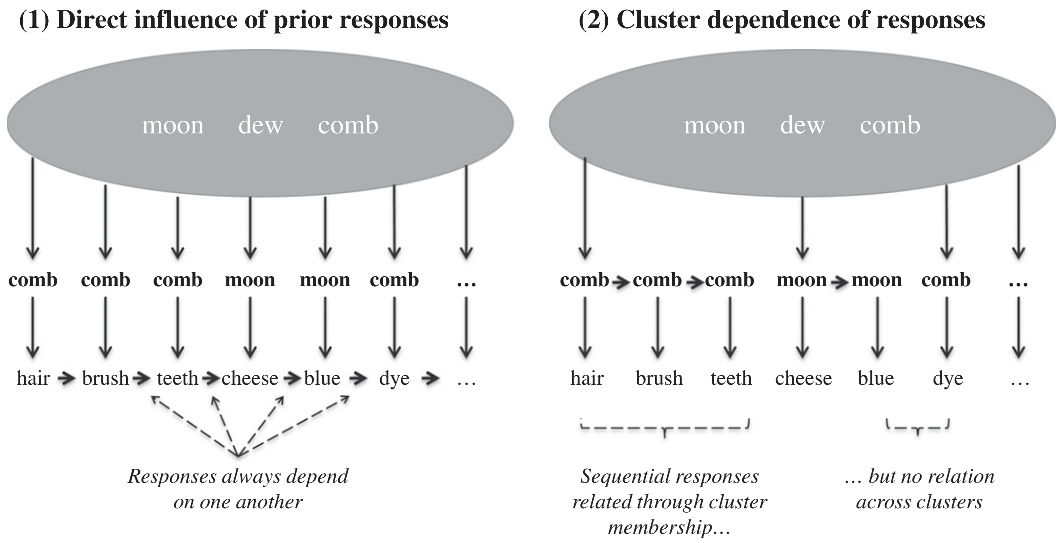


**Fig. 3.** Illustration of sequential dependence through direct influence (1) or cluster-level dependence (2). If responses are related through direct influence, then adjacent responses should always be more related than would be expected by chance. If responses are related by virtue of their cluster, then there will be a mixture of related adjacent responses within a cluster and unrelated responses across cluster boundaries, but averaging across both groups will show that adjacent responses are more related than chance.

by chance be less similar, and these responses will have longer inter-item response times and will be more likely to be assigned to different clusters. Furthermore, a slight increase in the rate of continued use of the same primary cue is also consistent with direct influence from prior responses. More specifically, direct relationships will naturally produce some longer runs with the same primary cue: because a pair of subsequent responses are similar to each other through direct influence, they will be more likely than chance to be most similar to the same cue.

If the sequential dependencies we observed arose from a cluster search process, then there would be two additional characteristics of the response chain. First, there should be 'breaks' when people switch cues, such that responses on either side of the switch should show only chance levels of similarity (e.g., retrieving items based only on global criteria during a cluster switch, as in the dynamic model of Hills, Jones, and Todd (2012)). Second, there should be no dependence of responses within cue clusters, such that adjacent words within a cluster should be no more similar than non-adjacent responses.[6]

If people make a clean switch between clusters, then we should be able to observe 'breaks' in their response chain with a cluster switch, where the first response of a new cluster is not influenced by the preceding response. Therefore if we examine response pairs with different primary cues (thus coming from different clusters), there should be no effect of adjacency on the relationship between those responses. This test showed that of the response pairs with different primary cues, adjacent responses were more similar than non-adjacent responses (0.1389 vs. 0.1133; $t(5977) = 13.05$, $p < 0.001$), indicating that adjacent responses are related even across cluster breaks.

Sampling from clusters would also predict no intracluster dependencies. Thus in a contiguous chain of responses that have the same primary cue, adjacent responses should be no more similar to one another than non-adjacent responses. However, the data suggest that adjacent responses are more similar even when the responses arise from the same primary cue (0.2471 vs. 0.2136; $t(6705) = 7.64$, $p < 0.001$), suggesting that the sequential dependence does not arise from clustering alone.

These results suggest that the sequential dependence between responses is at least in part driven by direct association between responses; although adjacent responses are more likely to share a primary cue, the fact that sequential dependence was still found within and across clusters suggests that this dependence cannot be due to reliance on cluster search alone.

### 3.5. The direction of local search

In computer science, local search algorithms often use gradients, or the information about the success of prior guesses, to systematically move towards the answer (e.g., gradient descent). Here we test whether something similar happens in the semantic search process for RAT problems. If people preferentially move towards the answer, their responses should become more similar to the answer over time on trials in which they find the correct answer. However, sequential dependencies could give the appearance that this occurs due to an artifact of path selection. Consider that local search is a kind of random walk in semantic space. There will be natural variation in the path taken by the random walk, and some paths will randomly arrive at the answer while others will not. By selecting only those

[6] This is not a prediction of the Hills et al. model which assumes local search within (but not across) clusters, but follows logically from the assumption that local search is driven solely by the clustered nature of semantic retrieval.

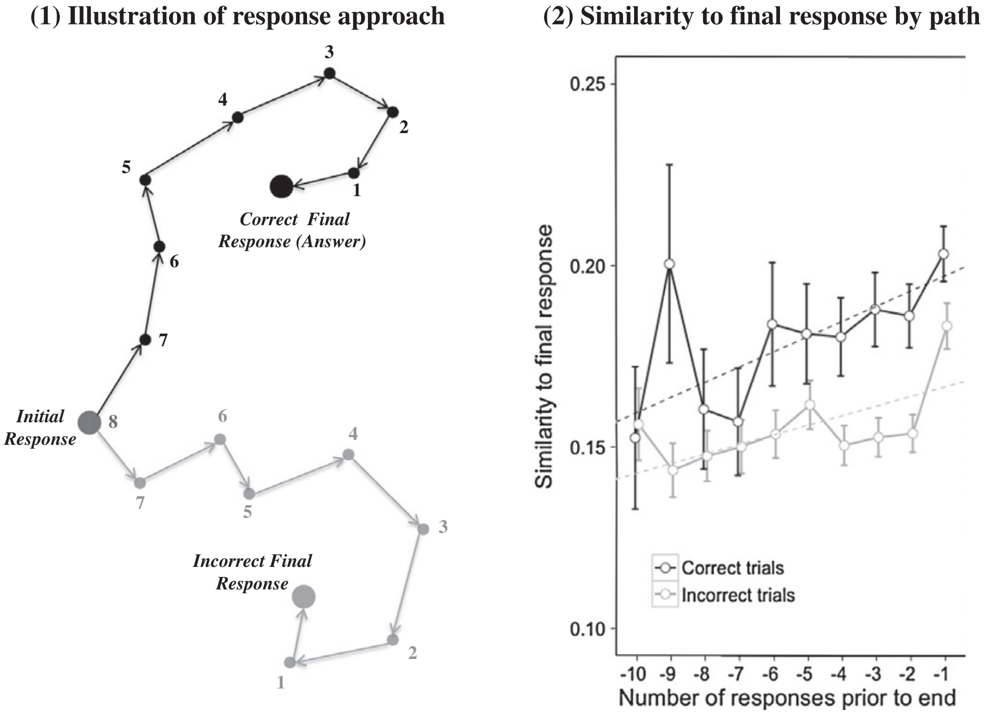


**Fig. 4.** (1) Illustration of how sequential dependencies correspond to a random walk in semantic space such that responses immediately prior to the final response, regardless of whether the final response is correct or incorrect, are likely to be more similar to the final response than responses occurring earlier in the sequence. (2) Average similarity between responses and the final response in the chain (the answer in 'correct' trials) as a function of the number of responses prior to the final response. Although responses become more similar to the final response with fewer intervening responses, there is no difference in the slope of this function when comparing 'correct' and 'incorrect' trials, suggesting that the approach is due to sequential dependence between responses, rather than a directed movement towards the answer. Lines represent best fitting linear trendline for each trial type, and bars represent standard error.

paths that randomly arrive at the answer, it will appear, on average, as if the path taken by the participant was one that was headed specifically towards that answer. Notably, this same logic applies to an analysis of incorrect trials if the similarity metric is applied in relation to the final (incorrect) response rather than the correct answer. Thus, the key to determining whether the search process hones in on the answer is a comparison of the path towards the answer for correct trials versus the path towards the final response for trials that were ended due to time constraints (see Fig. 4).[7] As seen in the figure, the sequences for both trial types became progressively more similar to the final response.

To test whether responses approached the answer more quickly than incorrect responses, we tagged each response by its location relative to the final response in the response chain (e.g., the penultimate response would be tagged as '-1', the response before that as '-2', etc.). We then measured the average similarity of responses to the final response for correct trials (which ended on the answer) and incorrect trials. We could then measure the rate of approach as the slope of the line that fits the similarity to the final response as a function of responses prior to the end, where greater slopes represent faster approach. For the subset of responses that ended on the answer there was a positive slope ($F(1,1265) = 7.32$, $p = 0.0069$) over the final 10 responses, suggesting that responses do become closer to the answer over time. However, this did not differ from the rate of approach to the final response in incorrect trials ($F(1,5859) = 0.077$, $p = 0.38$), which suggests that this approach is due to random walk selection effects rather than participants honing in on the answer.[8]

## 4. Discussion

To solve RAT problems, and multiply-constrained problems in general, people first search through a set of potential answers and then test those answers against each constraint for suitability. In this experiment, we investigated the search process, and found that people solve RAT problems by selecting a set of promising answers constrained primarily by one of the three cues at a time. In

[7] We excluded trials in which participants incorrectly marked their final response as the answer, leaving only trials terminated due to time constraints. Thus, the similarity between these two paths does not reflect a honing in on a specific word believed (incorrectly) to be the answer.

[8] Responses in correct chains did in general have greater similarity to the final response than responses in incorrect chains ($F(1,5859) = 34.9$, $p < 0.001$). This is likely because the final response in correct chains were answers, which were on average more related to randomly selected words than incorrect responses were (0.1371 vs. 0.1066; $t(20651) = 20.9$, $p < 0.001$). Furthermore, this effect disappears when using WAS metrics, in which answers are not more related to random words than responses. Thus this difference is a factor of answers being special words in the LSA space rather than reflecting a facet of the search process.

addition to using just one cue at a time, there were sequential dependencies in sequence of considered candidate answers, suggesting that prior guesses directly affect subsequent guesses (i.e., local rather than global search). Finally, we did not find evidence that this search process specifically hones in on the correct answer.

### 4.1. Combining constraints for memory search

One of the key features of multiply-constrained problems is that the set of potential answers is enormous before the constraints are imposed, so an exhaustive search of this space is impossible. Nevertheless, people quickly limit the search space to promising answers. This study shows that people find these promising answers by focusing primarily on one constraint at a time.

This finding at first seems to be at odds with prior models of episodic and semantic memory. Many of these models assume that multiple distinct cues interact to probe memory (Gillund & Shiffrin, 1984; Murdock, 1997; Ratcliff & McKoon, 1988). This joint cue representation is required to explain episodic memory phenomena such as recognition for multiple-item probes – e.g., if 'door', 'radio' and 'wall' are studied together, then recognition of 'door' is better if 'radio' and 'wall' are presented as additional retrieval cues than if 'door' is presented by itself (Clark & Shiffrin, 1987). But on the RAT, people retrieve potential answers primarily on the basis of a single cue, suggesting that memory cannot be probed using an arbitrary combination of cues. However, in these episodic memory paradigms, not only was memory cued with a joint cue, but the memories that were being retrieved were joint memories (i.e., the participant studied the three words jointly and then was cued with those same three words). In contrast, the set of cues in the RAT does not exist as a joint memory (e.g., it is unlikely that participants previously saw the words 'moon', 'comb', 'dew', and 'honey' together). Thus, it is not clear if these episodic memory results truly indicate an ability to use a joint cue on the fly-instead, it may be that prior study of the joint set of items allows the participant to learn the specific combination as new single conjoined representation.

Similar to the findings in episodic memory experiments, there has been prior research attempting to characterize how people combine cues to access semantic memory that came to different conclusions than this study. For instance, Rubin and Wallace (1989) found that when given both meaning and rhyme cues (e.g. 'a mythical being' and 'rhymes with ost') people would often come up with a response ('ghost') that would rarely or never be produced in response to one of the cues individually. This finding is used as evidence that memory works super-additively: the combination of the cues retrieves 'ghost' more often than an additive mixture of the cues. However, this and similar studies (e.g. Massaro, Weldon, & Kitzis, 1991) rely on a single, final response. Since the final response is the product of both a search and a test process, one possibility is that the search process proposes words by considering cues additively, while the test process accepts these proposals based on a multiplicative combination of cues. This account would offer a resolution to the tension between our finding that interim responses are related mainly to one cue alone and these findings that final answers are based on super-additive combinations of the two cues. Any interim words that are considered, but are not related to both cues, would be rejected by the test process and not produced in these tasks. For instance, the rhyme cue may cause 'most' to come to mind in both the single and dual-cue tasks; this will be an acceptable answer for just the rhyme cue (and so would be recorded), while it will be rejected in the dual-cue task (and so would not be captured). On the other hand, these tasks also differ from our RAT task in another important way: they require combining only two cues. People may therefore have attentional or memory limits that allow a conjunctive search based on two cues, but become overtaxed when three or more cues are presented. Further research is required to determine whether there is a qualitative difference between two-cue and three-cue memory search.

Our results raise the question of what causes the focus on a single cue. On the one hand, this could be a limitation of memory. However, it is not difficult to keep the three cue words in working memory, so this capacity constraint should not be the cause. But it is possible that there are additional attention or processing constraints that limit memory from retrieving based on more than one cue at a time. Conversely, this could be a strategy people use to find answers efficiently; perhaps searching based on individual cues arrives at an answer more quickly than searching based on a combination. Further research is required to disentangle these possibilities.

### 4.2. Exploration of potential answers

We found sequential dependencies between responses within a problem: responses tended to be semantically similar to the previous response, consistent with a local search process. Other studies have found sequential dependencies in tasks requiring semantic retrieval and production with a single constraint (Gruenewald & Lockhead, 1980; Hills et al., 2012; Howard & Kahana, 2002; Troyer, Moscovitch, & Winocur, 1997), and our results expand this finding by demonstrating that these dependencies exist even in tasks with multiple constraints.

We also found evidence that this dependence arises directly from relationships between responses. Prior studies have used categorical clustering as evidence of sequential dependencies (Bousfield, 1953; Troyer, Moscovitch, Winocur, Alexander, & Stuss, 1998), and some have hypothesized that the sequential dependencies directly result from a clustering search strategy (Hills et al., 2012; Pollio & Gerrow, 1968). However, we found evidence of sequential dependences both within and across clusters, which suggests that there is a direct influence of prior responses rather than sole reliance on cluster-level search.

This direct association of adjacent responses can arise from two distinct types of processes. First, this could be the result of sampling from an evolving search process. Under this account, search is a drift through semantic memory, and responses are intermittent samples from this process. This may be a facet of a general search process, as hypothesized by Hills, Todd, and Goldstone (2008). Second, this could be driven by a priming process without a specific functional significance; in this case, responding with a word will make that word more salient in semantic memory and therefore cause related words to be retrieved with greater ease. Both explanations are consistent with the response chains we observed in this study; different experiments will be required to determine which explanation underlies these sequential dependencies.

Regardless of the mechanism underlying these sequential dependencies, it is difficult to say whether that mechanism is a natural component of the search process or whether that mechanism only arises when participants are asked to overtly report their intermediate guesses. It is possible that the act of overtly producing a response changes the nature of the search process, serving to prime subsequent responses in a manner that would not occur if prior guesses were not overly reported. This is a limitation shared by all behavioral studies of semantic search, though made more explicit here because RAT problems require only a single correct response (as opposed to naming all animals that come to mind, for instance). But without a tool that would allow us to continuously measure the words under consideration without asking for a response, we must assume that the act of responding does not have a large impact on the search process.

Local search algorithms are the most common method for stochastic optimization in engineering applications (Spall, 2003), which might suggest that this dependence arises from a rational search process such as Markov Chain Monte Carlo (Griffiths, Vul, & Sanborn, 2012). And indeed, in many other tasks that require searching through semantic knowledge, responses that are related to prior responses are more likely to be viable candidates; e.g., when producing animal names, words associated with 'cat' are more likely to be animals as well. It is therefore possible that local search produces a more efficient exploration of the space of possible answers; however, on this RAT task we do not find evidence that local search causes faster convergence to the answer than would be expected to any other response in the chain.

This then begs the question – if people are searching through their semantic memory using a process similar to a random walk, how does anyone solve a RAT problem when there are so many words to consider and so few responses given? If subjects had only considered the words that they provided as responses, the probability of stumbling on the answer would be vanishingly small using this model. Instead, we believe that the responses that people provide are a small subset of all potential words considered; responses are the words that happened to be activated enough that subjects noticed and recorded them, but do not describe the full underlying path of semantic search. The full process is able to sample significantly more words, thus increasing the odds of stumbling on the correct answer.

### 4.3. Explicit versus implicit search

The current task was designed to investigate the process people use to search for answers in the RAT by investigating intermediate responses elicited from participants. Our procedure assumes that people have conscious access to this search process such that their intermediate guesses reflect the current status of the search. Suggesting that the search process underlying RAT performance may be implicit rather than explicit, prior studies have found that people can identify whether a combination of three cues in the RAT does or does not have valid solution within a second and a half of first viewing the cues even though they are unable to explicitly produce a solution within this time period (Bolte & Goschke, 2005; Topolinski & Strack, 2009). Similarly, Kounios and Beeman (2009) have found differences between deliberate and insight solutions to RAT problems. The current procedure can thus only address the conscious deliberate search for answers; it does not specify the nature of rapid implicit processes. It is possible that the explicit search process addressed in the current study is wholly unrelated to the implicit process that produces the correct answer on RAT problems.

However, we have reason to believe that these intermediate wrong answer guesses are indicative of the processes that support RAT performance. For instance, Gupta et al. (2012) used a version of the RAT in which participants were asked to give a single best guess within a short time period even if they knew that guess to be false. An analysis of these incorrect best guesses found that people who did better on the RAT were people who tended to give low frequency incorrect guesses. This supports the claim that overtly produced incorrect guesses relate to the processes of finding the correct solution to RAT problems.

### 4.4. Conclusion

From causal reasoning to analogy, to day-to-day tasks such as prioritizing work, many of the problems people face in naturalistic settings require finding the best answer or course of action from a huge set of potential answers, limited by only a handful of constraints. Finding an acceptable solution to these problems requires querying ones background knowledge in a quick and efficient fashion.

In this research we have begun to explore this process, finding that people search through their semantic knowledge primarily using one constraint at a time, and that direct sequential dependence occurs between search items. Of course, many questions still remain: is the focus on a single constraint a limitation of memory, or a strategy for solving these RAT problems in particular? Is direct sequential dependence due to semantic priming or a local search strategy? And how do people decide whether a potential answer is considered acceptable? Investigating how people solve multiply-constrained problems will provide further insight into how semantic memory is structured, and how people combine and use cues from their environment to retrieve the most task-relevant information. In general, understanding multiply-constrained search is necessary to explain how people are able to solve everyday problems so effortlessly.

## Acknowledgments

Edward Vul and Kevin Smith were supported by IARPA via DOI contract D10PC20023. The views and conclusions

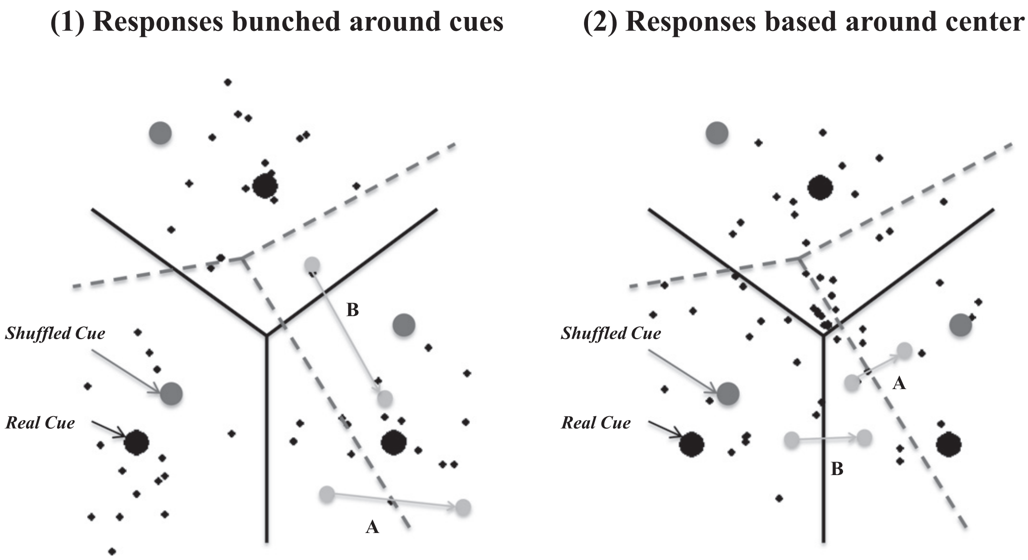


**Fig. A1.** Illustration of semantic space demonstrating test of response bunching. Responses were assigned to 'real' primary cues, with divisions delineated by the solid black lines, and 'shuffled' primary cues, delineated by the dashed grey lines. If responses bunch around the cues (1), then adjacent responses that cross the 'shuffled' lines but not 'real' lines (A) should be on average more similar than response pairs that cross 'real' but not 'shuffled' lines (B). If responses are generated based on an equal conjunction of the cues (2), then there should be no difference in adjacent responses.

contained herein are those of the authors and do not represent the official policies or endorsements of IARPA, DOI, or the U.S. Government. David Huber was supported by National Science Foundation Grant BCS-0843773.

## Appendix A. Addressing confounds in the test of cue usage

When we assign primary cues to each response (see Section 3.2), we cannot know which cue was used to constrain each response, so we estimated this primary cue by identifying the cue most similar to that response. We then found that adjacent responses sharing the same primary cue (within-cue pairs) were more similar than responses with different primary cues (across-cue pairs).

However, this assignment of primary cues to responses will favor greater similarity for within-cue word pairs due to chance factors; some adjacent pairs will be more similar than others by chance, and it is more likely that similar word pairs will be assigned the same primary cue. While we could not get rid of this structural bias, we could build it into our null hypothesis distribution, thus asking whether within-cue similarity is greater than would be expected by chance given the baseline similarity arising from cue assignment. We achieved this via across-problem randomization: to construct a null hypothesis distribution, the same across- and within-cue similarity measures were calculated except that the assignment of primary cues was based on the cues from another randomly selected problem ('shuffled cues'). Under this random assignment of cues to responses, the increase in similarity for within-cue response pairs as compared to across-cue response pairs was due solely to chance assignment of the shuffled primary cue. In addition, because the same words appeared in both the analysis of the effect and the null hypothesis, the specific semantic and lexical properties of the cues were controlled. Thus if the increase in similarity from within-cue to across-cue responses is greater when using the real cues as compared to the shuffled cues, then we have evidence for bunching of the responses around the real cues (see Fig. A1).

For this analysis, there were four classifications of word pairs, representing combinations of across- or within-cue as determined by real or shuffled cues. Of key interest were word pairs that were within-cue pairs for the real cues, and across-cue pairs for the shuffled cues (A-type response pairs in Figure A1) versus ones that were across-real-cues, but within-shuffled-cues (B-type responses). Both of these word pair types have a within- and an across-cue determination, so the same degree of similarity is imposed according to chance factors. However, if responses bunch around the cues, A-type pairs should be more similar since they are within the boundaries established by the real cues. A comparison of these similarity ratings showed that A-type response pairs were more similar than B-type response pairs (average similarities = 0.209 vs. 0.160; $t(2502) = 9.04$, $p < 0.001$). Because this analysis controls for the confound of primary cue assignment inherent in the original analysis, we can be confident that the results of this test are due to a feature of how people use the cues to produce responses and not due to the method by which we performed the test.

## Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2013.03.001.

## References


Bolte, A., & Goschke, T. (2005). On the speed of intuition: Intuitive judgments of semantic coherence under different response deadlines. *Memory & Cognition, 33*(7), 1248–1255.

Bolte, A., Goschke, T., & Kuhl, J. (2003). Effects of positive and negative mood on implicit judgments of semantic coherence. *Psychological Science, 14*(5), 416–421.

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology, 49*, 229–240.

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, 35*(4), 634–639.

Bowers, K. S., Regher, G., & Balthazard, C. (1990). Intuition in the context of discovery. *Cognitive Psychology, 22*, 72–110.

Cai, D. J., Mednick, S. A., Harrison, E. M., Kanady, J. C., & Mednick, S. C. (2009). REM, not incubation, improves creativity by priming associative networks. *Proceedings of the National Academy of Sciences, 106*(25), 10130–10134.

Clark, S. E., & Shiffrin, R. (1987). Recognition of multiple-item probes. *Memory & Cognition, 15*(5), 367–378.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic priming. *Psychological Review, 82*(6), 407–428.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433.

Ernst, M. O., & Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Science, 8*(4), 162–169.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19*, 141–205.

Gillund, G., & Shiffrin, R. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*(1), 1–67.

Graesser, A. I., & Mandler, G. (1978). Limited processing capacity constrains the storage of unrelated sets of words and the retrieval from natural categories. *Journal of Experimental Psychology: Human Learning and Memory, 4*(1), 86–100.

Griffiths, T. L., Vul, E., & Sanborn, A. N. (2012). Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science, 21*(4), 263–268.

Gruenewald, P. J., & Lockhead, G. R. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory, 6*(3), 225–240.

Gupta, N., Jang, Y., Mednick, S. C., & Huber, D. E. (2012). The road not taken: Creative solutions require avoidance of high frequency responses. *Psychological Science, 23*(3), 288–294.

Harkins, S. G. (2006). Mere effort as the mediator of the evaluation–performance relationship. *Journal of Personality and Social Psychology, 91*(3), 436–455.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4*, 967–992.

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

Hills, T. T., & Pachur, T. (2012). Dynamic search and working memory in social recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(1), 218–228.

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*(8), 802–808.

Hills, T. T., Todd, P. M., & Goldstone, R. L. (2010). The central executive as a search process: Priming exploration and exploitation across domains. *Journal of Experimental Psychology: General, 139*(4), 590–609.

Howard, M. W., & Kahana, M. J. (2002). When does semantic similarity help episodic retrieval? *Journal of Memory and Language, 46*(1), 85–98.

Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology, 52*(6), 1122–1131.

Kounios, J., & Beeman, M. (2009). The aha! moment: The cognitive neuroscience of insight. *Current Directions in Psychological Science, 18*(4), 210–216.

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*(2), 211–240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2&3), 259–294.

Landauer, T. K. (2007). LSA as a theory of meaning. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 3–34). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Martin, D. I., & Berry, M. W. (2007). Mathematical foundations behind latent semantic analysis. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 35–56). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Massaro, D. W., Weldon, M. S., & Kitzis, S. N. (1991). Integration of orthographic and semantic information in memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(2), 277–287.

McNamara, D. S., Cai, Z., & Louwerse, M. M. (2007). Optimizing LSA measures of cohesion. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 379–400). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Mednick, S. (1962). The associative basis of the creative process. *Psychological Review, 69*(3), 220–232.

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review, 104*(4), 839–862.

Pollio, H. R., & Gerrow, J. R. (1968). The role of rules in recall. *The American Journal of Psychology, 81*(3), 303–313.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95*(3), 385–408.

Rubin, D. C., & Wallace, W. T. (1989). Rhyme and reason: Analyses of dual retrieval cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(4), 698–709.

Russell, S., & Norvig, P. (2003). *Informed Search and Exploration, Artificial Intelligence: A Modern Approach*. Upper Saddle River, New Jersey: Pearson Education, Inc..

Spall, J. C. (2003). *Introduction to Stochastic Search and Optimization*. New York: Wiley.

Steyvers, M., Shiffrin, R., & Nelson, D. (2005). Word association spaces for predicting semantic similarity effects in episodic memory. In A. Healy (Ed.), *Experimental Cognitive Psychology and Its Applications* (pp. 237–249). Washington, DC: American Psychological Association.

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

Topolinski, S., & Strack, F. (2008). Where there's a will – there's no intuition. The unintentional basis of semantic coherence judgments. *Journal of Memory and Language, 58*, 1032–1048.

Topolinski, S., & Strack, F. (2009). Scanning the "fringe" of consciousness: What is felt and what is not felt in intuitions about semantic coherence. *Consciousness and Cognition, 18*, 608–618.

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

Troyer, A. K., Moscovitch, M., Winocur, G., Alexander, M. P., & Stuss, D. (1998). Clustering and switching on verbal fluency: The effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia, 36*(6), 499–504.

Vul, E., & Pashler, H. (2007). Incubation benefits only after people have been misdirected. *Memory & Cognition, 35*(4), 701–710.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The Educator's word frequency guide*. Touchstone Applied Science Associates.