Tracking the dynamics of divergent thinking via semantic distance: Analytic methods and theoretical implications

Richard W. Hass[1]

Abstract Divergent thinking has often been used as a proxy measure of creative thinking, but this practice lacks a foundation in modern cognitive psychological theory. This article addresses several issues with the classic divergent-thinking methodology and presents a new theoretical and methodological framework for cognitive divergent-thinking studies. A secondary analysis of a large dataset of divergent-thinking responses is presented. Latent semantic analysis was used to examine the potential changes in semantic distance between responses and the concept represented by the divergent-thinking prompt across successive response iterations. The results of linear growth modeling showed that although there is some linear increase in semantic distance across response iterations, participants high in fluid intelligence tended to give more distant initial responses than those with lower fluid intelligence. Additional analyses showed that the semantic distance of responses significantly predicted the average creativity rating given to the response, with significant variation in average levels of creativity across participants. Finally, semantic distance does not seem to be related to participants' choices of their own most creative responses. Implications for cognitive theories of creativity are discussed, along with the limitations of the methodology and directions for future research.

Keywords Creativity · Semantic distance · Divergent thinking · Fluid intelligence · Growth models

Perhaps the most common task paradigm for investigating creativity across psychological domains is the verbal divergent-thinking task. The best-known example is the alternative-uses task, during which a participant is given a prompt object (e.g., a brick) and asked to generate as many alternative[1] uses for the object as possible (i.e., uses *other* than for building a house or a wall) within some unit time. Task performance is generally quantified in terms psychometric summary scores: fluency, the number of nonrepeated items per participant; flexibility, the number of category switches among the responses; and originality, scored either in terms of the infrequency of the participants' responses within the sample or in terms of the judgments of a team of raters. In responding to a "brick" prompt, a participant might generate the following set of responses: to build a bridge, as a paperweight, or as a decoration. This participant's fluency score would thus be 3; according to the coding scheme for flexibility, the number of switches would be 2; and the participant's originality score would be some function of the relative frequencies or ratings of each of the responses.

In this article, I argue that the psychometric approach to divergent-thinking task analysis is limited in its ability to answer fundamental questions about the cognitive underpinnings of divergent thinking, and of creativity more broadly defined. A central question in both the creativity and cognitive psychology literatures is how existing knowledge is operated upon to produce novel solutions to common problems (e.g., Chan & Schunn, 2015; Chrysikou & Weisberg, 2005; Ward, Smith, & Vaid, 1997; Weisberg, 2006; Weisberg & Hass, 2007). The existing systems for summary scoring described above cannot, in themselves,

✉ Richard W. Hass
hassr@philau.edu

[1] Philadelphia University, Philadelphia, PA, USA

---

[1] The word "alternative" is used here as is standard in the current literature, though the task instructions vary across studies, including instructions to "be creative" (Nusbaum, Silvia, & Beaty, 2014), "give unusual responses," and so forth.

illuminate the process by which participants produce *each* response. As an example of the power of moving beyond summary scores, Gilhooly, Fioratou, Anthony, and Wynn (2007) used think-aloud protocols to track participants' retrieval strategies while they generated alternative uses for several objects. Protocol analyses revealed that early in response generation, people seemed to retrieve experienced uses (e.g., using a hanger to scratch your back), "unmediated" (p. 616) by long-term memory, but later in response generation, people generated either broad semantic uses (e.g., using a shoe as a weapon), uses based on the properties of an object (e.g., using a brick as a pillow), and in some cases disassembling the object and using its constituent parts (e.g., using only the laces of a shoe). This suggests that during alternative-uses tasks, and likely in divergent-thinking tasks more broadly, participants' cognition is dynamic. This is contrary to Guilford's (1967) psychometric explanation of divergent thinking as a unitary ability, and suggests that divergent-thinking tasks do not simply involve thinking in "multiple directions." Rather, response generation during divergent-thinking tasks might better be explained in terms of the dynamics of semantic memory retrieval.

As such, the central goal of this study was to probe whether the semantic analysis of divergent-thinking responses, via latent semantic analysis (Landauer & Dumais, 1997), is an appropriate method for examining changes in the conceptual content of responses as a function of response order. Apart from the trends described in the verbal protocols collected by Gilhooly and colleagues (2007), serial-order effects have been observed in divergent-thinking tasks (e.g., Beaty & Silvia, 2012). However, this analysis was the first to examine the serial-order effect in divergent thinking using a cognitive theory of semantic memory. It also represents a methodological step toward including divergent-thinking task analysis within the purview of cognitive techniques for measuring creative thinking.

The remainder of the article is organized as follows: First, the limitations of the summary score approach to divergent-thinking task analysis will be described more completely. This will include a review of recent attempts to amend divergent-thinking methodologies by removing the stipulation of multiple responses. I will argue that this is not a necessary amendment to the task. Instead, requiring multiple responses and then examining the dynamics of the semantic content of those responses across iterations can reveal much more about the cognitive underpinnings of divergent thinking than can single-response paradigms. In making this argument, important details of latent semantic analysis, including critiques of the theory, will be explained. Finally, the implications of such a method with regard to existing neuroscientific studies of divergent thinking will be discussed.

Limitations of summary scores and single-response paradigms

Divergent-thinking tasks and the scoring systems used to quantify performance were developed in the 1950s and 1960s, and generally remain unchanged. According to one estimate, approximately 55 % of studies of childhood creativity conducted between 1962 and 2014 defined creativity in terms of some combination of these scores (Hass, Toub, Yust, Hirsh-Pasek, & Golinkoff, 2015). The percentage of studies that have used the alternative-uses task, or some other measure of divergent thinking, is likely similar across studies using adult samples, though at the time of writing no exact estimate is available. Regardless, divergent-thinking tasks such as the alternative-uses task have become prevalent in recent neuroscientific research on creativity, with conflicting results regarding the importances of different brain areas (see Dietrich & Kanso, 2010). There is some consensus about the importances of both the executive network of the brain (i.e., the dorsolateral prefrontal cortex and the anterior cingulate cortex) and the brain's so-called *default mode* network (including the medial prefrontal cortex, posterior cingulate cortex, precuneus, and temporal parietal junction, among other areas; see, e.g., Beaty, Benedek, Silvia, & Schacter, 2016; Heinonen et al., 2016; Jung, Mead, Carrasco, & Flores, 2013) for divergent thinking. In addition, many researchers have demonstrated a positive relationship between measures of fluid intelligence, especially performance on Raven's Progressive Matrices, and divergent-thinking performance (e.g., Beaty & Silvia, 2012; Nusbaum & Silvia, 2011; Prabhakaran, Green, & Gray, 2014; Silvia, 2008). However, all of these analyses were limited by an exclusive focus on summary scores of divergent thinking.

The reason that summary scores limit our understanding of the underlying cognitive processing involved in divergent thinking is that the scoring systems were not devised to quantify the cognitive processes involved in *producing* responses. Instead, they were created to differentiate between individuals on the basis of latent intellectual factors theorized to be important for creativity (e.g., Guilford, 1967; Torrance, 1979). Cognitive psychologists largely sidestepped the problems with divergent-thinking summary scoring by developing single-response paradigms (but see Jansson & Smith, 1991). These paradigms include solving insight problems (e.g., Gupta, Jang, Mednick, & Huber, 2012; Kounios & Beeman, 2009) and solving ill-structured problems (e.g., Ward, 2008; Ward, Dodds, Saunders, & Sifonis, 2000, Exp. 2). Though these analyses were rooted in cognitive theory (usually in fixation, restructuring, and conceptual organization), participants generated only a single response or solution. If think-aloud protocols were not taken during solutions (cf. Chrysikou & Weisberg, 2005; Gilhooly et al., 2007), again, the idea generation process remained somewhat hidden.

As an illustration of the power of collecting multiple responses, Smith, Huber, and Vul (2013) asked participants to generate multiple candidate answers to remote-associates problems (Mednick, 1962; see also Bowden & Jung-Beeman, 2003). These problems consist of three cues (e.g., "pine," "crab," "sauce"), and participants must find a fourth word that can be combined with each of the three to form an English phrase or compound word (in this case, "apple"). Smith and colleagues showed evidence of sequential dependency among the response arrays, such that each response iteration was highly semantically related to the previous iteration, but semantic similarity declined as the iterations became more distant in time. They also showed that the three cue words defined a constrained semantic space in which participants searched for responses. That is, they were able to show that prior knowledge constrained the search for responses, but also that there was a level of randomness in people's navigation through their search spaces. This level of analysis of the task would not have been possible without multiple response iterations.

Finally, in addition to that of Gilhooly and colleagues (2007), two other studies (Beaty & Silvia, 2012; Heinonen et al., 2016) more recently have examined serial-order effects in divergent thinking. Both of these studies leveraged the fact that divergent-thinking tasks require response iteration. Beaty and Silvia found that originality generally increased as a function of response order on a task based on imagining alternative uses for a brick. They also found that people scoring high on fluid intelligence showed less of a serial-order effect. That is, high fluid intelligence scores were associated with greater originality *earlier* in participants' response arrays. Beaty and Silvia further claimed that an executive control process operates on semantic memory activation to help people avoid high-frequency (i.e., low-originality) responses during divergent thinking. Heinonen and colleagues did not examine the response content, but showed that participants responded at slower rates as responding continued (see also Hass, 2016). However, since neither study examined semantic aspects of the responses themselves, it remains to be determined whether fluid intelligence is directly associated with semantic memory access.

Semantic memory and creativity

Other authors have described a fundamental relationship between creative thinking and semantic memory (e.g., Abraham, 2014; Beaty, Silvia, Nusbaum, Jauk, & Benedek, 2014; Kenett, Anaki, & Faust, 2014; Ward, 2008). It should also be noted that single-response paradigms can certainly test semantic-memory hypotheses (e.g., Chrysikou & Thompson-Schill, 2011; Prabhakaran et al., 2014; see also Green, 2016). In an article advocating such a methodology,

Prabhakaran and colleagues asked people to generate single verbs as creative responses to noun prompts. The authors analyzed the semantic relations between the prompts and the responses using latent semantic analysis (LSA; see, e.g., Landauer & Dumais, 1997). LSA is a distributional semantics technique in which the meaning of a word is captured in terms of its co-occurrences with other words in the documents comprising a large linguistic corpus. The result is a vector representation of a word in terms its "occurrence" in a reduced set of "latent" documents (factors). This reduced representation has been shown to account for the results of many studies of semantic memory, notably for word association (for a review, see Landauer & Dumais, 1997), but has also been criticized as not being an optimal model for learning and language development (Steyvers & Tenenbaum, 2005). Prabhakaran and colleagues found that participants' mean semantic distance scores, on a subset of trials in which they were explicitly told to be creative, significantly correlated with summary scores on traditional divergent-thinking tests, and also with judges' scores of creativity on a story-writing task. Additionally, the "creative" semantic distance scores correlated with measures of intelligence and with accuracy on a three-back working memory task.

Though Prabhakaran and colleagues (2014) did not track semantic distance across successive response iterations, the relationship between semantic distance and other creativity measures is encouraging. Indeed, a few authors have experimented with the use of LSA to quantify semantic distance within the context of divergent-thinking tasks (Dumas & Dunbar, 2014; Forster & Dunbar, 2009; Harbison & Haarmann, 2014). All of those studies have shown that the semantic distance of divergent responses relates to the judgments of creativity or originality given to those items by human raters. Taken together with the construct validity that Prabhakaran and colleagues presented regarding LSA-derived semantic distances, such a measure seems to have the potential to illuminate cognitive processing during divergent thinking.

The present study

As was discussed, summary scores of divergent-thinking performance seem to obscure the underlying idea generation process. A potential solution would be to analyze the semantic properties of successive responses to illuminate the idea generation process with regard to human semantic memory. Two important points of departure separate the present study's use of LSA for semantic analysis and the three studies discussed above. First, the previous studies did not examine temporal patterns of semantic distance. Given the results presented by Gilhooly and colleagues (2007), one would imagine that the semantic distance between the divergent-thinking prompt and

subsequent responses would not remain constant (see also Beaty & Silvia, 2012). Second, the present study was focused on how successive divergent-thinking responses related, semantically, to the prompt concept (e.g., a brick). This was more in line with the technique of Prabhakaran and colleagues (2014), who computed the semantic distance between the nouns used as prompts (e.g., "umbrella") and the responses (e.g., "poking"). In contrast, Forster and Dunbar (2009) compared participants' divergent-thinking responses to a set generated by an independent group of participants who were instructed to generate "uncreative" responses. This is important, because distance in semantic space is relative to the points of comparison. To make LSA-derived distances useful for creative-cognition studies, the target concepts from which the participants' responses "diverge" must be fixed and must represent the type of general knowledge that is tapped during divergent thinking. Thus, the prompts used in this study were represented in semantic space in terms of composite dictionary definitions (see Table 1).

Finally, the analysis was carried out using a preexisting dataset, collected and analyzed by Silvia and colleagues (2008) for the purposes of systematizing divergent-thinking originality scoring. The subjective rating system used to score originality was shown to be reliable, and provides a good metric against which semantic distance can be compared. That dataset also included measures of fluid intelligence—used currently to replicate Beaty and Silvia's (2012) results regarding the interaction of fluid intelligence and the serial-order effect—and participants' choices of their "top two" most creative responses on each divergent-thinking task. The latter offered an opportunity to explore the relationship between semantic distance and participants' perceptions of the creativity of their own responses.

To summarize, the analytic goals for this study were (1) to explore the relationship between subjective creativity scores and semantic distance scores by using LSA and a fixed representation of the category of the divergent-thinking prompt; (2) to test the generalizability of the serial-order effect in terms of semantic distance and probe its relation to fluid intelligence; and (3) to provide preliminary analysis of participants' own response evaluations in terms of semantic distance.

Method

Data preparation and variable definitions

Existing data The dataset was data collected as a part of the Creativity and Cognition Project at the University of North Carolina, Greensboro (see Silvia et al., 2008), and was obtained from that first author via his website (www.uncg.edu/~p_silvia) and subsequent personal communications. A complete description of the data collection procedure appears in the article by Silvia and colleagues (2008) and will not be repeated here, except for some relevant details that are summarized below.

Divergent-thinking data A total of 226 undergraduates completed a battery of pencil-and-paper tests in groups of one to ten under supervision of laboratory experimenters. Of interest to the present analyses are the responses on four divergent-thinking tasks—alternative uses for a *brick*, alternative uses for a *knife*, instances of *round things*, and instances of *things that make noise*. The instructions for the two alternative-uses tasks were as follows:

> For this task, you should write down all of the original and creative uses for a brick that you can think of. Certainly there are common, unoriginal ways to use a brick; for this task, write down all of the unusual, creative, and uncommon uses you can think of. You'll have three minutes (Silvia et al., 2008, p. 72).

The instructions for the instances tasks were identical, except that they called for participants to think, for example, of "unusual, creative, and uncommon" things that are round.

Three trained raters independently gave subjective creativity ratings (henceforth, simply *creativity ratings*) to each response. Participants also circled the *top two* responses in their arrays for each task. For the present analysis, the creativity ratings were averaged across raters to simplify the linear modeling. The top-two designation was retained as a categorical variable for the purpose of exploring the relationships among top-two choice, semantic distance, and response order.

Table 1 Phrases used to construct the LSA representations of the four divergent-thinking prompts

Divergent-Thinking Prompt	LSA Phrase Representation
Brick (alternative uses)	a small, hard block of baked clay that is used to build structures such as houses and sometimes to make streets and paths
Knife (alternative uses)	a usually sharp blade attached to a handle that is used for cutting or as a weapon
Things that make noise	a loud or unpleasant sound that someone or something makes or unwanted electronic signals that harm the quality of something
Things that are round	shaped like a circle ball, or a cylinder, having curves rather than angles

Fluid intelligence, verbal fluency, and strategy fluency In addition, the dataset included the measures of fluid intelligence, verbal fluency, and strategy formation analyzed by Nusbaum and Silvia (2011). Again, only the relevant details of that study are reproduced here. Fluid intelligence was measured with three tasks, which evaluate the ability to adapt to rule changes: Ravens Advanced Progressive Matrices (18 items, 12 min total), a letter sets task (identifying which set of letters violates a rule held by the others; 16 items, 4 min), and a paper-folding task (deciding how a piece of paper would look if folded, punched with a hole, and then unfolded; 10 items, 4 min). Each participant's performance was scored in terms of numbers of correct items on each measure. For the purpose of analysis, these scores were transformed into *z* scores and averaged to create a single *fluid intelligence* variable for each participant.

Four tasks measured verbal fluency: naming words that begin with the letter "f," words that begin with the letter "m," animals, and occupations. Participants generated verbal responses for 2 min per item. These fluency scores were averaged for each participant and then, due to significant skewing, log-transformed for the present analysis, to form a single *verbal fluency* variable for each participant.

Three other tasks measured strategies for producing verbal fluency responses: strategies for thinking of body parts, foods, and countries. As in the verbal-fluency tasks, participants were given 2 min per category. These scores were not skewed, and thus were averaged and transformed into *z* scores to create a single *strategy* variable for each participant.

Semantic distance Latent semantic analysis was performed using the tools hosted at lsa.colorado.edu (for more information, see Landauer & Dumais, 1997, and the documentation on the website). Analysis was performed using the data from the TASA corpus, which was compiled to represent the semantic information a human would learn between first grade and entry to college. The "one-to-many" LSA tool was used to compare each divergent-thinking response from the dataset to a target phrase—a composite description of the divergent-thinking prompt compiled from the Merriam Webster Dictionary (www.merriam-webster.com/dictionary; see Table 1) in document space. In comparing documents (i.e., phrases), LSA first computes the centroid of the word vectors in each document—here, a kind of semantic blending or average—and then computes the cosines between the vectors, which can range from −1 to 1. This represents the *similarity* of two vectors, such that the cosine of the angle between two identical vectors is 1, the cosine of two orthogonal (i.e.,

unrelated) vectors is 0, and the cosine of two vectors pointing in opposite directions is −1. The cosine similarity values were then transformed into *distances* by subtracting each from 1, which follows procedures used by other researchers (e.g., Dumas & Dunbar, 2014; Prabhakaran et al., 2014).

Prior to the analysis, all responses were manually spell-checked, and a set of stop-words[2] was removed using functions from the "tm" package (Feinerer, Hornik, & Meyer, 2008) in the R statistical programing environment (R Development Core Team, 2016).

One issue not addressed by previous authors was whether or not the 300-factor LSA model is best for this type of data, though it works well for representing the meanings of single words (Landauer & Dumais, 1997). Thus, before testing the hypotheses, the cosine similarities were computed and converted to distances for LSA solutions using 50, 100, 150, 200, 250, and 300 factors, all from the TASA corpus.

Figure 1 plots the strengths of the correlations between the distance and originality scores across the different factor solutions. The highest correlations between LSA distances and rated originality were found on the two alternative-uses tasks. Interestingly, the 100-factor LSA distances correlated highest with originality ratings from the brick task, whereas the 100-factor solution did not perform well on the other tasks. The 300-factor solution seemed to be a good compromise for the alternative-uses tasks, because this solution yielded the highest correlation with LSA distances on the knife task.

The 300-factor distances did not correlate very highly with creativity ratings on the two instances tasks. Indeed, these tasks required participants to list instances of objects, and it has been suggested that participants approach these tasks as category/semantic-fluency tasks rather than as divergent-thinking tasks per se (Hass, 2015). Coupled with the large differences in the strengths of the correlations across the two instances tasks for the 300-factor solution, and the possibility that these tasks might not be interpreted as requiring creative thinking, they were dropped from further analysis.[3] Figure 2 shows the frequency distributions of the 300-factor LSA semantic distance scores for the two alternative-uses tasks. Table 2 lists descriptive statistics for both the creativity ratings, and the semantic distance scores for each prompt.

Results

All statistical analyses were performed using R (R Development Core Team, 2016). Because of the skewness found in both the distance and originality scores, the Bayesian multilevel modeling procedures given by Finch, Bolin, and Kelley (2014), and implemented using the

---

[2] The LSA analyzer ignores stop-words, so the dictionary definitions were not prepared in this way. The reason to remove stop-words from the responses was so that a smaller number of unique responses from the participants' data could be fed into the LSA analyzer.

[3] The exclusion of these tasks did not change the results of either of the hypothesis tests included in the next section.
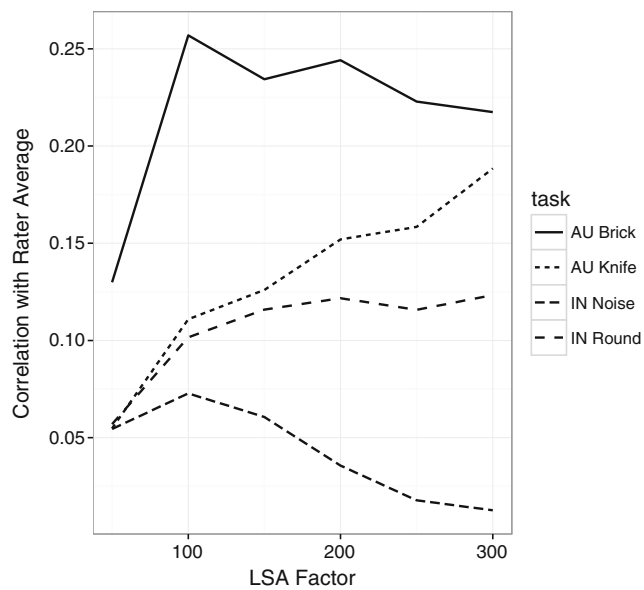
Fig. 1 Summary correlations of semantic distances with subjective creativity ratings across different latent semantic analysis (LSA) factor solutions
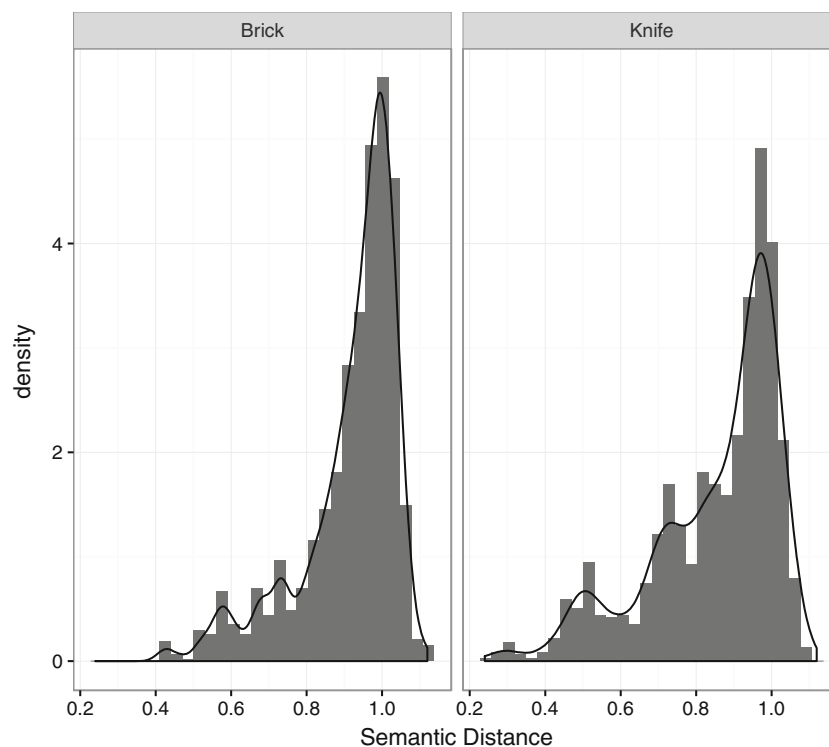
with a burn-in of 3,000 and a thinning of 10—were used, and all models were checked for convergence and autocorrelation of the estimates. In all cases the models converged, with low autocorrelation of the estimates across iterations.

Examining the distance–creativity relationship

The first model examined the relationship between semantic distance and creativity ratings with a random intercept for each participant. This provided a better estimate of the true correlation between originality and semantic distance than those given in Fig. 1. The posterior mean for the distance coefficient was .51 (95 % Bayesian credible interval: .39, .64), showing that semantic distance significantly predicted creativity ratings in the positive direction. There was significant variance in the participant-level intercepts, such that the posterior mean for the variance estimates was .047 (95 % Bayesian credible interval: .036, .061). The latter is evidence of significant individual differences in the average creativity ratings per participant, and explains the difference between the correlations in Fig. 1 and the regression coefficients from the multilevel model.

The second model examined the effect of task as a level-2 variable and the cross-level interaction between task and distance, again with a random effect of participant and the creativity rating as the criterion variable. As in the first model, the Creativity Rating ~ Semantic Distance slope was significant. No cross-level interaction was apparent between task and distance, meaning that there was no significant variation in the

"MCMCglmm" package (Hadfield, 2010) in R, were used to examine the effects of interest: the relationship between semantic distance and originality ratings, and the serial-order effect. The MCMCglmm algorithm uses noninformative normal priors for estimating coefficients and intercepts, and Wishart priors were used for the variance/covariance components. The default MCMCglmm settings—13,000 iterations

Fig. 2 Frequency distributions and densities of semantic distances on all four divergent-thinking tasks

Table 2 Descriptive statistics for semantic distances (latent semantic analysis 300-factor solution) and creativity ratings across the two alternative uses (AU) tasks, at the level of each individual response

Task	$N_{responses}$	M	SD	Median	Skew	Kurtosis
AU – Brick						
Semantic distance (300 factors)	1,738	0.92	0.13	0.96	−1.47	1.81
Creativity rater average	–	1.68	0.56	1.67	1.16	2.45
AU – Knife						
Semantic distance (300 factors)	1,682	0.85	0.17	0.91	−1.11	0.57
Creativity rater average	–	1.71	0.53	1.67	1.14	3.08

Table 3 Bayesian estimates for the relationship between distance and originality as moderated by task

Model Term	Posterior Mean	Lower CI	Upper CI	pMCMC
Model for the Intercept				
Grand mean	1.04	0.84	1.23	–
Knife task	0.25	0.03	0.49	.018[*]
Model for the Originality - Distance Slopes				
Mean Originality - Distance slope	0.70	0.48	0.89	<.001
Knife task	−0.20	−0.44	0.06	.130
Variance Estimates				
Intercept variance per participant	0.047	0.034	0.059	–
Level 1 variance	0.239	0.227	0.251	

Limits are for the 95 % Bayesian credible intervals (CIs)

linear relationship between creativity ratings and semantic distances across the two tasks (see Fig. 3). It is important to note that Fig. 3 is aggregated across participants with variations in mean creativity rating, so the noise present in the figure is accounted for by the model (see Table 3).

Serial-order effect

A growth model for semantic distance was assembled to probe linear and nonlinear serial-order effects. Orthogonal linear and quadratic terms for response order were constructed using the "poly" function in R. For the first model, semantic distances were regressed on the linear and quadratic response order

terms, with a random intercept for each participant. There was a significant linear serial-order effect (posterior mean coefficient = .62, $p < .001$; 95 % credible interval: .30, .94), but no significant quadratic serial-order effect (posterior mean coefficient = −.18, $p = .25$; 95 % credible interval: −.47, .15).

Fluid intelligence, verbal fluency, and strategies Since only the linear serial-order effect was significant, the remaining tests of the effects of fluid intelligence, verbal fluency, strategy, and prompt utilized only the linear serial-order term. In
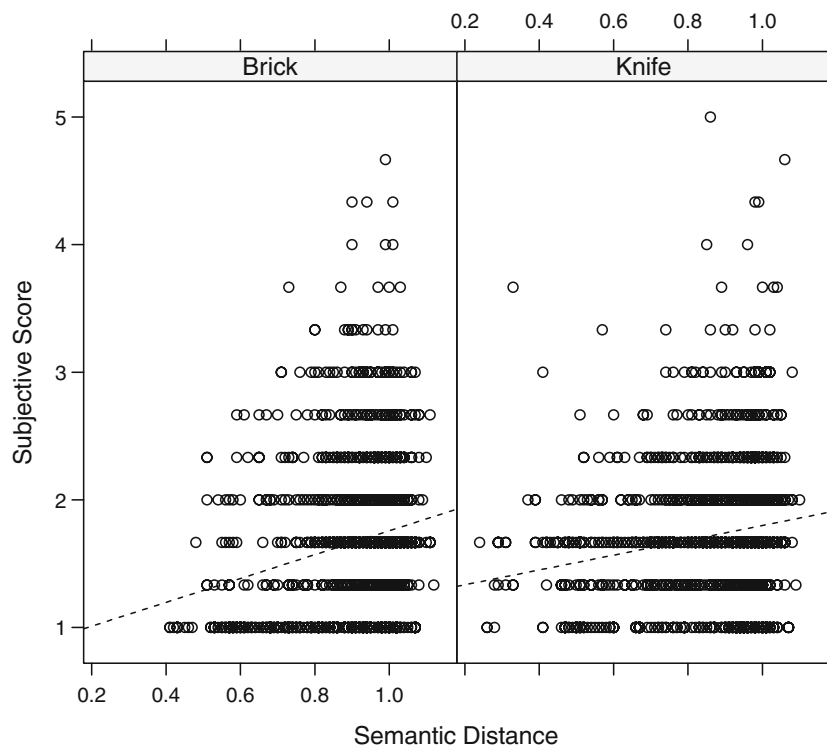
Fig. 3 Scatterplots of semantic distances and subjective scores. Plots are aggregated across participants and do not show variation in the participant-level intercepts

addition, the linear term was rescaled by setting the first response to 0. This allowed for a straightforward interpretation of the intercept estimates as estimates of the *initial* distance, or the semantic distance of the first response.

Table 4 lists the results of two models that built on the basic linear growth model for distance across response orders. The top of the table gives the estimates for a serial-order model with fluid intelligence, verbal fluency, and strategy scores predicting initial semantic distance, and also cross-level interactions between the three intelligence measures and response order. The results were surprising, in that the serial-order effect vanished (posterior mean = .008), whereas fluid intelligence significantly predicted the initial semantic distances in the positive direction.

The bottom of the table shows that when the effect of task was added to the model, the serial-order effect remained nonsignificant, but the effect of fluid intelligence remained significant. Additionally, the table shows that participants began the knife task by giving less distant responses from the prompt than in the brick task. The deviance information criterion (DIC) is given for each model, with the second model producing a lower (more negative) result, which favors that model.

Top two Finally, it is of interest to explore whether or not people are sensitive to variations in the semantic distances of their own responses when they make choices of their top two responses. This analysis is purely descriptive and exploratory, since there is much more to be understood about how people

Table 4 Bayesian estimates for two models of semantic distance

Model Term	Posterior Mean	Lower CI	Upper CI	*p*
Gf-only Model (DIC = −2,890.26)				
Model for Initial Semantic Distance				
Grand mean	0.856	0.676	1.034	–
Fluid intelligence	0.018	0.004	0.031	.008*
Verbal fluency	0.005	−0.053	0.065	.852
Strategy fluency	0.012	−0.004	0.026	.118
Model for Slopes				
Mean slope	0.008	−0.023	0.039	.564
Fluid intelligence	-8.86×10^{-5}	−0.002	0.003	.986
Verbal fluency	−0.002	−0.012	0.008	.724
Strategy fluency	-5.37×10^{-4}	−0.003	0.002	.704
Variance Components				
Intercept variance per participant	0.002	0.001	0.003	
Level-1 variance	0.022	0.021	0.023	
Gf+ Task Model (DIC = −3,028.77)				
Model for Initial Semantic Distance				
Grand mean	0.897	0.714	1.049	–
Fluid intelligence	0.017	0.005	0.030	.016*
Verbal fluency	0.003	−0.052	0.059	.940
Strategy fluency	0.012	−0.002	0.026	.106
Knife task	−0.075	−0.090	−0.060	<.001*
Model for Slopes				
Mean slope	0.004	−0.021	0.032	.782
Fluid intelligence	-5.51×10^{-6}	−0.003	0.002	.988
Verbal fluency	-7.39×10^{-4}	−0.010	0.007	.874
Strategy fluency	-3.59×10^{-4}	−0.003	0.002	.756
Knife task	0.003	−0.001	0.006	.060
Variance Components				
Intercept variance per participant	0.0017	0.0011	0.0024	–
Level-1 variance	0.021	0.019	0.022	–

The model at the top of the table is nested within the model at the bottom of the table. Rescaling of the predictors yielded estimates of the semantic distance of the first response, and for the rate of change as iterations progress. CI, 95 % Bayesian credible interval; DIC, deviance information criterion (lower [= more negative] is better)

monitor and evaluate the results of their own response generation processes. Figure 4 shows the cross-tabulations of frequencies of the top-two and non-top-two responses by response order. Chi-square tests of independence showed that participants' choices of their top two did not depend on response order on either the brick task [$\chi^2(22) = 20.08$, $p = .58$] or the knife task [$\chi^2(18) = 28.78$, $p = .05$]. Regarding semantic distance, Fig. 5 shows that the distribution of semantic distances for top-two scores was more compact for top-two responses, but that the medians of both distributions were relatively equal.

Discussion

The goal of this analysis was to describe a theoretical and methodological framework for understanding the response generation processes that operate during divergent-thinking task performance. Some of the results were consistent with

prior research, but others were not. First, LSA-derived semantic distances correlated with creativity ratings on only two of the four divergent-thinking tasks examined in the dataset. This is consistent with prior research, but the correlations were lower than in previously reported results (e.g., Harbison & Haarmann, 2014), in which LSA-derived distances correlated with subjective creativity ratings in the range of $r = .40$. However, testing the relationship via multilevel modeling revealed that semantic distance was a significant predictor of subjective creativity ratings when the nesting of responses within participants was part of the model. That is, there was a significant positive relationship between semantic distance and the creativity rating *at the level of each response*.

The positive relationship between fluid intelligence and the initial semantic distance of responses is consistent with the results presented by Beaty and Silvia (2012). However, in the present analysis the relationship between fluid intelligence and initial semantic distance rendered the serial-order effect nonsignificant. This was likely due to the strong relationship between fluid intelligence and semantic distance, and also to the fact that Beaty and Silvia used a 10-min time limit, which may have intensified the serial-order effect. On the other hand, it could be that the change in metric—from creativity ratings to semantic distance—was responsible for the lack of a strong serial-order effect. That is, it might be that semantic distance serves to anchor the beginning of the iteration process a certain distance away from the concept represented by the prompt, as a sort of heuristic to ensure the novelty of responses. After that adjustment, successive iterations may not yield linear increases in semantic distance. The exploratory analysis of the nature of participants' top two responses suggested that people were just as likely to choose early as to choose later responses as their most creative ones. That is, originality and semantic distance are not equivalent, nor should they be. Originality is a subjective construct, likely
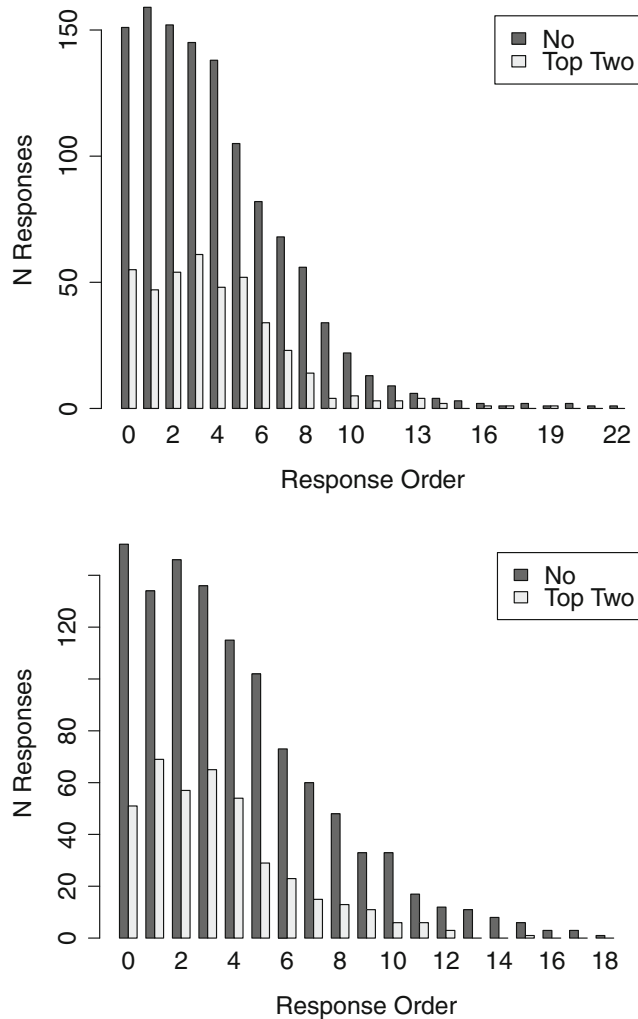
Fig. 4 Frequency of top-two and non-top-two designations by response order for the brick task (top) and the knife task (bottom)
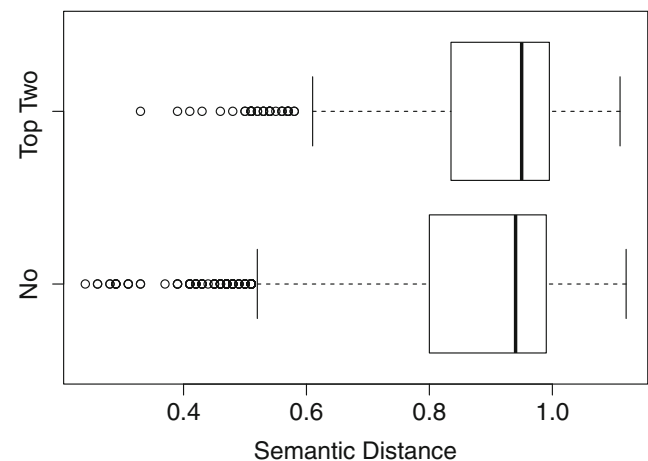
Fig. 5 Boxplot of semantic distances for top-two and non-top-two response designations

influenced by cognitive and cultural forces (Csikszentmihalyi, 1999). However, the goal of the analysis was not to equate semantic distance and originality, but to explore whether semantic distance changed as a function of response order.

Implications of the semantic distance–fluid intelligence relationship

Like Beaty and Silvia's (2012) results, the present analysis has revealed more links between semantic memory and fluid intelligence. But what role does fluid intelligence play in the idea generation process? As was described above, in a large body of research fluid intelligence and other, related cognitive abilities (e.g., switching, updating, and inhibition) have been found to predict divergent-thinking summary scores (see Benedek et al., 2014, for a review). These relationships are central to recent neuroscientific explanations of creative cognition. As was stated earlier, an emerging theory of creativity and the brain is that the structures that comprise the default mode network support the type of self-directed thought required during divergent thinking, whereas executive regions of the brain are involved in the direction and evaluation of the procedure as a whole. If it is assumed that the executive control–default mode network theory is accurate, that means that one role of the executive system might be to adjust the semantic content of potential responses, *at the start* of the divergent-thinking process. This is consistent with data linking the generation of highly unusual divergent-thinking responses to activity in the inferior frontal gyrus and temporal poles (Abraham et al., 2012; Chrysikou & Thompson-Schill, 2011). Thus, the results of the present analysis suggest that if fluid intelligence measures the functioning of executive control centers in the brain, then the better those areas function, the more likely one is to begin the iteration process with distant ideas. Indeed, Kenett, Beaty, Silvia, Anaki, and Faust (2016) showed that levels of fluid intelligence and levels of creative achievement—assessed via the Creative Achievement Questionnaire (Carson, Peterson, & Higgins, 2005)—were differentially related to semantic memory organization, as assessed via semantic network analysis. Particularly, participants who demonstrated high fluid intelligence and high levels of creative achievement tended to have more flexible and less structured semantic networks.

Limitations and remaining questions

What remains to be explained is how the remainder of the iteration process unfolds during divergent-thinking tasks. According to the present results, participants continue to iterate responses that are distant from the prompt, but not necessary qualitatively *better* or quantitatively *more distant* than those earlier in the iteration process. Again, the latter finding is supported by the exploratory analysis of the serial

distribution of top-two responses. What semantic distances seem to be able to capture is the difference between very close associates of the prompts and those that are unrelated. The final sections of this article will cover a few key questions that emerged from the potential limitations of this analysis.

Regarding LSA The validity of the present results rests wholly on the validity of the LSA-derived distances for this kind of data. Given that many of the results—particularly the relationships between semantic distance and both originality and fluid intelligence—are consistent with those obtained in prior research, the semantic distance metric presented here has convergent validity. One of the strengths of using LSA, as compared to semantic network analysis (see, e.g., Kenett et al., 2014), is that it can be applied to phrases, sentences, and even paragraphs. Of course, there are other ways to assess the semantic relations among phrase-level data. For example, Chan and Schunn (2015) used topic models (e.g., Griffiths, Steyvers, & Tenenbaum, 2007) to examine combination distances among contributions to the online platform OpenIDEO. This may be a fruitful path for future analysis. Indeed, Fig. 1 shows that the creativity ratings related differently to semantic distance measures across the tasks. However, this is only problematic if LSA-derived distances are to *replace* subjective scores. This was not the intent of this analysis. Rather, the intent here was to measure the semantic relations between divergent-thinking responses and task prompts. The reason to opt for LSA based on the TASA corpus is that alternative-uses tasks are domain-general. Thus, the TASA corpus seemed to be the most fitting way to represent the general knowledge that people might have accessed to generate responses (for more on this corpus, see the documentation at lsa.colorado.edu).

Association norms or network analysis? As has been discussed previously, people tend to generate more unique responses to divergent-thinking items if they have more distributed semantic networks (Kenett et al., 2014). As was mentioned, one drawback to network analysis is that it is difficult to apply to phrase-level data, which are generated during divergent thinking. Instead, Kenett and colleagues (2014; Kenett et al., 2016) assessed semantic memory organization by analyzing participants' responses on a category fluency task. However, there is no reason why future studies should not combine the two approaches. That is, what seems most necessary at this point is to examine whether individual differences in semantic memory organization predict semantic distances on divergent-thinking tasks. It is likely that there will be some relationship, and one recommendation from this study is that others begin to examine such data.

Secondary data analysis Finally, the dataset used for the analysis was not generated for the purposes of cognitive

analysis. Specifically, reaction time data are missing, such that it was not clear whether individual differences in response rates and interresponse times could be related to semantic distance or fluid intelligence. Hass (2016) presented preliminary evidence that interresponse times increase across iterations, so further research will be necessary to fully characterize the iteration process in divergent thinking. This dataset was selected because the system of creativity ratings was shown to be reliable (Silvia et al., 2008), and also because of inclusion of the fluid intelligence data. However, because the divergent-thinking tasks were administered amidst a battery of other tests, order effects and fatigue may have been confounding variables for this particular analysis. Still, interpretable effects were found, and this concern, although worthy of consideration, does not seem to be serious.

Concluding remarks The goal of this article was to present a first step toward understanding response generation during divergent thinking, using a metric that is rooted in cognitive theory. The results supported the utility of semantic distance as one approach to modeling response generation. The results also showed support for an executive or controlled process operating on response generation, though much more research will be necessary to properly test that theory. Regardless, researchers are urged to begin to think about divergent-thinking response generation as an interesting focus for research on creative thinking.

References

Abraham, A. (2014). Creative thinking as orchestrated by semantic processing vs. cognitive control brain networks. *Frontiers in Human Neuroscience, 8*(95), 1–6. doi:10.3389/fnhum.2014.00095

Abraham, A., Pieritz, K., Thybush, K., Rutter, B., Kröger, S., Schweckendiek, J., … Hermann, C. (2012). Creativity and the brain: Uncovering the neural signature of conceptual expansion. *Neuropsychologia*, *50*, 1906–1917. doi:10.1016/j.neuropsychologia.2012.04.015

Beaty, R. E., Benedek, M., Silvia, P. J., & Schacter, D. L. (2016). Creative cognition and brain network dynamics. *Trends in Cognitive Sciences, 20*, 87–95. doi:10.1016/j.tics.2015.10.004

Beaty, R. E., & Silvia, P. J. (2012). Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks. *Psychology of Aesthetics, Creativity, and the Arts, 6*, 309–319. doi:10.1037/a0029171

Beaty, R. E., Silvia, P. J., Nusbaum, E. C., Jauk, E., & Benedek, M. (2014). The roles of associative and executive processes in creative cognition. *Memory & Cognition, 42*, 1186–1197. doi:10.3758/s13421-014-0428-8

Benedek, M., Jauk, E., Fink, A., Koschutnig, K., Reishofer, G., Ebner, F., & Neubauer, A. C. (2014). To create or to recall? Neural mechanisms underlying the generation of creative new ideas. *NeuroImage, 88*, 125–133. doi:10.1016/j.neuroimage.2013.11.021

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, & Computers, 35*, 634–639. doi:10.3758/BF03195543

Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the Creative Achievement Questionnaire. *Creativity Research Journal, 17*, 37–50. doi:10.1207/s15326934crj1701_4

Chan, J., & Schunn, C. D. (2015). The importance of iteration in creative conceptual combination. *Cognition, 145*, 104–115. doi:10.1016/j.cognition.2015.08.008

Chrysikou, E. G., & Thompson-Schill, S. L. (2011). Dissociable brain states linked to common and creative object use. *Human Brain Mapping, 32*, 665–675. doi:10.1002/hbm.21056

Chrysikou, E. G., & Weisberg, R. W. (2005). Following the wrong footsteps: Fixation effects of pictorial examples in a design problem-solving task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1134–1148. doi:10.1037/0278-7393.31.5.1134

Csikszentmihalyi, M. (1999). Implications of a systems perspective for the study of creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 313–335). New York: Cambridge University Press.

Development Core Team, R. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from http://R-project.org

Dietrich, A., & Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological Bulletin, 136*, 822–848. doi:10.1037/a0019749

Dumas, D., & Dunbar, K. N. (2014). Understanding fluency and originality: A latent variable perspective. *Thinking Skills and Creativity, 14*, 56–67. doi:10.1016/j.tsc.2014.09.003

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software, 25*(5), 1–54. Retrieved from www.jstatsoft.org/v25/i05/

Finch, W. H., Bolin, J. E., & Kelley, K. (2014). *Multilevel modeling using R*. New York: CRC Press.

Forster, E. A., & Dunbar, K. N. (2009). Creativity evaluation through latent semantic analysis. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 602–607). Austin: Cognitive Science Society.

Gilhooly, K. J., Fioratou, E., Anthony, S. H., & Wynn, V. (2007). Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology, 98*, 611–625. doi:10.1348/096317907X173421

Green, A. E. (2016). Creativity, within reason: Semantic distance and dynamic state creativity in relational thinking and reasoning. *Current Directions in Psychological Science, 25*, 28–35. doi:10.1177/0963721415618485

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114*, 211–244. doi:10.1037/0033-295X.114.2.211

Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

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*, 288–294. doi:10.1177/0956797611429710

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R Package. *Journal of Statistical Software, 33*(2), 1–22. Retrieved from www.jstatsoft.org/v33/i02/

Harbison, J. I., & Haarmann, H. (2014). Automated scoring of originality using semantic representations. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of COGSCI 2014* (pp. 2337–2332). Retrieved from https://mindmodeling.org/cogsci2014/papers/405/index.html

Hass, R. W. (2015). Feasibility of online divergent thinking assessment. *Computers in Human Behavior, 46,* 85–93. doi:10.1016/j.chb.2014.12.056

Hass, R. W. (2016). Conceptual expansion during divergent thinking. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 996–1001). Austin: Cognitive Science Society.

Hass, R., Toub, T. S., Yust, P., Hirsh-Pasek, K., & Golinkoff, R. M. (2015). *What is creativity in young children?* Philadelphia: Poster presented at Society for Research in Child Development.

Heinonen, J., Numminen, J., Hlushchuk, Y., Antell, H., Taatila, V., & Suomala, J. (2016). Default mode and executive networks areas: Association with the serial order in divergent thinking. *PLoS ONE, 11,* e0162234. doi:10.1371/journal.pone.0162234

Jansson, D. J., & Smith, S. M. (1991). Design fixation. *Design Studies, 12,* 3–11.

Jung, R. E., Mead, B. S., Carrasco, J., & Flores, R. A. (2013). The structure of creative cognition in the human brain. *Frontiers in Human Neuroscience, 7,* 330. doi:10.3389/fnhum.2013.00330

Kenett, Y. N., Anaki, D., & Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Frontiers in Human Neuroscience, 8*(407), 1–16. doi:10.3389/fnhum.2014.00407/abstract

Kenett, Y. N., Beaty, R. E., Silvia, P. J., Anaki, D., & Faust, M. (2016). n, fluid intelligence, and creative achievement. *Psychology of Aesthetics, Creativity, and the Arts*. doi:10.1037/aca0000056

Kounios, J., & Beeman, M. (2009). The aha! moment: The cognitive neuroscience of insight. *Current Directions in Psychological Science, 18,* 210–216. doi:10.1111/j.1467-8721.2009.01638.x

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

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review, 69,* 220–232. doi:10.1037/h0048850

Nusbaum, E. C., & Silvia, P. J. (2011). Are intelligence and creativity really so different? *Intelligence, 39,* 36–45. doi:10.1016/j.intell.2010.11.002

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,* 423–432. doi:10.1037/a0036549

Prabhakaran, R., Green, A. E., & Gray, J. R. (2014). Thin slices of creativity: Using single-word utterances to assess creative cognition. *Behavior Research Methods, 46,* 641–659. doi:10.3758/s13428-013-0401-7

Silvia, P. J. (2008). Another look at creativity and intelligence: Exploring higher-order models and probable confounds. *Personality and Individual Differences, 44,* 1012–1021. doi:10.1016/j.paid.2007.10.027

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., … Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts, 2,* 68–85. doi:10.1037/1931-3896.2.2.68

Smith, K. A., Huber, D. E., & Vul, E. (2013). Multiply-constrained semantic search in the Remote Associates Test. *Cognition, 128,* 64–75. doi:10.1016/j.cognition.2013.03.001

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29,* 41–78. doi:10.1207/s15516709cog2901_3

Torrance, E. P. (1979). *The search for satori and creativity.* Buffalo: Creative Education Foundation Press.

Ward, T. B. (2008). The role of domain knowledge in creative generation. *Learning and Individual Differences, 18,* 363–366. doi:10.1016/j.lindif.2007.07.002

Ward, T. B., Dodds, R. A., Saunders, K. N., & Sifonis, C. M. (2000). Attribute centrality and imaginative thought. *Memory & Cognition, 28,* 1387–1397. doi:10.3758/BF03211839

Ward, T. B., Smith, S. M., & Vaid, J. (Eds.). (1997). *Creative thought: An investigation of conceptual structures and processes.* Washington, DC: American Psychological Association.

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

Weisberg, R. W., & Hass, R. (2007). We are all partly right: Comment on Simonton. *Creativity Research Journal, 19,* 345–360.