Idea Generation and Goal-Derived Categories

Richard W. Hass (Richard.Hass@jefferson.edu), J. Colin Long, Joshua Pierce
College of Humanities and Sciences, Thomas Jefferson University
Philadelphia, PA 19144 USA

Abstract

Semantic search and retrieval of information plays an important role in creative idea generation. This study was designed to examine how semantic and temporal clustering varies when asking participants to generate ideas about uses for objects compared with generating members of goal-derived categories. Participants generated uses for three objects: brick, hammer, picture frame, and also generated members of the following goal-derived categories: things to take in case of a fire, things to sell at a garage sale, and ways to spend lottery winnings. Using response-time analysis and semantic analysis, results illustrated that all six prompts generally led to exponential cumulative response-time distributions. However, the proportion of temporally clustered responses, defined using the slope-difference algorithm, was higher for goal-derived category responses compared with object uses. Despite that, overall pairwise semantic similarity was higher for object uses than for goal derived exemplars. The effect of prompt on pairwise semantic similarity is likely the result of context-dependency of exemplars from goal-derived categories. However, the current analysis contains a potential confound such that special instructions to give "common and uncommon" responses were provided only for the object-uses prompts. The confound is likely minimal, but future work is necessary to verify that these results would hold when the confound is removed.

Keywords: Creativity; Divergent Thinking; Goal-Derived Categories; Latent Semantic Analysis; Semantic Memory

Creative cognition researchers often highlight the contributions of memory structure and process to creative idea generation. Though theories vary widely in explaining how existing knowledge is actually used to support the generation of creative ideas and products, there is sufficient evidence to suggest that in both laypeople (Ward, 2008), and in eminent creators (Weisberg, 2006) creative thinking operates within the bounds of an individual's system of knowledge. This study was designed to extend recent work (Hass, 2017a) exploring the degree of semantic clustering found among ideas generated when participants complete divergent thinking tasks. Divergent thinking tasks are heavily used as a proxy for creative thinking in a variety of behavioral (Snyder, Hammond, Grohman, & Katz-Buonincontro, 2019) and neuroscientific (Dietrich & Kanso, 2010) studies. There is a general consensus that dynamic interplay among executive search and control processes and semantic memory organization enables the generation of creative ideas (cf. Abraham & Bubic, 2015; Beaty, Christensen, Benedek, Silvia, & Schacter, 2017; Chrysikou & Thompson-Schill, 2011).

The central aim of the study was to extend prior results (e.g., Hass, 2017a; Hass & Beaty, 2018) by comparing semantic processing during object-uses generation to the generation of exemplars from goal-derived categories (Barsalou, 1985). Generating uses for objects is the core feature of the Alternative Uses task (Wilson, Guilford, Christensen, & Lewis, 1954), one of the most popular divergent thinking tasks, and its validity as a psychometric measure of creative thinking is enhanced by illuminating the underlying cognitive processes operating while people perform it. Creative thinking has also been described as related to goal-derived knowledge (Chrysikou, 2006), so it is natural to explicitly examine potential similarities between generating uses for objects and generating exemplars of goal-derived categories. The paper is structured as follows: first, research on the relationship between semantic memory retrieval and idea generation will be reviewed, along with a brief discussion of how divergent thinking tasks like object-uses generation relate to goal-derived category recall or generation tasks. Then, the analysis is presented in three phases: an analysis of cumulative response-time functions across conditions, an analysis of temporal clustering of responses, and an analysis of the semantic similarity of pairs of responses across two prompts, one from each condition.

Knowledge and creative generation

As mentioned, cognitive accounts of creativity tend to differentially emphasize the importance of associative processes of semantic organization and executive control of thought (cf. Chrysikou & Thompson-Schill, 2011; Mednick, 1962). In an early theoretical account, (Mednick, 1962) suggested that creative idea generation is underpinned by associative networks that afford more remote connections among concepts. Recent studies of creative thinking have shown support for this view, illustrating that individuals with flexible semantic networks tend to perform better on creative cognitive tasks and report a greater number of creative achievements (e.g. Kenett, Beaty, Silvia, Anaki, & Faust, 2016). Additional studies have highlighted the influence of executive control on the remote association process. For example (Beaty, Silvia, Nusbaum, Jauk, & Benedek, 2014) showed that the fluency and originality of uses for objects was almost equally well predicted by measures of remote association and associative flexibility, the latter thought to be an index of executive control over lexical association. Similarly, (Hass, 2017b) showed that the degree to which creative uses for objects were semantically distant from the core of the prompt object concept was positively related to fluid intelligence.

Goal-derived Categories A key aspect of this study was to propose that goal-derived category exemplar generation can serve as a basis for understanding object-uses generation. Goal-derived categories constructed during goal-directed activities (Barsalou, 1983, 1985), like deciding which chair to sit on in a cafe or which personal belongings to keep from a childhood home. These categories can be distinguished from "natural" or "taxonomic" categories in several ways, though, we focus on two. First, goal-derived categories are constructed when performing decision-making tasks; defined by personal objectives and constrained by the environment or immediate context. Second, it can be argued that goal-derived categories such as "things to take with you in case of a fire" are not as well-established in memory as categories such as breakfast foods (Barsalou, 1983). Omelets and pancakes are within the same category (breakfast foods) because they are edible, made with eggs, served warm, are eaten in the morning, and relatively straightforward to cook. Attributes like those, such as times of consumption and ingredients, which are the basis of category discrimination (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), do not co-occur as frequently in goal-derived categories. In addition, goal-derived categories are not as well used in the literature on semantic search and retrieval, so this analysis provides novel information about memory search dynamics when people name members of goal-derived categories. More importantly for this analysis, Barsalou (1985) suggested that retrieval of conceptual information for category processing involves generation of multiple conceptual representations, each held in working memory, when the category is encountered in normal life. This reliance on multiple conceptual representations could account for the effects reviewed above relating object-uses generation to fluid intelligence. Thus, comparing goal-derived category search to object uses search serves the dual purpose of exploring how context-dependent organization and executive control might interact during idea generation.

The current study

The primary focus of the current analysis was on semantic clustering. Because there are no established category norms for the prompts used in this study (cf. Troyer, Moscovitch, & Winocur, 1997), clusters were first identified using the slope-difference algorithm (Gruenewald & Lockhead, 1980). Latent semantic analysis (Landauer & Dumais, 1997) was then used to quantify the semantic similarity among sequential pairs of responses. The slope-difference algorithm identifies potential semantic clusters in terms of the difference between an actual IRT and the expected IRT given a mathematical relation between response time and output total. It was expected that slope-difference clusters would be more prevalent in the goal-derived response arrays, and that the pairwise semantic similarity of within-cluster responses would be higher in goal-derived response arrays. The reasons to expect that goal-derived response arrays would be more clustered than object-uses arrays are two-fold. First, response totals are usually quite low when people generate uses for objects, and though clusters appear, the number of responses per cluster is usually small. As cluster size decreases, output total should follow (Herrmann & Pearle, 1981), and the lack of success in finding newly retrieved clusters will ultimately lead to search termination (Raaijmakers & Shiffrin, 1981). Second, memory search is often described as a multiply-constrained problem (e.g., Polyn, Norman, & Kahana, 2009; Smith, Huber, & Vul, 2013), with multiple sources of information vying for attention in the process. It is plausible that goal-derived category generation is less constrained than object-use generation, such that a single context-dependent goal (e.g., "items to sell at a garage sale") remains in mind. This should enable the integration of contextual and semantic information in more efficient manner than in object-use generation, where the goal may change from response to response (e.g., "use a brick as a weight" $\rightarrow$ "use a brick as a pencil holder").

As will be described, the prompt used for object-uses in the current study was to "think of common *and* uncommon uses", designed to provide a more natural comparison to the generation of category exemplars (i.e., the word "creative" was not used in the instructions). That is, several studies have shown that instructing participants to "be creative" decreases fluency (output total), while increasing the average originality of their responses (Forthmann et al., 2016; Nusbaum, Silvia, & Beaty, 2014). Since the primary interest in the current study was the nature of the category itself (e.g., use of an object v. goal-derived category) and *not* whether participants were trying to engage in creative thought, we felt the special instruction was warranted. However, as we discuss, the inclusion of this "common and uncommon" instruction was not used in the goal-derived conditions, which presents a confound. The nature of our results do not suggest the confound is serious, it is important to keep in mind.

Method

Participants

A total of 32 participants were recruited from undergraduate psychology courses. Participants were offered extra credit or chocolate in compensation for their time. Participants ranged in age from 18 to 25 years old, and the demographics were consistent with a traditional undergraduate university in the northeastern United States. All recruitment and consent procedures were approved by the university's Institutional Review Board.

Materials

Participants completed the tasks using a custom Matlab interface on an Apple iMac. Instructions and prompts appeared as text on white background above a text-box where participants entered responses. Instructions were displayed and read to participants prior to each of three task blocks, the first

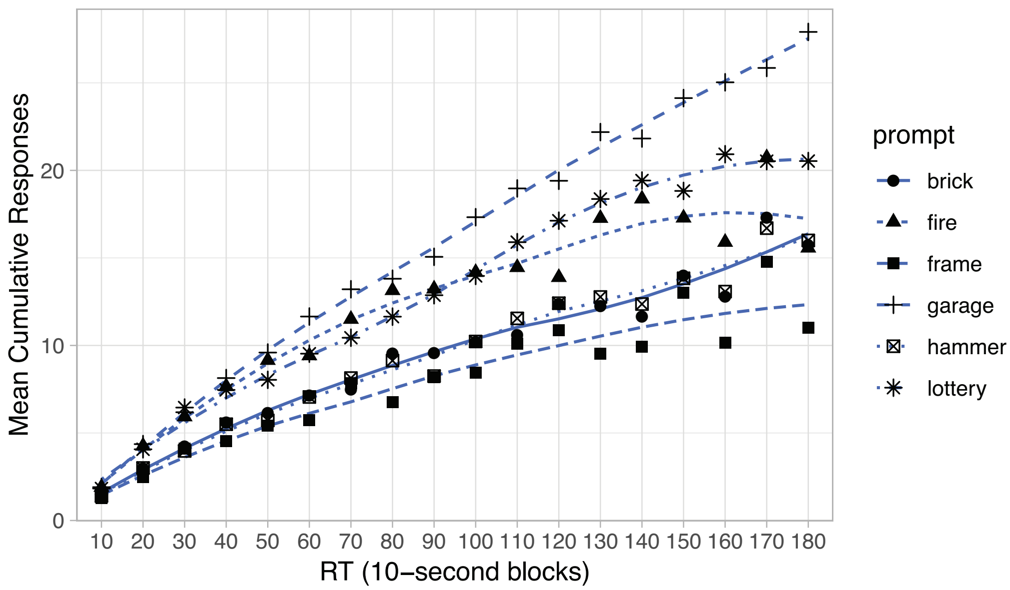

Figure 1: Mean number of cumulative responses per 10-second blocks for each of the six prompts. Object-uses prompts: brick, (picture) frame, hammer; Goal-derived prompts: fire = things to take in case of fire, garage = things to sell at a garage sale, lottery = ways to spend lottery winnings. Note: lines represent loess fits, not the exponential fits used in the next section.

of which was a practice block (naming colors). Instructions were not visible during response generation, but the prompt remained displayed for the entire duration of each response-generation interval. Demographic information was obtained using a pencil-and-paper survey after the experiment finished.

Procedure

Participants were greeted by the experimenter, and were told that the experiment was designed to test memory. The experimenter read general instructions about how the computer system worked and instructed participants to type responses on the computer keyboard, and to enter responses by pressing the return button. Participants then practiced this by naming colors (at least 3) for 30 seconds. The experimenter then answered any questions before the experiment began. Matlab recorded the time of the first keypress of each response, the time between the first key-stroke and the response entry, and the actual response.

The tasks were presented in two blocks of 3 prompts each, with a break in between each block. Both the order of the blocks and the order of presentation of prompts within the blocks were randomized by Matlab code. All participants responded to each prompt in each block. Each response interval was three minutes in length to permit valid comparison among the two prompt conditions (goal-derived categories, and object-uses prompts). The goal-derived category prompts began with “name examples of” and ended with one of the three categories: things to spend lottery winnings on, things to take from your house if it caught on fire, and things to sell at a garage sale. The object-use prompts began with “name common and uncommon uses for a” and ended with one of the three prompt objects: brick, hammer, and picture frame.

The entire prompt phrase remained on the screen above the text-entry box for the entire 3 minutes. When 3 minutes expired, the screen displayed a message indicating that the next prompt was loading for 5 seconds before the next prompt appeared. After the first and second blocks, instructions for the next block appeared on the screen, and the participant was given a 1-2 minute break before beginning the next block. After the final block, a thank-you message appeared and the participant filled out the demographic questionnaire, and the experimenter answered any questions the participant may have had. The entire process lasted between 20 and 25 minutes for each participant.

Analyses and Results

All analyses were conducted using the R Statistical Programming Language, and all data and algorithms are available for download (https://osf.io/fvne2/). Response times were defined in terms of the time (since presentation of the prompt) of the first key-press of each response, to be consistent with studies using voice-keyed response recording. Prior to analysis, data were examined for repeated responses and malfunctions in Matlab’s execution of the experiment. Three participants were excluded due to Matlab malfunctions reducing the final sample size to 29. Repeated responses were those that were identified as the same response given more than once by the same individual to a specific prompt. When repeats were identified, the RTs for those responses were removed from the data set. Participants gave a total of 1746 responses to the three goal-derived prompts after the removal of 23 repeated responses. Finally, participants gave a total of 1012 responses

Table 1: Average response totals per category and intercorrelations (Spearman's ρ). Object Uses prompts are in the top half of the table. All correlations are significant, with $p \leq .01$, except the correlation between garage sale and hammer totals ($p = .06$).

			ρ				
Prompt	M	SD	1	2	3	4	5
1. Brick	12.76	4.66	-				
2. Hammer	11.86	5.09	.64	-			
3. Frame	10.28	4.40	.72	.78	-		
4. ... sell at garage sale	24.17	7.40	.35	.67	.46	-	
5. ... take from fire	17.17	5.33	.67	.52	.48	.53	-
6. ... do with lottery winnings	18.86	6.69	.50	.57	.46	.65	.55

to the object-use prompts after the removal of 11 responses.

Fluency and Cumulative Response Times

The mean number of cumulative responses was computed in 10-second blocks and plotted in Figure 1. Clearly there is nonlinearity, and not surprisingly, fluency is higher for the goal-derived prompts compared with the object-use prompts. The shape of the distributions in these plots is consistent with those found in normal memory retrieval studies. Table 1 further illustrates that response totals are uniformly lower for objects uses prompts, and that there is a relatively large degree of correlation among output totals.

Clustering

Clusters were identified using a modification the Slope Difference Algorithm (Gruenewald & Lockhead, 1980), that uses an exponential function rather than the hyperbolic function used by Gruenewald and Lockhead:

$$R(t) = N(1 - e^{-\frac{t}{\tau}}) \quad (1)$$

This is the "two parameter" exponential, with N being the estimated asymptote, or number of responses generated with an unlimited amount of time, and τ being the inverse of the rate parameter λ in an exponential distribution. Thus, τ is parameterized as the estimated mean response time.

The Slope-Difference algorithm works as follows: given the estimated N and τ parameters for each participant, calculate the difference between the predicted and observed instantaneous rates of change in responding. The predicted rate of change is just the derivative of Equation 1 calculated with each participant's parameters and the cumulative response times of that participant. The observed instantaneous rate of change is just the reciprocal of each inter-response time (IRT) (i.e., for R = cumulative number of responses, $\frac{\Delta R}{\Delta t} = \frac{1}{IRT}$). Gruenewald and Lockhead (1980) provided support for the validity of the algorithm, such that large, positive differences between observed and predicted rates were indications that responding was faster than predicted, and thus, faster than expected responses qualify as being within clusters. The threshold for slope-differences being categorized as "switches" was .10, which is the same as used by Grunewald and Lockhead.

To obtain slope-difference values, Equation 1 was first fit to each participant's cumulative response-time distribution *per prompt*, using ordinary nonlinear least-squares estimation. Parameter values along with response times were used to compute predicted rates of change to be differenced from the actual rates of change. Clusters were then identified as any IRT with a slope difference value less than .10, the threshold used by Gruenewald and Lockhead (1980). Exponential parameter estimates were not optimal for 1-3 participants per prompt, and data from those participants were excluded for the cluster analysis of each prompt.

Figure 2 shows that the proportion of responses identified as within cluster was significantly greater for goal-derived categories compared with object uses, $\chi^2(1) = 48.27, p < .001$. Of the 998 object-use responses, 18.8% were identified as within-cluster, while 31.3% of the 1567 goal-derived responses were identified as within-cluster.

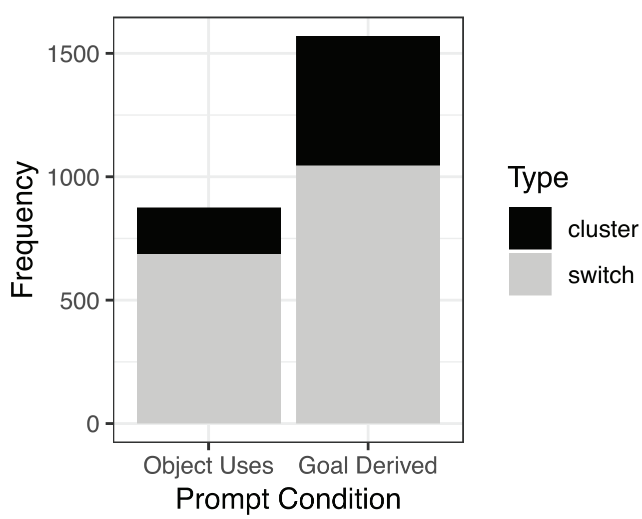

Figure 2: Number of responses classified as within cluster, or as a switch between clusters by the slope difference algorithm for the two types of prompts. See text for proportions.

Pairwise Semantic Similarity

The validity of the slope difference algorithm rests on further semantic analysis of its results. Here, the central question was whether responses in clusters corresponding to goal-derived category were more semantically similar than those in clusters corresponding with object-use generation. Pairs of sequential responses were analyzed for semantic similarity using the tools at the UC Boulder website (lsa.colorado.edu). The General Reading corpus, with 300 factors, was chosen as the basis for comparisons, and the term-to-term comparator was used.

Mixed-effects regression was used to examine the main-effects of clustering (within cluster v. between cluster response) and prompt-type (goal-derived v. object use) on LSA-derived cosine similarities, and the interaction of the two fixed effects. A random intercept term was added to account for participant variation, and another to account for variations across the 6 prompts. Table 1 contains the results of the analysis including 95% confidence intervals for the fixed and random effects terms. Rather surprisingly, on average the pairwise similarity of responses to the goal-derived prompts was less than the average pairwise similarity of object-uses responses. However, the slope difference algorithm seems to distinguish between semantic clusters such that on average, within-cluster responses were less similar (in terms of pairwise similarity) than between cluster responses. Figure 3 illustrates that there may be a small interaction between prompt type and clustering, and in Table 2, the estimate is a slightly smaller difference in similarity of clustered and non-clustered responses for object uses compared with goal derived categories, though zero remains a plausible value for the interaction.

A slightly different result is obvious if one plots pairwise similarity as a function of IRT. Figure 4 shows that, at the level of individual pairs of responses, the relationship between IRT and similarity is not linear, and that for a great many pairs of responses on all six prompts, there is a substantial degree of variability in pairwise similarity for short IRTs. A closer look at Figure 4 reveals that the garage prompt has the highest concentration of low-similarity pairs. This an interesting result on its own and is likely the result of context dependency for that prompt, as will be discussed next.

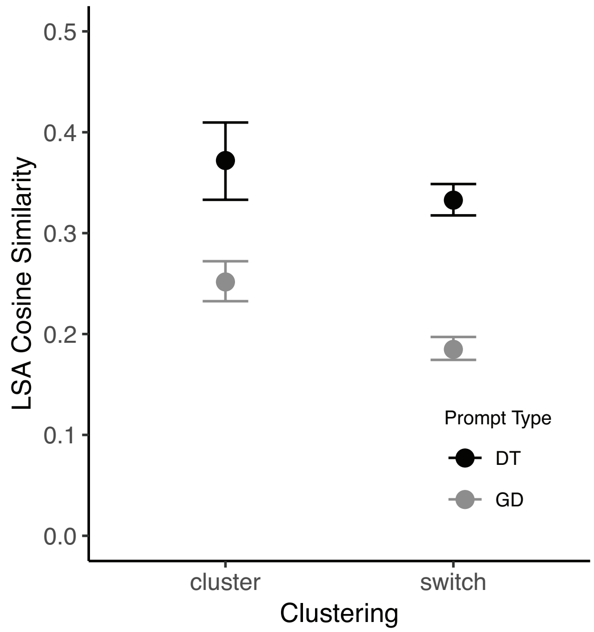

Figure 3: Average pairwise semantic similarity per prompt and per cluster category (in cluster v. switch). Bars are bootstrapped 95% confidence limits.

Discussion

The purpose of this study was to probe the differences between object-uses generation and goal-derived category exemplar generation in terms of semantic search and retrieval. Using measures of clustering and similarity, this analysis illustrated that there may be two key differences between responding to these two types of prompts. First, output totals for goal-derived categories were much higher than object uses, and also contained a greater proportion of faster than expected IRTs, identified by the slope difference algorithm.

Though semantic analysis of adjacent pairs of responses showed that the slope-difference clusters are indeed semantic clusters, pairwise similarity among goal-derived exemplars was surprisingly less than the similarity among adjacent pairs of object-use responses. One explanation for this result is that semantic relationships *across* clusters of goal-derived exemplars may be minimal because of the dependence of semantic similarity on context (e.g., Barsalou, 1982). Of course, that characterization might also be said of object-uses. More importantly, Hass (2017a) illustrated that LSA-derived cosine similarities may not accurately represent context-dependent relationships between object uses. Indeed, the main difference between the two types of prompts is that goal-derived prompts identify a context (e.g., a garage sale), which all items must relate to in some way, while object-uses prompts identify an exemplar (e.g., a brick) to which responses must relate. While object-uses responses will likely have context-dependency, it is also likely that context dependency will be greater among goal-derived categories such as those in this study, as the context itself is the main constraint on conceptual activation. That is, consider the example discussed in the introduction: electronics to sell at a garage sale. Say a participant activates electronics as a concept and exploits it for a bit, what is the likelihood that the next conceptual representation activated will be highly similar to electronics in a context-independent fashion? Contrast that with the activation of the attribute “heavy” in generating uses for a brick. What is the likelihood that the next conceptual representation used after “heavy” is going to be semantically similar to “heavy” in a context-independent sense. It seems plausible that the semantic similarity among all activated concepts

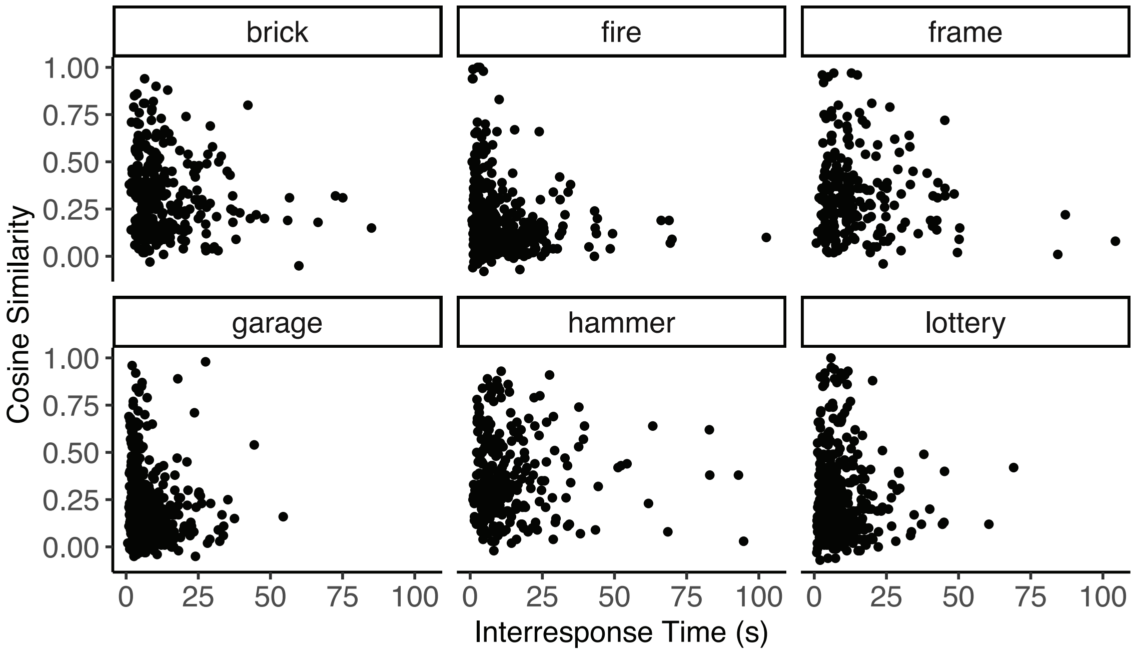

Figure 4: Scatterplots of the IRT-similarity relationship across all size prompts. Object uses prompts are in the top row, goal-derived prompts are in the bottom row

Table 2: Results of the mixed-effects regression with pairwise similarity as the dependent variable. The baseline prompt-type condition was object-use, the baseline cluster condition was Within Cluster.

			Confidence Interval	
Fixed Effects	b	*t*	2.5%	97.5%
Intercept	0.38	14.65	0.33	0.42
Prompt Type	-0.12	-3.88	-0.18	-0.06
Cluster	-0.04	-2.24	-0.08	-0.005
prompt $\times$ switch	-0.03	-1.49	-0.07	0.01
Random Effects	σ		2.5%	97.5%
Participant	0.04		0.03	0.06
Prompt	0.03		0.01	0.05
Residual	0.20		0.19	0.20

in the garage context might be lower than the semantic similarity among activated concepts in the context of a use for a brick because of the dependence on the garage context. The latter conclusion is still highly speculative, but it suggests that this is a fruitful avenue for future research to follow, as it will likely illuminate how semantic information is organized and used in both kinds of tasks.

Limitations and future directions

In this study, participants were explicitly instructed to *think of common and uncommon uses for objects* in an effort to obtain a greater total number of responses generated across the object uses prompts (i.e., the word "creativity" was not present in the instructions). In the recent study by Hass (2017a), participants were instructed to think of *creative* uses for objects, and indeed, their response totals were, on average, lower than the current study (about 7 responses). So it is likely that the instruction to be creative may limit the semantic similarity of clustered output when generating object uses. Indeed, the major motivation for the choice to avoid the word creative was to provide a baseline for future studies that would vary instructions, including "be-creative" conditions (e.g., Forthmann et al., 2016), and strategy inductions (e.g., Unsworth, Brewer, & Spillers, 2013). However, since participants were only given the "common and uncommon" instructions in one condition, the effect of prompt type on semantic similarity is confounded by the differing instructions. Specifically, our use of the phrase "common *and* uncommon" uses in the object use condition may have confused participants, or led some participants to approach the task differently from others, with some potentially assuming that they should be creative, or only think of uncommon uses. We believe that this can be remedied in future studies by changing all prompts to be of the form, "think of things to ..." and then appending the prompt (e.g., ... to sell at a garage sale; ... to do with a brick). Participants can then be instructed to perform the two tasks in the ways just mentioned (e.g., creatively, or using a certain search strategy), without the confound currently present. However, the current results are still informative, and it is likely that the confound presented by the "common or uncommon" phrasing was minimal.

References

Abraham, A., & Bubic, A. (2015, March). Semantic memory as the root of imagination. *Frontiers in Psychology*, *6*, 1–5.

Barsalou, L. W. (1982, January). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*(1), 82–93.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211–227.

Barsalou, L. W. (1985, October). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 629–654.

Beaty, R. E., Christensen, A. P., Benedek, M., Silvia, P. J., & Schacter, D. L. (2017, March). Creative constraints: Brain activity and network dynamics underlying semantic interference during idea production. *NeuroImage*, *148*(C), 189–196.

Beaty, R. E., Silvia, P. J., Nusbaum, E. C., Jauk, E., & Benedek, M. (2014, June). The roles of associative and executive processes in creative cognition. *Memory & Cognition*, *42*(7), 1186–1197.

Chrysikou, E. G. (2006). When Shoes Become Hammers: Goal-Derived Categorization Training Enhances Problem-Solving Performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 935–942.

Chrysikou, E. G., & Thompson-Schill, S. L. (2011, April). Dissociable brain states linked to common and creative object use. *Human Brain Mapping*, *32*(4), 665–675.

Dietrich, A., & Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological Bulletin*, *136*(5), 822–848.

Forthmann, B., Gerwig, A., Holling, H., Celik, P., Storme, M., & Lubart, T. (2016, July). The be-creative effect in divergent thinking: The interplay of instruction and object frequency. *Intelligence*, *57*, 25–32.

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

Hass, R. W. (2017a, September). Semantic search during divergent thinking. *Cognition*, *166*, 344–357.

Hass, R. W. (2017b, October). Tracking the dynamics of divergent thinking via semantic distance: Analytic methods and theoretical implications Richard W. Hass. *Memory & Cognition*, *45*, 233–244.

Hass, R. W., & Beaty, R. E. (2018). Use or consequences: Probing the cognitive difference between two measures of divergent thinking. *Frontiers in psychology*, *9*, 2327.

Herrmann, D. J., & Pearle, P. M. (1981). The proper role of clusters in mathematical models of continuous recall. *Journal of Mathematical Psychology*, *24*(2), 139–162.

Kenett, Y. N., Beaty, R. E., Silvia, P. J., Anaki, D., & Faust, M. (2016). Structure and Flexibility: Investigating the Relation Between the Structure of the Mental Lexicon, Fluid Intelligence, and Creative Achievement. *Psychology of Aesthetics, Creativity, and the Arts*, *10*, 377–388.

Landauer, T. K., & Dumais, S. T. (1997, April). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

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

Nusbaum, E. C., Silvia, P. J., & Beaty, R. E. (2014). Ready, Set, Create: What Instructing People to "Be Creative" Reveals About the Meaning and Mechanisms of Divergent Thinking. *Psychology of Aesthetics, Creativity, and the Arts*, *8*(4), 423–432.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134.

Rosch, E. H., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Smith, K. A., Huber, D. E., & Vul, E. (2013, July). Multiply-constrained semantic search in the Remote Associates Test. *Cognition*, *128*(1), 64–75.

Snyder, H. T., Hammond, J. A., Grohman, M. G., & Katz-Buonincontro, J. (2019). Creativity measurement in undergraduate students from 1984–2013: A systematic review. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 133–143.

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

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013, June). Strategic search from long-term memory: An examination of semantic and autobiographical recall. *Memory*, *22*(6), 687–699.

Ward, T. B. (2008, November). The role of domain knowledge in creative generation. *Learning and Individual Differences*, *18*(4), 363–366.

Weisberg, R. (2006). *Creativity: Understanding innovation in science, problem solving, and the arts*. Wiley.

Wilson, R. C., Guilford, J. P., Christensen, P. R., & Lewis, D. J. (1954). A factor-analytic study of creative-thinking abilities. *Psychometrika*, *19*(4), 297–311.