Research Article

Psychological Science
23(3) 288–294

DOI: 10.1177/0956797611429710
http://pss.sagepub.com

The Road Not Taken: Creative Solutions Require Avoidance of High-Frequency Responses

Nitin Gupta[1], Yoonhee Jang[1], Sara C. Mednick[2], and David E. Huber[1]
[1]Department of Psychology and [2]Department of Psychiatry, University of California, San Diego

Abstract

To investigate individual differences in creativity as measured with a complex problem-solving task, we developed a computational model of the remote associates test (RAT). For 50 years, the RAT has been used to measure creativity. Each RAT question presents three cue words that are linked by a fourth word, which is the correct answer. We hypothesized that individuals perform poorly on the RAT when they are biased to consider high-frequency candidate answers. To assess this hypothesis, we tested individuals with 48 RAT questions and required speeded responding to encourage guessing. Results supported our hypothesis. We generated a norm-based model of the RAT using a high-dimensional semantic space, and this model accurately identified correct answers. A frequency-biased model that included different levels of bias for high-frequency candidate answers explained variance for both correct and incorrect responses. Providing new insight into the nature of creativity, the model explains why some RAT questions are more difficult than others, and why some people perform better than others on the RAT.

Keywords

remote associates test, word frequency, creativity, word association, individual differences, associative processes, language, semantic memory

Received 4/12/11; Revision accepted 9/6/11

Real-world problems are often complex and ill defined. For instance, imagine that you open the refrigerator to figure out what to cook for an unexpected dinner guest, but, to your horror, you have only potatoes, tomatoes, onions, and eggs. In a flash of insight, you figure out the solution—a Spanish omelet! From among the many thousands of possible dishes, how did you manage to find this solution? And why is it that some people are much better than others at finding creative solutions to complex problems?

In 1962, S. A. Mednick defined creativity as "the forming of associative elements into new combinations, which either meet specified requirements or are in some way useful" (p. 221). On the basis of this definition, Mednick developed a test of creativity called the remote associates test (RAT). Each question on the RAT is composed of three apparently unrelated cue words that associate to or associate from a fourth word, which is the correct answer (e.g., cues: *surprise*, *line*, and *birthday*; answer: *party*). Early work established that RAT performance correlates with traditional measures of IQ (M. T. Mednick & Andrews, 1967) and predicts originality during brainstorming (Forbach & Evans, 1981; although see Kray, Galinsky, & Wong, 2006). More recently, the RAT has been used to measure the effects of manipulations related to creativity. For instance, the RAT has been used to study intuition and incubation (Bowers, Regehr, Balthazard, & Parker, 1990; Topolinski & Strack, 2008, 2009; Vul & Pashler, 2007), the role of affect during problem solving (Fodor, 1999; Isen, Daubman, & Nowicki, 1987), implicit learning during REM sleep (Cai, Mednick, Harrison, Kanady, & Mednick, 2009), and the relation between synaesthesia and creativity (Sitton & Pierce, 2004; Ward, Thompson-Lake, Ely, & Kaminski, 2008), to name just a few examples.

Despite the demonstrated utility of the RAT, there are no formal (i.e., computational) models specifying why some people perform better than others on the RAT. Here, we report a study in which we tested one account of individual differences in the search process that takes place while people solve these problems. The RAT search process was initially investigated in a study of performance evaluation. By interjecting lexical

Corresponding Author:
David E. Huber, Department of Psychology, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0109
E-mail: dhuber@ucsd.edu

decisions following presentation of the RAT cue words and by manipulating various delays, Harkins (2006) determined that the threat of performance evaluation increased fixation on incorrect words closely associated to the cue words, thus blocking access to the correct remote associate. The current study built on this work by mathematically formalizing the search process within a norm-based semantic space. This allowed separate analysis of each RAT question and each individual. With this formalization, we tested the hypothesis that highly creative individuals, as measured by the RAT, are able to access remote associations because they are not biased to consider only high-frequency words. It follows that difficult RAT questions are ones in which the cue words are associated with incorrect high-frequency words, which makes it difficult to access the correct answer. We tested this hypothesis empirically by examining the relation between the word frequency of responses, both correct and incorrect, and performance on the RAT, and we tested it theoretically by modeling individual differences at the level of each participant and each RAT question.

Method

To test the role of word frequency, we performed a RAT experiment in which participants had only 30 s for each RAT question and were encouraged to guess before the response deadline. For this experiment, we wanted RAT questions for which the correct answer was easily predicted using only associative information without regard to word frequency or other factors. By selecting RAT questions in this manner, we defined a norm-based model of what people should do, which we contrasted with a frequency-biased model of observed behavior.

Norm-based model

Because RAT questions are based on remote associations, many of the associations between cues and responses are missing in Nelson, McEvoy, and Schreiber's (1998) association norms.[1] To fill in these missing associations, we used the Word Association Space (WAS) of Steyvers, Shiffrin, and Nelson (2004). To create the WAS, they compressed the 5,018 words contained in the association norms of Nelson et al. into a reduced 400-dimensional representation by using singular value decomposition (SVD). Although 400 dimensions defines a vast space, this number of dimensions is much less than 5,018; thus, latent associations are revealed in the WAS. For each RAT question, we calculated the euclidean average within the WAS for the three cue words and used this *cue centroid* to define the best guess at an answer. The 5,018 words in the association norms were ordered according to their similarity to the cue centroid, and performance of the model was measured as the relative position of the correct answer in this list (i.e., the percentage of words that were more dissimilar from the cue centroid than the correct answer was). As is commonly done with semantic spaces based on SVD (e.g., Landauer & Dumais, 1997), we used the cosine of the angle between points in the WAS to calculate semantic similarity.

Participants

One hundred thirty-two undergraduate students at the University of California, San Diego, were recruited and received credit for psychology courses in return for their participation.

Materials

Collecting RAT questions from three different sources (Bowden & Jung-Beeman, 2003; Bowers et al., 1990; S. A. Mednick, 1962), we identified 178 RAT questions for which all three cue words and the answer existed in the WAS. For each of these questions, the similarity between the cue centroid and each of the 5,018 words in the WAS was calculated. The experimental stimuli consisted of the 48 RAT questions (see Table S1 in the Supplemental Material available online) that produced the highest performance for the norm-based model, subject to the constraint that no cue word or answer word could appear in more than 1 RAT question. By including only these RAT questions, we ensured that the cue centroid was a good indicator of the correct answer for every question. Four additional RAT questions were used for practice trials, which were not analyzed.

Procedure

For each RAT question, participants were presented with the three cue words and asked to find the common associative link among them. The cue words were presented horizontally in the center of a computer screen, in the same left-to-right order as they appear in Table S1 in the Supplemental Material. Participants were asked to type their best guess at an answer within the allotted time of 30 s. If they provided no response within the first 15 s, a string of asterisks appeared over the cue words, indicating that only 15 s remained. After a participant typed in a response and pressed the "enter" key, the trial was stopped regardless of whether the entire 30 s had elapsed. The back-space key was enabled so that participants could retype a response prior to the end of a trial. Answers were recorded whenever the "enter" key was pressed or when 30 s elapsed. For each participant, the 48 RAT questions were randomly divided into four blocks containing 12 questions each. Between blocks, participants performed an unrelated task for 10 min.

Word frequency

In keeping with a recent theory of insight proposed by Topolinski and Reber (2010), our account of RAT performance supposes that answers are considered if they spring to mind easily. In other words, we propose that responses are biased toward words with high fluency. Griffiths, Steyvers, and Firl (2007) recently studied word fluency by asking participants to report

the first response that came to mind when they were given a first letter (e.g., "a" for "apple"). Griffiths et al. applied different frequency measures to the results of this experiment, finding that traditional measures of written word frequency, such as the Kucera and Francis (1967) norms, failed to explain word fluency. In contrast, PageRank values for individual words, as determined from the association norms, provided a better measure of fluency. For this application, the PageRank of a word is the probability of visiting that word during a random walk through the associative links between words. However, a very good approximation to PageRank is the sum of the in-links, which in this case meant the sum of the association strengths of all the cue words that associate to a particular target word. Griffiths et al. termed this *associate frequency* (AF) and found it to yield results nearly indistinguishable from those obtained with PageRank. Therefore, we used AF as our measure of word frequency.

Frequency-biased model

We assumed that solving a RAT question entails three aspects of the search process: (a) the order in which candidate words are evaluated in the search set, (b) whether the correct answer is recognized as being the correct answer if it is evaluated, and (c) guessing upon quitting the search process. To capture individual differences in the search process, we assumed that the order of the search set is biased by word frequency. We implemented this assumption with the weighted similarity measure, $f(w)^F \times \text{similarity}(w, C)$, where the similarity between word w and the cue centroid, C, is calculated with the cosine angle in the WAS, $f(w)$ is the AF of word w, and F is an individual difference parameter that determines the role of frequency. The search set included all 5,018 words in the WAS ordered by weighted similarity from high to low. A geometric distribution was used to model the probability of quitting the search process at each position in the search set; the geometric parameter p denotes the probability of quitting after each word in the set. For a word with rank i, the probability that the search reached that word is given by $(1-p)^{i-1}$.

Search evaluates each word to determine whether it is the correct answer. If the correct answer is evaluated, it is recognized as being the correct answer with probability R, in which case the search process stops. R provides a second measure of individual differences, allowing for the possibility that some individuals may not carefully consider whether a word meets the task requirements. If the correct answer is at rank position k, the probability that it is reached prior to termination of the search and also recognized as the correct answer is calculated as follows:

$$P_{\text{recognized}} = (1-p)^{k-1}R.$$

If participants are aware that recognition of the answer is less than perfect, a good guess may lie among words already evaluated and rejected. More specifically, participants may recognize one or two of the associations between an evaluated word and the cues, but fail to appreciate all three. Thus, we assume that participants are biased to guess on the basis of the direct associations between previously evaluated words and the cues. Across the three cues for a RAT question, the combined associative strength of a word with rank i (denoted as W_i) is as follows:

$$C(i) = \varepsilon + \sum_{j=1,2,3} A(\text{Cue}_j \rightarrow W_i) + A(W_i \rightarrow \text{Cue}_j),$$

where Cue_j denotes the jth cue for the RAT question and A denotes the associative strength (with direction) based on the association norms. A small fixed offset, ε (set to 10^{-5}), is added to make $C(i)$ nonzero for words that do not associate with any of the cues.

The probability of guessing a word with rank i is proportional to the product of the combined associative strength and the probability that quitting did not occur prior to search reaching that word:

$$P_{\text{guess}}(i) \propto (1-p)^{i-1}C(i)^q,$$

where q is a parameter reflecting the relative importance of the cue associations in the guessing process. Because a guess is needed only if the correct answer is not recognized, the probability of guessing word i is as follows:

$$P_{\text{guess}}(i) = (1 - P_{\text{recognized}}) \frac{(1-p)^{i-1}C(i)^q}{\sum_{j=1..N}(1-p)^{j-1}C(j)^q},$$

where N is the total number of words in the lexicon (5,018 in this case). All incorrect words ($i \neq k$) are produced only via guessing, whereas the probability of reporting the correct answer is given by

$$P(k) = P_{\text{recognized}} + P_{\text{guess}}(k),$$

which captures the possibility that the correct answer might be given as a guess. Collectively, these equations define a discrete probability distribution for responding with any word in the lexicon.

The parameters p and q were estimated by fitting the experimental data reported in Table S1 (i.e., both correct and incorrect responses). These values were optimized to maximize the log likelihood (LL), which is computed as follows:

$$LL = \sum_{r=1..48} \sum_{i=5,018} O_r(i) \log[P_r(i)].$$

LL thus ranges over the 5,018 possible responses to each of the 48 RAT questions. For RAT question r, $P_r(i)$ denotes the predicted response probability of the word with rank i, according to the equations of the frequency-biased model we just

outlined, and $O_r(i)$ denotes the observed response probability for that word.

Results

Experimental results

Table S1 shows all responses to all 48 RAT questions. The questions received an average of 35 different responses (SD = 9.93), and there were large accuracy differences across the 48 questions (M = .32, SD = .21). Accuracy varied greatly across participants as well (M = .32, SD = .12). The mean response time until the first key press was 11,913 ms (SD = 4,113) for correct responses and 15,301 ms (SD = 2,258) for incorrect responses.

According to our theory, individuals who perform poorly on the RAT should be biased to give high-frequency responses. As predicted by this account, there was a negative correlation across individuals between average accuracy and average AF of the words given as responses, $r(130) = -.45$, $p < .01$. However, this correlation is potentially misleading because the average AF of responses by people with high accuracy largely reflects the frequency of the correct answers, and the average AF of the correct answers was 2.31, which is relatively low. The key question was whether there was a correlation across individuals between average accuracy and average AF of incorrect responses specifically. According to our hypothesis, the reason that people perform poorly is that they are biased to consider high-frequency incorrect words, which blocks access to low-frequency correct answers (i.e., the correct answer is further down the search list). As predicted, there was a negative correlation across individuals between average accuracy and average AF of incorrect responses, $r(130) = -.26$, $p < .01$. This result supports the claim that a tendency to consider high-frequency words impairs performance on the RAT.

Results for the norm-based model

Figure 1 summarizes the performance of the norm-based model for the 178 RAT questions for which all three cues and the correct answer were in the association norms. Specifically, the graph shows the distribution of questions for a performance measure based on similarity to the average of the three cues (i.e., the cue centroid). Performance for each question was the percentage of the 5,018 words in the association norms that produced lower similarity values than the similarity value of the correct answer. As the figure shows, for the majority of the RAT questions, no more than 5% of the words in the WAS were closer to the centroid than was the correct answer. Average performance of the norm-base model was 91.1% across all 178 RAT questions and 99.5% for the 48 RAT questions used in our experiment. Because the answer to a RAT question is associatively related to each of the cue words, it is possible that a single randomly chosen cue word from the set of three cue words might yield even better performance than the cue centroid. To test this possibility, we calculated the model's performance on each question for each of the three cues separately and found that average performance dropped to 83.1%. Thus, a very good strategy when performing the RAT is to use the cue centroid as the basis for identifying the answer to each question.

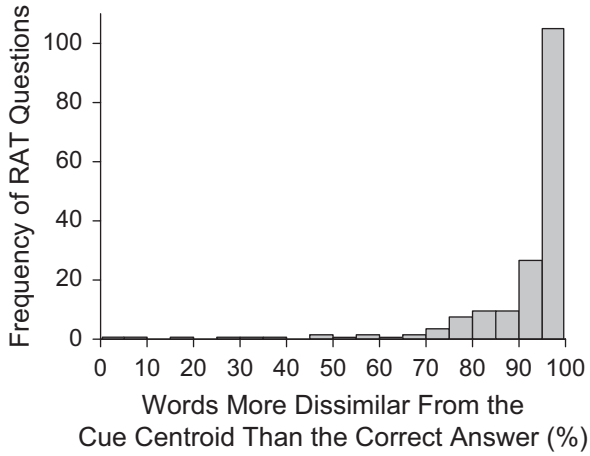

Fig. 1. Results for the norm-based model for 178 remote associates test (RAT) questions. The model's performance was based on the average of the semantic positions of the three cue words (i.e., the cue centroid) within the Word Association Space (WAS) created by Steyvers, Shiffrin, and Nelson (2004). All 5,018 words contained in the norms of Nelson, McEvoy, and Schreiber (1998) were ordered according to their similarity to the cue centroid (i.e., the cosine of the angle between the cue centroid and each word in the WAS). Performance of the model was measured as the percentage of the 5,018 words that had lower similarity values than the similarity value of the correct answer. The graph shows the frequency distribution of the 178 questions for this performance measure.

Results for the frequency-biased model

We assumed that individual differences arise from the parameters F and R, which respectively capture the extent to which individuals are biased to consider high-frequency words and the ability of individuals to recognize the correct answer if it is reached during the search process. In particular, individuals with a high value of F are those who place high-frequency words closer to the top of their search list compared with low-frequency words, and consequently tend to have insufficient time to consider a low-frequency correct answer. Figure 2 shows a specific example of how this reordering of candidate words can affect response probabilities. This example concerns the question that was the most difficult (2% accuracy): the one with *panel* as the correct answer in response to the cues *jury*, *door*, and *side*. If an individual uses the cues without regard to word frequency (F = 0), then *panel* is near the top of the search list, as indicated by the solid line in the figure. However, *panel* is a very low-frequency word (AF = 0.10), and, as shown by the dotted line, the observed response probability was lower for *panel* than for several other answers, which suggests that *panel* was placed farther down the search list than those other words. In contrast, the word *room* associates to some of the cue words and is much higher in frequency

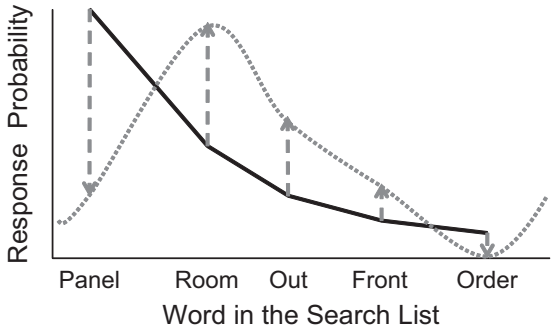

Fig. 2. Illustration showing how a bias toward high-frequency words affects performance on the remote associates test (RAT). Response probabilities are shown for the most difficult RAT question in the experiment: with the cues *jury*, *door*, and *side*, and the answer *panel*. The words along the *x*-axis are a representative selection of the words near the top of a search list based on similarity to the cue centroid (i.e., the norm-based value that should be used to order the search list). The solid line is proportional to the words' similarity to the cue centroid, and the dotted line is proportional to the observed response probabilities for these same words. The values for the solid and dotted lines are rescaled in the figure to place them on the same *y*-axis. The difference between these two lines (dashed arrows) can be accounted for by the frequency-biased model. In this model, the frequency values for candidate words are used to reorder the search list, with high-frequency words placed toward the beginning of the list. The rank-order position of the words in this list is the important determinant of performance. Thus, because *panel* is very low frequency, the observed response probability for this word is low even though this word is very similar to the cue centroid; in contrast, *room* is very high frequency, so this word's observed response probability is high even though this word is not as similar to the cue centroid.

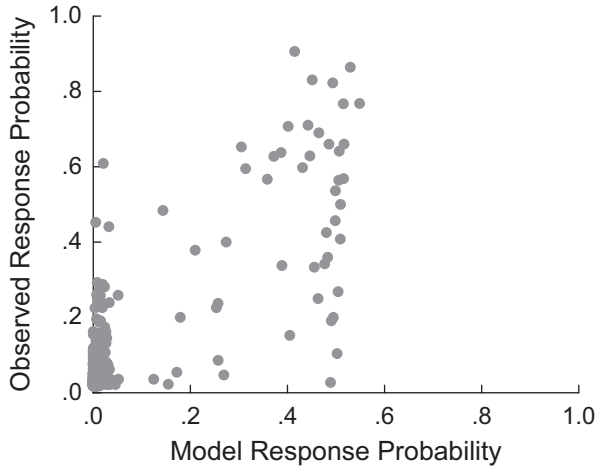

Fig. 3. Scatter plot showing the relation between observed response probabilities and the frequency-biased model's response probabilities for the 48 correct responses and the 499 incorrect responses that were contained in the association norms (results averaged across all individuals). The model parameters *p* and *q* were fit to the data using maximum likelihood. The model was at least 10% correct for all 48 RAT questions, and so the 48 correct answers are represented by the 48 points with model response probabilities greater than .1.

(AF = 3.83). Thus, for individuals with a frequency bias ($F > 0$), *room* is placed toward the top of the search list, which results in a higher probability that *room* is given as a guess.

Individual differences contributing to the group data in Table S1 were modeled by allowing a range of values across *R* and *F*. All combinations of four values of *R* (.2, .4, .6, and .8) and four values of *F* (0, 1, 2, and 3) were used to create 16 hypothetical individuals. The model produced response probabilities for all 5,018 words in the lexicon for each hypothetical individual on each of the 48 RAT questions. These probabilities were averaged across the hypothetical individuals to produce a response distribution across words for each RAT question. A comparison between predicted and observed response probabilities was then made for all of the 547 different words in Table S1 that also appear in the association norms. The same values of *p* (.002) and *q* (0.43), as determined by a maximum likelihood fit of the data, were used for all 16 hypothetical individuals. These parameters, respectively, captured search depth and how strongly the guessing process was driven by association with specific cues. Figure 3 is a scatter plot showing the relation between observed response probabilities and the model's response probabilities for the 48 correct answers, as well as the 499 different incorrect responses. The percentage of variance accounted for across both correct and incorrect responses was 65%; across the correct responses specifically, the percentage of variance accounted for was 21%. The log likelihood of the model was –171, which corresponds to a chi-square of 180, an impressively good fit ($p = 1$) considering that this is a fit of 547 different categories of response with only two free parameters.

Model comparison

By itself, an impressive fit does not constitute evidence in favor of a model (e.g., Roberts & Pashler, 2000), and consideration of alternative models is a critical component of formal modeling. Thus, we compared the frequency-biased model with one that included individual differences in neither *R* nor *F* or in only the *R* parameter. The former is equivalent to the norm-based model when *F* is set to 0 and *R* to 1 (i.e., an individual who is unbiased by frequency and who never fails to recognize the correct answer when it is considered). For this model, with these *F* and *R* parameter values, the optimal values of *p* and *q* were .003 and 0.34, respectively. This model produced a chi-square of 328—an increase of 148 over the chi-square for the frequency-biased model, even though the two models have the same number of free parameters. As noted, the frequency-biased model accounted for 21% of the variance for the 48 correct answers. However, only 2% of the variance for the correct answers was accounted for when *F* was set to 0. Given that the frequency-biased model becomes the norm-based model as regards the 48 correct answers when *F* is 0,[2] this result indicates that similarity to the cue centroid does not explain why some RAT questions are easy and others are difficult.

To allow nested model comparisons, we considered a model with 4 hypothetical individuals for whom *F* was set to 0 and *R* was set to .2, .4, .6, or .8. This model produced a chi-square of 204—an increase of 24 over the chi-square for the frequency-biased model, but a decrease of 124 compared with the norm-based model. This result is evidence that individuals

differ both in their ability to recognize the correct answer and in their tendency to consider high-frequency responses. We note that there may be different specific model implementations of these two effects that are equally compatible with our theory. For instance, consider a model in which participants all place high-frequency words at the top of their search list (i.e., high *F*), but differ in their latency to reject incorrect responses, which determines whether they will have time for the search process to reach low-frequency words further down the search list. We implemented this alternative by assuming different values for the search parameter (*p*) while a single *F* value was optimized to the data. This alternative was nearly identical to the frequency-biased model that included individual differences in *F* and fit the data nearly as well. However, this alternative is merely a different technique for implementing individual differences in the proportion of high- versus low-frequency words that are considered within the allotted time, and so this is also a frequency-biased model.

Individual differences

Although there were insufficient data to apply the frequency-biased model to each individual's responses, we examined individual differences by looking at the relation between average accuracy and average frequency. Figure 4 plots these individual differences along with corresponding values for the 16 hypothetical individuals of the frequency-biased model. As the figure shows, the range of observed individual differences roughly fills in the same triangular region of individual differences hypothesized by the frequency-biased model. An individual's observed average accuracy and average frequency can be used to estimate that individual's *R* and *F* values using a linear approximation (see Fig. 5). As the figure shows, the chosen a priori range of *F* values for the 16 hypothetical individuals underestimated the range of observed *F* values (i.e., some people were even more strongly biased to give high-frequency answers than the a priori range allowed for). The range of *F* in the model could be increased, and presumably this would improve the model's fit. However, this would be a post hoc adjustment, in which case the range of *F* values should be thought of as a free parameter.

To summarize, the frequency-biased model provided the best account of the data. According to this model, individuals who perform poorly on the RAT might or might not be biased to consider high-frequency words (i.e., there are individual differences in *F*). If they are biased to consider high-frequency words, they will produce incorrect guesses that are high-frequency words. If they are not biased to consider high-frequency words, their low accuracy is explained by low *R* (failure to recognize that the correct answer satisfies the task requirements). However, to perform well, they must have both a high value of *R* and a low value of *F*. That is, they must be able to recognize that a potential answer meets the specified requirements, and, more important, they must not be biased to consider only high-frequency candidate answers.

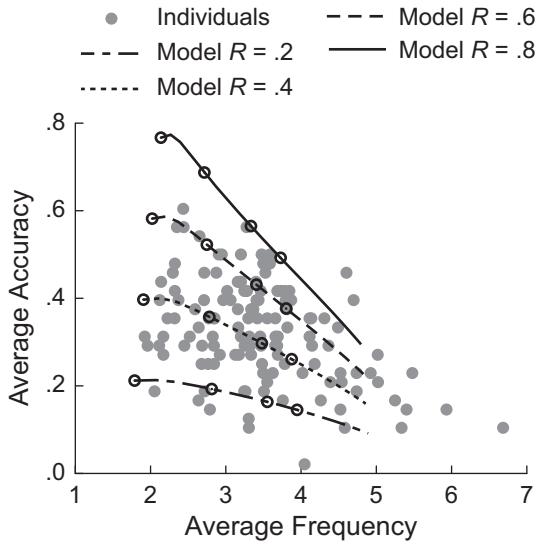

Fig. 4. Individual differences as revealed by the relationship between average accuracy on the 48 remote associates test (RAT) questions and average associate frequency of responses, collapsed over correct and incorrect responses. Each of the 132 filled dots represents the observed behavior of a participant. The open circles represent the 16 hypothetical individuals in the frequency-biased model: all combinations of four values of *F* (0, 1, 2, and 3) and four values of *R* (.2, .4, .6, and .8). *F* captures individual differences in bias to consider high-frequency words, and *R* captures individual differences in recognizing whether a candidate word associates to all three cues. For each value of *R*, the plotted line fills in model behavior for a full range of *F* values. For every combination of *R* and *F*, 48 different discrete probability distributions (1 per RAT question) over the 5,018 words in the association norms were used to calculate average accuracy and average frequency of response.

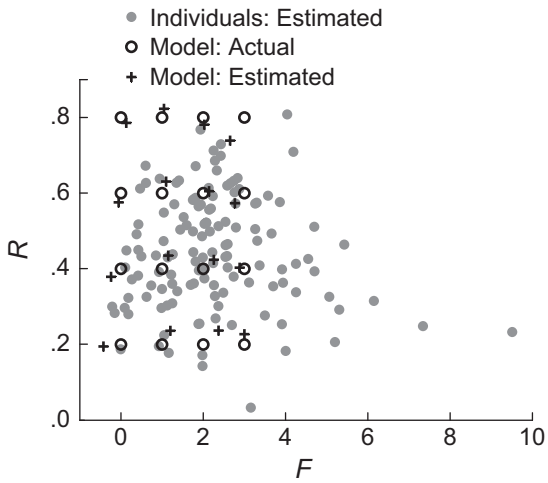

Fig. 5. Scatter plot of the *R* and *F* values estimated for each of the 132 participants. The graph also shows the *R* and *F* values of the 16 hypothetical individuals in the frequency-biased model, as well as estimates of these values. Estimates were based on a linear approximation calculated from average accuracy and average frequency, which are plotted in Figure 4. (A comparison of the actual and estimated values for the hypothetical individuals indicates the accuracy of the linear approximation.) The following equations were used: average frequency = $a + (b \times F)$ and average accuracy = $R/(1 + (c \times F))$. A least squares fit based on the 16 hypothetical individuals determined the values of the parameters: $a = 2.06$, $b = 0.63$, and $c = 0.19$. Negative values of *F* are theoretically possible, and indicate an individual who is biased to consider low-frequency responses.

Conclusions

Using the RAT as a test of creativity, we tested the hypothesis that creative solutions require avoidance of high-frequency candidate answers. We performed this test by conducting an experiment in which participants responded quickly, giving many incorrect responses. As predicted, individuals who performed poorly responded with high-frequency incorrect words. We modeled both correct and incorrect responses with a high-dimensional word association space created from association norms. This model performed well on all of the RAT questions, whereas human participants found some questions much more difficult than others. We modeled individual differences by assuming that the process of searching for the correct answer is biased by word frequency and that the correct answer may not be recognized even when considered. This model provided a good account of all responses, and it explained why some RAT questions are more difficult than others. It is unclear whether these results will generalize to creative solutions of nonlinguistic problems, but we hope that this well-specified process model of creative problem solving will promote additional tests of this account in other domains.

Acknowledgments

The first author performed the modeling work, and the second author performed the experimental work.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was supported by National Science Foundation Grant BCS-0843773 to David E. Huber.

Supplemental Material

Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

Notes

1. Throughout all of the reported analyses, the only association norms used were those of Nelson et al. (1998).

2. Because the *R* parameter serves only to rescale the response probability for correct responses versus the incorrect responses, the percentage of variance for correct responses that is accounted for is independent of *R*.

References

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

Bowers, K. S., Regehr, G., Balthazard, C., & Parker, K. (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, USA*, *106*, 10130–10134.

Fodor, E. M. (1999). Subclinical inclination toward manic-depression and creative performance on the Remote Associates Test. *Personality and Individual Differences*, *27*, 1273–1283.

Forbach, G. B., & Evans, R. G. (1981). The Remote Associates Test as a predictor of productivity in brainstorming groups. *Applied Psychological Measurement*, *5*, 333–339.

Griffiths, T. L., Steyvers, M., & Firl, A. (2007). Google and the mind: Predicting fluency with PageRank. *Psychological Science*, *18*, 1069–1076.

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

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

Kray, L. J., Galinsky, A. D., & Wong, E. M. (2006). Thinking within the box: The relational processing style elicited by counterfactual mind-sets. *Journal of Personality and Social Psychology*, *91*, 33–48.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

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.

Mednick, M. T., & Andrews, F. M. (1967). Creative thinking and level of intelligence. *Journal of Creative Behavior*, *1*, 428–431.

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

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme and word fragment norms, 1999*. Retrieved from http://w3.usf.edu/FreeAssociation/

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*, 358–367.

Sitton, S. C., & Pierce, E. R. (2004). Synesthesia, creativity and puns. *Psychological Reports*, *95*, 577–580.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 237–249). Washington, DC: American Psychological Association.

Topolinski, S., & Reber, R. (2010). Gaining insight into the "aha" experience. *Current Directions in Psychological Science*, *19*, 402–405.

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.

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

Ward, J., Thompson-Lake, D., Ely, R., & Kaminski, F. (2008). Synaesthesia, creativity and art: What is the link? *British Journal of Psychology*, *99*, 127–141.