

***THE CASE FOR CORE IDEAS IN THE FUZZY FRONT END:
VERBOSITY INCREASES PERCEIVED CREATIVITY***

Laura J. Kornish
Leeds School of Business
University of Colorado Boulder
kornish@colorado.edu

Sharaya M. Jones
Leeds School of Business
University of Colorado Boulder
sharaya.jones@colorado.edu

February 2019

ABSTRACT

At the “fuzzy front end” of an innovation process, organizations typically consider dozens, or even hundreds, of raw ideas. Selecting the best ones to focus on is a double challenge: evaluating so many ideas is large undertaking, and the ideas in their raw form permit only noisy evaluations. Recent work presents a scalable data-driven approach for identifying patterns of the words used in good ideas. The main finding from that work is that the most creative ideas contain a typical balance of close and far associations between the elements of the idea. In this paper, we explore a simpler explanation: longer idea descriptions, which tend to have a more typical balance, are perceived as more creative. In light of that simpler explanation, we reconsider the question about how close and far associations predict evaluations of ideas. To do that, we consider core ideas, ideas that are stripped down to their essential elements. In contrast to the recent work showing balance is best, and earlier work touting “moderate incongruity,” we find that close associations between the essential elements of the idea are the best predictors of idea appeal.

Keywords: innovation, idea generation, creativity, new product development, combinations, close and far associations

We gratefully acknowledge the helpful input from John Lynch, Page Moreau, Mike Palazzolo, and Jameson Watts.

At the “fuzzy front end” of an innovation process, organizations typically consider dozens, or even hundreds, of raw ideas. Sorting through a large set of ideas to identify the most promising ones is a significant challenge. Academics have tackled this challenge from various angles. There are studies that examine the success of an idea based on its origin (Girotra, Terwiesch, and Ulrich 2010, Goldenberg, Lehmann, and Mazursky 2001, Poetz and Schreier 2012, Rosenzweig, Mazurksy, and Tellis 2015), its semantic attributes (Eliashberg, Hui, and Zhang 2007, Toubia and Netzer 2017), its more general attributes (Åstebro and Elhedhli 2006), and its fundamental structure (Goldenberg, Mazursky, and Solomon 1999).

Like the authors of those papers, we are also concerned with insights for identifying promising ideas. We build on the recent work by Toubia and Netzer (2017), which analyzes the semantic network properties of raw ideas. That work examines the strength of association between elements of an idea. They find that the most creative ideas have a balance of close and far associations.

Their work is especially interesting because it provides a scalable, data-driven method to analyze *all the combinations* within an idea. Handling all the combinations is a departure from an existing paradigm in marketing scholarship, where authors have found support for *moderate incongruity* of a *single, central combination* for a product (e.g., Meyers-Levy and Tybout 1989). Toubia and Netzer’s approach therefore relaxes an embedded conceptual constraint in the congruity work that ideas have two focal aspects, and makes it more widely relevant to real sets of raw ideas.

However, we show how a forte of Toubia and Netzer’s work—that the methods can be applied to any set of ideas expressed in words—is also a vulnerability. The vulnerability arises because in looking at all combinations of words in an idea, we obscure patterns between the

combinations in the *core idea*, the shortened version of the idea that retains only the essential elements. Even stripping out stop words (the, a, of) and generic words (is, does, has), ideas differ in the nature and number of peripheral words. Some ideas include descriptions of the origin story, some list of possible uses of the idea or optional features, and some simply elaborate more than others. Toubia and Netzer's account is that ideas with a more prototypical balance of close and far associations are rated as most creative. Our simpler account is that longer idea descriptions tend to be rated as more creative (and also have a more prototypical balance).

Our evidence shows that when we strip ideas down to their essential elements, *close* associations predict purchase intent. This result contrasts with Toubia and Netzer's findings, that a prototypical balance of close and far associations predicts creativity but not purchase intent. It also contrasts with the conclusions supporting *moderate incongruity* in the work focused on two-part concepts.

Influential scholarship in ideation uses raw ideas as composed by the idea generators or designers. This element of realism in the studies contributes to the credibility that the study results will apply in practical settings. But in this paper, we show that the practice of using the raw ideas introduces confounds among the level of detail, the core idea itself, and the creative talents of the idea contributor. And that this confound is especially troubling when measuring a dependent variable of creativity.

This paper flows as follows. First, we introduce two data sets and show that the method of Toubia and Netzer (2017) easily replicates with them. Second, we explain our concerns about the confounding role of the idea description length and show mediation analyses. Third, we present evidence that the length of an idea description affects perceived creativity. Fourth, we show that the main result of Toubia and Netzer attenuates or disappears when working with core

instead of raw ideas. Finally, we use multiple approaches to show that close connections between the elements of core ideas predict appeal.

REPLICATIONS OF TOUBIA AND NETZER (2017)

The work of Toubia and Netzer (2017) contains a powerful and general technique for predicting perceived creativity of ideas and suggesting ways to increase it. They use a semantic network approach to analyze raw ideas, and their main results show how word patterns in the idea descriptions predict creativity.

We provide a brief explanation of their approach and main result and show replications. Their study focuses on combinations in each idea, specifically on the strength of association between all pairs of words within an idea. They predict that the *best ideas* will have a *balance of close and far associations*. They treat ideas as semantic networks: the words are nodes in the network, and the strength of association between any two words is the edge weight in the network. They look at the distribution over those edge weights. Their main finding is that *ideas that have semantic networks with edge weight distributions closest to a “prototypical” distribution are perceived as most creative*.

The prototypical distribution derives from a corpus of text specifically related to the innovation topic. That corpus can be a set of ideas or the text of results pages from a Google search on the topic. The strength of association between words in the corpus can be determined many ways; Toubia and Netzer use the Jaccard index in their primary analysis. (The Jaccard index for a pair of words is the ratio of the number of documents in the corpus in which the

words appear together to the number of documents in which at least one of the words appears.) The prototypical distribution gives a precise meaning for what constitutes “balance” between close and far associations.

We replicate their analysis on two sets of raw ideas for which we have the full text of the ideas plus creativity and purchase intent ratings.

Study 1a, Frosting Ideas. The first data set contains ideas for new frosting products (frosting in the sense of icing typically used to cover cakes and cookies). A panel of “creative consumers” who contract with a new products consulting firm generated the ideas in response to this prompt:

For this assignment, you are a brand manager for a consumer products company that has a successful line of frosting products. The frostings come in different flavors and are primarily used for covering cakes, cupcakes, and cookies. The products can be found in grocery store chains, usually in the baking aisle. Your task is to come up with new frosting products, that don't currently exist, that you would sell in the same grocery stores.

There were 248 ideas in total, generated by 61 people. We randomly selected a set of 96 for evaluation. We collected creativity and purchase intent ratings from Mechanical Turk workers.

Study 1b, Household Products Ideas. The second data set is based on one of the idea sets from Kornish and Ulrich (2014), household products from Quirky.com, a community product development website. We obtained the data set, including the text of the ideas. The data set has 100 ideas randomly selected from the entire universe of thousands of ideas submitted to Quirky.

The purchase intent responses were from a general population panel purchased from the vendor Qualtrics as part of the original data set. We independently collected creativity ratings from Mechanical Turk workers.

Semantic Network Examples. The key relationship in the analysis is between a *idea evaluations*—such as creativity or purchase intent—and the *distance of the edge weight distribution for an idea from the edge weight distribution for the “prototypical” distribution*. We illustrate *edge weight distribution* and *prototypical distribution* in Figure 1.

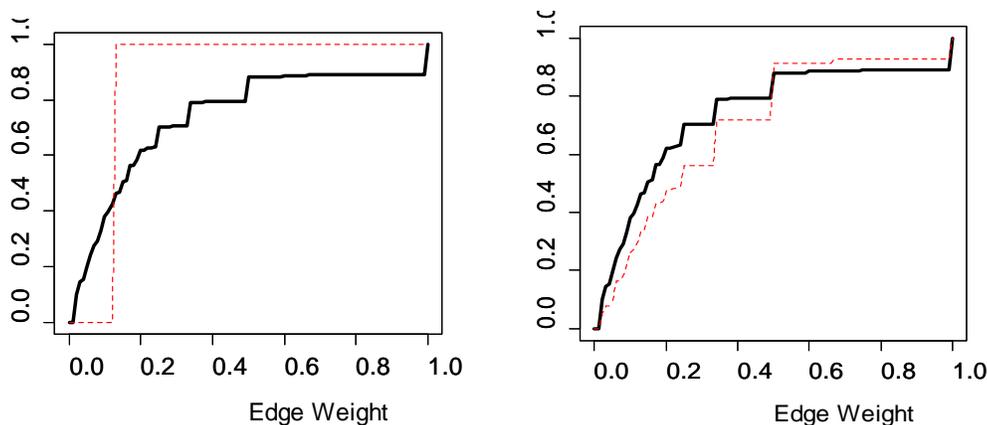
Figure 1 shows edge weight distributions for two ideas from the frosting data set. Both graphs in the figure show the prototypical cumulative distribution function of edge weights as a solid black line: the vertical value is the fraction of edge weights that are less than or equal to each horizontal axis value. The edge weights are between 0 and 1, with 1 meaning perfect association (the word on one side of the edge never appears without the word on the other side of the edge). Each graph also has a dashed line, representing the edge weight distribution for a single idea. The idea on the left is “Cookie dough” and the idea on the right is “Frosting with the ability to be easily drizzled over a cake for a ganache coating. Maybe it needs to be heated up in some way or is already pourable and packaged to pour over a cake and harden slightly in the air temperature (think like Magic Shell hardens over ice cream).”

The network for the idea “Cookie dough” has only two nodes, and therefore one edge. The single edge creates the single-stepped shape of the dashed-line cumulative distribution function in the left graph. The network for the other, longer idea has many nodes and therefore

many edges. The cumulative distribution has more steps and the steps are smaller than in “Cookie dough.”

Figure 1: Cumulative Distribution Functions for Two Frosting Ideas

Each graph shows the prototypical edge weight cumulative distribution function (CDF) as a solid line. (In this example, the baseline semantic network is created from the ideas themselves.) The vertical value of the CDF is the fraction of edges weights in the semantic network that are less than or equal to the edge weight value on the x-axis.



We followed the explanation of Toubia and Netzer’s procedure to create the baseline semantic network for the topic and then the semantic subnetworks for each idea. Across studies, they demonstrate their effect with both hand-corrected text mining (e.g., separating or combining stems and removing generic terms from the network) and fully automated (as in their Study 5). We used the fully automated approach. We removed stop words from the semantic networks; we used the list of “SMART” stop words from the R package *tm*. That list of stop words includes generic verbs like became, goes, and follows; prepositions like after and between; and more typical stop words like the, he, and she. We used the “SMART” list rather than the shorter “English” stop words list to automate removal of generic terms. Also following their procedure, we used two variations on the reference corpus: the ideas themselves and Google results.

Table 1 shows the results from the simplest version of the analysis for these two data sets. In the results in the table, we use the ideas themselves to form the baseline semantic network. We also replicated the results using other controls that they report such as the minimum, maximum, and average edge weights and node frequencies, and using the other method for building the baseline semantic network, the text on the pages for the top search results of a Google query. The results with controls and the different baseline semantic network are in Appendix A. The unit of analysis is an idea, but two of the frosting ideas were a single word: “butterscotch” and “cinnamon.” Single word ideas have a single node and zero edges, so cannot be analyzed for the effects of edge weights, leaving 94 frosting ideas.

Table 1: Frosting and Household Products Ideas Analysis
Baseline semantic network derived from the ideas themselves.

	Frosting Ideas (Study 1a)		Household Products Ideas (Study 1b)	
	Dependent Variable	Dependent Variable	Dependent Variable	Dependent Variable
	Creativity	Purchase Intent	Creativity	Purchase Intent
Constant	3.68*** (0.082)	3.19*** (0.10)	3.45*** (0.096)	0.42*** (0.02)
Distance from prototypical distribution	-1.08*** (0.24)	-0.49 (0.28)	-2.42** (0.73)	-0.17 (0.15)
Adj. R ²	0.17	0.02	0.09	0.00
N	94	94	100	100

. p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Discussion. Toubia and Netzer’s main result replicates: an idea’s distance from the prototypical edge weight distribution is a highly significant predictor of the rated creativity. Confirming their finding, we also did not find a strong relationship for predicting purchase intent.

DISTANCE FROM PROTOTYPICAL PREDICTS CREATIVITY, BUT SO DOES LENGTH

Table 1 shows the simple relationship between the distance from the prototypical distribution and creativity. In Toubia and Netzer’s (2017) Table 2, they summarize many studies with a long set of controls. As shown in the variations in Appendix A, the main effect of distance and creativity does not always survive inclusion of all the controls. A plausible explanation for that discrepancy is that we are working with smaller data sets, with N close to 100, while their Ns are 173 to 1735, with the median N at 261.

In examining the relationships between the controls and the idea evaluation measures (purchase intent and creativity), we notice a strong role for the length of the idea description. We show subsets of the correlation matrices in Table 2.

Table 2: Correlation Matrix Subsets for Frosting and Household Products Ideas

Frosting with Ideas as Baseline					Household Products with Ideas as Baseline				
	(1)	(2)	(3)	(4)		(1)	(2)	(3)	(4)
(1) Creativity	1.000				(1) Creativity	1.000			
(2) Purchase Intent	0.117	1.000			(2) Purchase Intent	0.225	1.000		
(3) Dist from Proto Distribution	-0.424	-0.167	1.000		(3) Dist from Proto Distribution	-0.293	-0.095	1.000	
(4) Num Characters	0.414	0.120	-0.464	1.000	(4) Num Characters	0.307	0.057	-0.448	1.000

In both data sets, the correlation between the number of characters in an idea description and perceived creativity has about the same magnitude as the correlation between the distance from the prototypical distribution and creativity. (See bolded pairs in the correlation matrices.) In other words, the distance from the prototypical edge weight distribution predicts creativity of an

idea, but so does the length of the description. Further, the magnitude of the correlation between number of characters and distance (outlined in the correlation matrix) is similar or bigger.

Toubia and Netzer note, “It is important to control for the number of nodes and number of characters in the idea, as larger semantic subnetworks tend to have smoother distributions of edge weights, which tend to be more prototypical” (p. 8). That point is apparent in the comparison of the distributions of two ideas in Figure 1. As shown in their Table 2, their effect survives the inclusion of these controls.

In our own analyses, the focal relationship does not always remain significant with the inclusion of controls for idea description length. There is a good reason to expect that idea description length could influence perceived creativity of an idea. Extent of elaboration is a measure used to assess an *individual’s* creativity in idea generation (Torrance 1966). Ideas themselves could also be judged as more creative when they include more detail.

These observations raise the statistical question of whether idea description length serves as a mediator of the relationship between an idea’s distance from the prototypical distribution and perceived creativity. With our data sets, we find that it is. We examined both the frosting and household products data sets, with both the ideas and Google results as baseline corpora. Mediation analysis (using 5000 resamples, Preacher and Hayes 2004; Hayes 2012; Zhao et al. 2010) revealed partial mediation in the frosting data set using the ideas themselves as the reference corpus ($b_{axb} = -0.33$, 95% confidence interval: -0.60 to -0.15), and full mediation when using the Google search results as the reference corpus ($b_{axb} = -0.27$, 95% confidence interval: -0.50 to -0.11). Similarly, full mediation was found for the household products data set with the ideas as the reference corpus ($b_{axb} = -0.78$, 95% confidence interval: -1.55 to -0.21) and with the Google search results as the reference corpus ($b_{axb} = -0.50$, 95% confidence interval: -0.97 to -

0.18). Appendix B shows variations on the analysis. Taken together, these results suggest that ideas with a more typical balance of novel and familiar elements are considered more creative because they are longer.

DOES LENGTH PREDICT CREATIVITY? CORRELATIONAL EVIDENCE

To follow up on our conjecture that longer idea descriptions are perceived as more creative, we start with descriptive, correlational evidence. In the next sections, we present results from controlled experiments to establish causation.

Across nine data sets, we see a positive and highly significant relationship between idea description length (in characters) and rated creativity. The relationship between length and purchase intent is also positive, but much weaker. In this section we describe the data sets and present the correlational data.

Study 2, Data. Seven of the data sets come from Toubia and Netzer (2017). These data sets are available to researchers from *Marketing Science*. The topics of these sets include oral care, health related smartphone apps, and insurance. The data sets include metrics but not the text of the ideas. The other two data sets are the frosting ideas and the household products.

Results. In all nine data sets, we observe that longer descriptions are significantly positively correlated with higher creativity ratings. The relationship between length and purchase intent is less clear. In a few of the data sets, they are significantly positively correlated. The overall relationship from a meta-analysis of these nine studies between length and purchase intent is statistically significant. Using the R package *meta*, the function *metacor* reports a 0.13 overall

correlation between length and purchase intent for a random effects model ($p < 0.001$), and a 0.35 correlation between length and creativity ($p < 0.0001$).

Table 3: Summary of Correlations from Nine Studies (Study 2)

# of ideas in the data set and data set topic (TN indicates Toubia and Netzer 2017)	Correlation of length and purchase intent	Correlation of length and creativity
96 frosting products	0.12 ($p=0.26$)	0.42***
100 household products	0.05 ($p=0.59$)	0.31**
276 new insurance products related to aging (TN Study 1a)	0.10 ($p=0.11$)	0.34***
271 new insurance products related to financial security (TN Study 1b)	0.12.	0.44***
251 new insurance products related to unemployment (TN Study 1c)	0.12.	0.40***
555 health-related smartphone apps (TN Study 2)	0.15***	0.14***
173 health-related smartphone apps (TN Study 3)	0.21**	0.44***
220 oral care for women over 40 (TN Study 4)	-0.04 ($p=0.52$)	0.16*
648 health related smartphone apps (TN Study 6)	0.28***	0.46***

. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

With correlational evidence, we can't conclude whether longer ideas are perceived as more creative or ideas that are more creative take more words to express. There could even be a third explanation, that the more creative people both use more words and have more creative ideas. We next turn to experimental data to understand if longer ideas are perceived as more creative.

DOES LENGTH INCREASE PERCEIVED CREATIVITY? EXPERIMENTAL EVIDENCE

To establish evidence for a causal relationship—that making an idea longer increases the perception of creativity—we ran experiments manipulating the length of ideas.

In Study 3, we separate the effects of length and the inherent merit of the idea by evaluating three different versions of ideas. We used an original version, a lengthened version that added no information, and a lengthened version that added concrete but unsurprising details.

By collecting ratings on all three versions, we can see if the lengthened versions of the same idea are rated higher on purchase intent and creativity than the original version. We expected that the lengthened idea with the details would rate higher than the original, at least on creativity, based on the correlational evidence. And this study also allows us to see whether the lengthened but no more informative version is rated more like the original version or more like the lengthened version with details. If the lengthened but no more informative version is rated like the one with details, this would tell us that people use the length of an idea description as a signal of creativity.

Study 3, Data. In this study, we used a set of ideas for products for the college student market. The set of ideas were a random set of 29 ideas from a larger set that students generated for a course project for “physical products for college student market with retail price of under \$50.”

We worked with a subset of the ideas (only 29 rather than the full set, which had 290) for a few reasons beyond the obvious one of frugality. First, we have found that our respondent pool attends well to a task of reading and rating about 30 ideas. When the set of ideas is higher, closer to 50, attention filter pass rates drop off. Second, when all respondents rate the same set of ideas, we can look at consistency of the ratings across raters. Third, with a sample size of 29, we will only detect effects of at least moderate strength. At $N=29$, only correlations of greater than 0.37 are significant at $p<0.05$.

In this study, we created two variations on each idea, both longer than the original idea. One of the longer versions (Version B) was extended by repeating information in the original idea (Version A) and adding only obvious statements. The other longer version (Version C) included

additional concrete details, but ones that were straightforward elaborations on the original idea. The longer versions were 1.4 to 2.3 times as long as the original version.

Here is an example of the three versions of one of the ideas.

- Version A: *This product involves using durable plastic to replicate a large sized cardboard box, with in-built slots that allow shelves to create separate compartments within. This will solve the problem of storing fragile items with everything else. (239 characters)*
- Version B: *This product involves using durable plastic to replicate a large sized cardboard box, with in-built slots that allow shelves to create separate compartments within. This box will be made of plastic so it is durable and it will be pretty large. This compartment box will solve the problem of storing fragile items with everything else because the separate compartments will keeps fragile items in place. (403 characters, 1.7 times as long as Version A)*
- Version C: *This product involves using durable plastic to replicate a large sized cardboard box, with in-built slots that allow shelves to create separate compartments within. The compartments will have removable pads that line them. This will solve the problem of storing fragile items with everything else. The box should also have a handle with an ergonomically designed grip to make transporting the box easy. (403 characters, same as Version B, adding the unsurprising but concrete details about the pads and handle)*

Manipulation Check. First we checked that our construction of the three versions was perceived as intended. For each of the 29 ideas, we asked a few people to compare either Versions A and B

of the 29 ideas or Versions A and C. The people were randomly assigned to one of the conditions. For both conditions, the question was as follows. “The two versions describe the same idea, but the descriptions are not exactly the same. Considering the content of the descriptions, how big a DIFFERENCE do you see in the two versions?” The response options were (1) no or essentially no difference, (2) a small difference, (3) a moderate difference, and (4), a large difference. Our manipulation was confirmed. The average rating of the six people who compared versions A and B was 1.84 (i.e., between no difference and a small difference). The average rating of the three people who compared versions A and C was 2.31 (i.e., between a small difference and a moderate difference). The paired t-test for the 29 ideas gives $p < 0.0001$. This confirms our manipulation in two ways: for both comparisons, the differences are not perceived as large; and the difference between Versions A and B is smaller than the difference between Versions B and C.

Idea Ratings. Next we collected creativity, uniqueness, and purchase intent ratings of all three versions of the 29 ideas from Mechanical Turk workers. Each worker was randomly assigned to rate one block of 29 ideas, including exactly one of the three versions for each idea, with the versions interleaved in the blocks (i.e., block 1 had the pattern A, B, C, A, etc., block 2 had the pattern B, C, A, B, etc., and block 3 had the pattern C, A, B, C, etc.) We received 50-52 responses that passed the attention filter in each block.

Results. Table 4 shows the differences in the means on each of three measures—creativity, purchase intent, and uniqueness. The p-values are for the paired t-tests. We observe no statistical difference in purchase intent across the three versions. The original version (A) is less creative than the longer versions (B and C).

Table 4: Differences in Evaluations for Versions of Ideas for College Students (Study 3)

	Creativity	Purchase Intent	Uniqueness
Version A (Original)- B (Extended)	-0.10 ($p < 0.05$)	0.00 ($p = 0.97$)	-0.07 ($p = 0.12$)
Version A (Original)- C (Extended with Detail)	-0.17 ($p < 0.05$)	0.00 ($p = 0.99$)	-0.15 ($p < 0.05$)
Version B (Extended)- C (Extended with Detail)	-0.08 ($p = 0.23$)	0.00 ($p = 0.98$)	-0.07 ($p = 0.20$)

Discussion. Controlling for the content of the idea, we find that longer is perceived as more creative. Ideas that are expressed with more words, even if there is no additional information in the longer version, are considered more creative. This experimental evidence helps us determine the direction of the relationship between additional words and rating of creativity.

DOES LENGTH INCREASE PERCEIVED CREATIVITY? EXPERIMENTAL REPLICATIONS

We ran two other studies with a similar structure to Study 3 to examine the effect of the length of idea descriptions on the ratings of the ideas. We briefly report on them here and include more details in Appendix C. Both of these studies also used a random subset of 29 ideas from the 290 ideas for college student products. In Study 4, the three versions of the ideas were the original idea, a shortened version, and a lengthened-but-essentially-the-same-idea version. In Study 4, the lengthened version was lengthened in similar but less strict way than Study 3, mixing repetition, use cases, unsurprising details, and options to add length. In Study 5, the three versions of the ideas were the original idea, a lengthened version (in a similar manner to Study 4), and an alternative “wacky” lengthened version (explained below). The results were similar to Study 3 and the correlational evidence.

In Study 4, we collected ratings from 54 to 61 Mechanical Workers (those passing the attention filter) in each block. We collected ratings on four variables: need, purchase intent, creativity and uniqueness. For need, purchase intent, and creativity, average ratings by idea were

highest for the long version, then the original version, then the short version. (For uniqueness, short was less unique than the original, but the difference between original and long was not statistically significant.) Thus, we replicate the pattern that longer descriptions are considered more creative.

For purchase intent, Study 4 *does* show an effect of length; Study 3 did not. In the meta-analysis of our observational data, we did see a significant effect of length on purchase intent.

In Study 5, we first lengthened each idea in a similar way to Study 4. Then we used a procedure for lengthening the ideas in a “wacky” way. We selected a random page from Wikipedia using <https://en.wikipedia.org/wiki/Special:Random> and incorporated the subject of the page into the description of the idea in a way that still conformed to natural speech patterns, no matter how big the stretch for the content. Examples of the three versions of an idea:

- Version A (original idea): *Gloves with retractable tips that the user can pull back and thereby expose his fingertips, so as to allow normal use of his hands.* (131 characters)
- Version B (lengthened version including typical use case and unsurprising options): *Gloves with retractable tips that the user can pull back and thereby expose his fingertips, so as to allow normal use of his hands. This type of flexible design would be useful for people using smart phones in places where the weather is cold. The gloves would come in a variety of designs or colors.* (300 characters)
- Version C (lengthened wacky Wikipedia version): *Gloves with retractable tips that the user can pull back and thereby expose his fingertips, so as to allow normal use of his hands. A snail can stick his body out of his shell or retract it back into his shell. You*

should be able to do the same with putting your fingers in and out of your gloves. (297 characters) The random Wikipedia page was the one for Pteromeris, a snail.

We collected ratings on four variables—need, purchase intent, creativity, and uniqueness—from 58 to 64 Mechanical Turk workers (those passing the attention filter) in each block.

The first lengthened versions were considered more creative but did not have higher purchase intent or need measures than the original versions. The wacky Wikipedia versions, which were just as long as the first lengthened versions, were considered more unique than the original versions *but not more creative*. Need and purchase intent were not significantly different across the three blocks.

Additional consistent information, even if not incremental in substance—raises perceived creativity. Length added by incongruous information did not increase creativity or reduce purchase intent.

WORKING WITH CORE IDEAS

The results we presented above show that the perceived creativity of an idea description is sensitive to the length of the description. We see a much weaker effect of length on purchase intent. Likewise, Toubia and Netzer found strong results for predictions of creativity and some—but not complete—support for predictions of purchase intent. We find it not surprising, but still intriguing, to see such different patterns for creativity and purchase intent. Not surprising because creativity and purchase intent measure different constructs. But intriguing because, in the context of new products, they both represent desirable attributes. Toubia and Netzer (2017) invoke “beauty in averageness” to explain why prototypicality predicts desirability in the form of

creativity. But both originality (via creativity) and appeal (via purchase intent) plausibly represent “beauty” of new product ideas.

One remedy we propose for eliminating the distortion from peripheral detail in the ideas is to work with the *core ideas*. The standard definition of *core*—“the central or most important part of something”—applies in this context. Kornish and Ulrich (2014) define “raw ideas” as “the opportunity as conceived at the outset of the innovation effort in a specific organizational context. A raw idea for an innovation is often expressed in words or with a simple visual depiction.” A raw idea is raw in the sense of being in the original form expressed by the person who contributed it. A core idea is a condensed version of the raw idea, retaining only the essential elements. A core idea is stripped down so that we cannot remove further detail without materially changing the idea.

Mechanically, Toubia and Netzer’s method can be applied to core ideas. However, we don’t start with a strong prediction about whether the key result will obtain, that is, *whether the predictive power of the prototypical balance between close and far associations still holds for core ideas, not just the raw ones*. In this section, we show that key relationship between distance from prototypical distribution and creativity does *not* hold for the frosting and household products core ideas sets.

Study 6a and 6b, Semantic Network Analysis with Core Ideas. We created core versions of all the ideas in the frosting and household products data sets.¹ Mechanical Turk workers each

¹ In a study shown in Appendix D, we demonstrate that untrained assistants can reliably extract the core idea of a longer idea description.

rated a random subset of 29 ideas for purchase intent, creativity, and uniqueness. (In these studies, different people rated different ideas.) For the frosting ideas (Study 6a), an average of 55 people rated each of the 96 ideas (with a range of 39 to 72 raters). For the household products (Study 6b), an average of 39 people rated each of the 100 ideas (with a range of 27 to 53 raters).

Table 5 shows the results from the simplest analysis. After removing the “SMART” stop words, we retained all stems from the raw ideas for the baseline semantic network. Variations appear in Appendix A.

Table 5: Frosting and Household Products Core Ideas Analysis

Baseline semantic network derived from the original ideas (the raw ideas).

	Frosting Ideas (Study 6a)		Household Products Ideas (Study 6b)	
	Dependent Variable Creativity	Dependent Variable Purchase Intent	Dependent Variable Creativity	Dependent Variable Purchase Intent
Constant	3.45*** (0.17)	2.97*** (0.12)	3.50*** (0.13)	2.68*** (0.11)
Distance from prototypical distribution	-0.40 (0.33)	-0.15 (0.24)	-0.57 (0.33)	-0.43 (0.29)
Adj. R ²	0.00	0.00	0.02	0.01
N	94	94	100	100

. p<0.1, * p<0.05, ** p<0.01, *** p<0.001

With the core ideas, the distance from prototypical edge weight distribution only marginally predicts the creativity of the ideas for household products and does not predict creativity for frosting.

Discussion. There are three challenges to applying Toubia and Netzer’s method to core ideas.

The three challenges are equal treatment of all words, lumpy distributions, and identifying a relevant corpus for the baseline.

The first challenge of applying the method is that it *treats all words in an idea description the same*. But not all words are equally crucial to the idea. For example, one of the frosting ideas is for a frosting to be mixed with whipped cream. That idea also suggests alternative mixers: “or some dairy product (sour cream, cream cheese, yogurt).” Sour cream is a peripheral, not an essential, element of the idea. The list of dairy options are closely associated with one another. These peripheral words distort the balance of close and far associations.

Peripheral words in descriptions cannot be easily dismissed. The first reason is that in the longer descriptions, *most* of the words are peripheral. Say we need 5 out of 20 words in an original description to convey the core idea; that means that a majority, 15 out of 20, are peripheral. This majority becomes even bigger when we look at word pairs, as there are only 10 pairs of two essential words (10 is $5 \cdot 4 / 2$) and 180 pairs that include at least one peripheral word (180 is $20 \cdot 19 / 2 - 10$). The second reason we cannot quickly dismiss the role of the peripheral words is that as we showed in Study 3, peripheral content in idea descriptions adds length, systematically increasing the perception of creativity.

The second challenge for applying the method to core ideas is the small (few-node) networks of the semantic networks. To illustrate why a small semantic subnetwork interferes with the interpretation of distance from the prototypical distribution, we return to the examples from our frosting data set in Figure 1. Recall that these graphs show the prototypical cumulative distribution function of edge weights (solid black line) and the edge weight distribution for a single idea (dashed line). The idea on the left is “cookie dough” and the idea on the right is a much longer idea. Short descriptions like “cookie dough” have fewer nodes, so the step sizes in the distribution are bigger. A distribution with pronounced steps will tend to be farther away

from a prototypical distribution, which is generally filled with small steps, as it is created from an average of many distributions.

The third challenge for applying the method to core ideas is that it's harder to find an appropriate baseline corpus. The baseline corpus, either the text of the Google search results or the original, longer ideas themselves, contains many peripheral words. And words that are peripheral for one idea may well be essential for another idea. (For example, in the frosting concept, say one person has a frosting that can be spread on fruit and mentions examples of fruit, like banana; another person's core idea is banana-flavored frosting.) Thus, these baseline corpora do not provide robust information about the strength of association for elements of *core* ideas.

Both empirically and logically we conclude that the Toubia and Netzer's method does not help us understand what *patterns of close and far associations within the core of an idea* predict the evaluations of that idea. However, the question about those patterns is still relevant and important for core ideas. In the next section, we present some ways we tackled it.

MEASURING STRENGTH OF ASSOCIATION IN CORE IDEAS

The application of Toubia and Netzer's method to core ideas did not show how the strength of key associations predicts idea evaluations. Thus, we use different approaches to examine the relationship. We present the results of two studies of the relationship between the strength of key associations and idea evaluations, one on the core frosting ideas and another on a set of the college student product ideas.

Study 7a, Decomposition, Classification, and Suitability of Frosting Ideas. We decomposed the core versions of the 96 frosting ideas into their constituent parts. We understand that this

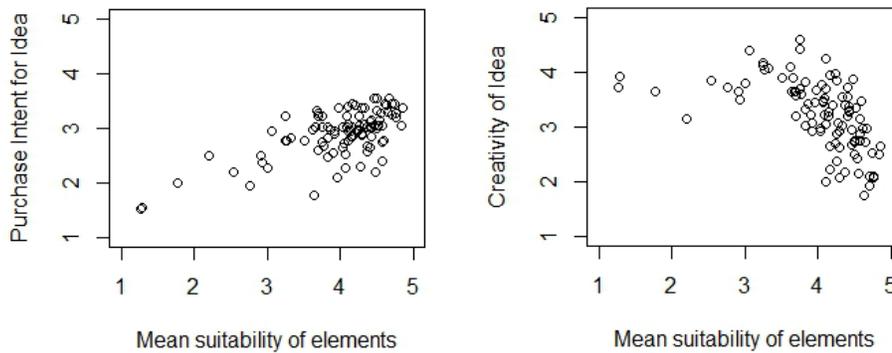
approach is not a scalable, big-data approach. Rather, it is our more direct but labor intensive way to understand the relationship between strength of associations and idea evaluations.

In addition to decomposing, we also classified each element into one of nine types, specific to the frosting category: a flavor (chocolate), an appearance (tie-dye), a consistency (whipped), an attribute (with vitamins), an add-on (sprinkles), a theme (patriotic), a use (for vegetables), an accessory (interchangeable tips), or packaging (squeezable tube).

Rather than finding the association between all pairs of elements (over 10,000 comparisons), we asked a much easier question, how “suitable” each element is for the frosting category. We note that not all ideas actually use the word frosting, so mechanically asking about all word pairs would overlook important relationships. For example, one of the ideas is “butterscotch,” meaning butterscotch flavored frosting. The association of the flavor (butterscotch) with the category (frosting) is a key combination. Asking about suitability for the category addresses this variation in the way people expressed the ideas. An average of 27.0 people rated each element for suitability (with a range of 24 to 30).

Results. The results show clear patterns for the relationships between suitability of an idea’s elements for the category and evaluations of the idea. The correlation between mean suitability (average across raters and elements of an idea) and purchase intent is 0.65 ($p < 0.0001$, $N = 96$). For creativity, correlation is -0.53 ($p < 0.0001$). Figure 2 contains the scatter plots showing these relationships. Unlike Toubia and Netzer, we don’t see support for the balance between close and far; instead, ideas with more appeal have closer associations in the sense of suitability. Unlike Myers-Levy and Tybout (1989), we don’t see support for moderate incongruity.

Figure 2: Mean Suitability of Frosting Idea Elements vs. Evaluations



Discussion. Raw ideas vary in how many elements a description has, and even determining the number of elements is a subjective call. In our decomposition, we treated multi-word phrases expressing a single idea as a single element. For example, we treated “strawberry banana” and “mint chocolate chip” each as a single element. These choices reflect our belief about the key combinations in the idea: mint and chocolate are not a key combination in the idea of mint chocolate chip frosting. It is the flavor and the category (frosting) that form the key combination.

Study 7b, Decomposition and Rating of College Student Product Ideas. We conceptually replicate this result with a different data set and a different method of assessing strength of associations. In this replication, we again selected a random set of 29 ideas from our list of 290 college student products. We had four workers from Mechanical Turk decompose the ideas into two elements, guiding them to consider an idea as either as a combination of two or more independent products, the addition of a feature or features to a base product, or any two elements of the idea: a feature, function, benefit, characteristic, setting, etc. One of the authors examined the four decompositions to determine the consensus decomposition.

With the ideas and decompositions in hand, we collected data in two waves. In the first wave, 54 people rated the ideas. In the second wave, 40 people did not see the ideas, just the two elements, and gave ratings for similarity and strength of association between the two elements.

Results. Closer associations are perceived as better ideas. The correlation between the average rating of “how good is this idea” and “how strong is the association between (Element 1) and (Element 2)” for the N=29 ideas is 0.39 ($p < 0.05$). (As explained in Appendix C, the respondents who answered the association question did not see the idea descriptions, just the elements.) We see the same results ($r = 0.43$, $p < 0.05$) using the similarity measure from the second wave instead of the association measure.

Discussion. In studying strength of association and evaluations of ideas, we have to be cautious of a familiarity effect. Two elements of an idea that typically appear together in existing products (e.g., a refrigerator with an ice maker, a phone with a camera) are also likely to be rated as good ideas. We do not believe that is the issue with this set of products, as the combinations of elements aren't standard issue for any of the product categories. For example, ideas in the set include the hairbrush with a removable layer and sunglasses with an ice pack sleeve. Appendix C shows several of the ideas.

Overall Discussion. Toubia and Netzer's method is robust: we easily replicated them with our two data sets. However, our reanalysis supports a different interpretation of their results. Because their method operates largely on associations of peripheral details, it can be used to improve the perceived creativity of the idea description (as they demonstrate in their Study 6). But this result does not mean that a prototypical balance between close and far associations for the *essential*

elements predict better ideas. In fact, the evidence we have suggests something else, that close associations make for good ideas.

CONCLUSION

In this work, we demonstrate that different ways of expressing the same fundamental idea systematically change the evaluation of that idea. In particular, longer ideas are perceived as more creative than shorter ones. This is true even when the longer version contain no more information than the shorter one, and especially when the longer version contains concrete and unsurprising details. Other measures of idea evaluation, like purchase intent, are less susceptible to such distortion. These findings are important because of three trends in innovation: the emphasis on creativity, the use of crowdsourcing, and the use of text-based data analysis.

Moreau and Engeset (2016) make the case for the importance of creativity for success in career pursuits. We don't dispute that ability to find superior solutions is valuable, and that creativity helps one do that better. However, when it comes to evaluating ideas for new products, we find that perceived creativity is more sensitive to cosmetic differences than measures of appeal like purchase intent, goodness, and strength of need.

In idea selection tasks for new products, there is general agreement in practice and in the academic literature about criteria for selection. In concept screening and concept testing, it has long been standard to ask a purchase intention question of prospective consumers (Moore 1982, Ulrich and Eppinger 2015). Moore (1982) describes the typical practice as using a primary scale measuring appeal, such as "intention to purchase, interest, or, liking," and "possibly a few secondary scales such as uniqueness or believability." Other scholars also use multiple measures,

e.g., Diehl and Stroebe (1987) use originality and feasibility, Dahl and Moreau (2002) use originality and perceived value, and Girotra, Terwiesch, and Ulrich (2010) primarily analyze business value and purchase intent, but also examine feasibility, novelty, and specificity.

Runco and Jaeger (2012) emphasize that the standard definition of creativity itself has two elements, which they label as originality and effectiveness. In our data, we see that uniqueness and appeal jointly predict creativity, but that the standardized weight on uniqueness is much higher than the weight on appeal, that is, that creativity measures originality more than it measures effectiveness. (Analyses omitted for brevity.) Our results warn against placing too much weight on creativity or uniqueness, as the perceptions of those are sensitive to uninformative filler in the description. In a survey of academic papers on idea generation in top marketing journals, we see that creativity is a commonly measured dependent variable.

Table 6: Use of Creativity Measures in Marketing Articles about Idea Generation

These results reflect articles that rate product ideas. We searched in Web of Science in five journals, from 1999 to 2019, for these terms: “new products,” “idea generation,” or “ideation.” See full list of papers in Appendix E.

	Papers with Creativity, Originality, or Novelty as a Focal Measure	Total Papers
J. Consumer Res.	1 (25.0%)	4
J. Marketing	5 (71.4%)	7
J. Marketing Res.	7 (63.6%)	11
Marketing Science	3 (60.0%)	5
Management Science	2 (15.4%)	13
Total	18 (45.0%)	40

A second important trend in the innovation landscape is crowdsourcing. Crowdsourcing is the practice of having many people contribute ideas for an innovation effort. When ideas are written by many different people, there is great variation in style. Evaluating the raw ideas is a challenge. Kornish and Ulrich (2014) show that getting 50 purchase intent ratings from the

general population is a reasonably good way to predict market outcomes. The results in the current paper suggest that the more standardized the submitted idea descriptions are, the better. One important way to standardize is on length. Large innovation enterprises “optimize” ideas and present them in a consistent format, but this refinement is usually done after the fuzzy front end, once the set of ideas have been narrowed to a manageable set.

A third trend in all of business and society, not just innovation, is greater availability and use of data to support decision making. The work of Toubia and Netzer (2017) provides a general, scalable way to analyze the content of the ideas. Indeed, they have shown a robust pattern. But our story is a simpler account: idea description length predicts perceived creativity.

Surface details matter in many domains. Cosmetic—as opposed to essential—elements of idea descriptions have a real impact on the way the ideas are evaluated. This is true whether the ideas are analyzed with textual analysis or evaluated by human raters. In this paper, we have described considerations to help make us better able to evaluate which raw ideas hold promise.

REFERENCES

- Åstebro T, Elhedhli S (2006) The Effectiveness of Simple Decision Heuristics: Forecasting Commercial Success for Early-Stage Ventures. *Management Science* 52(3):395-409.
- Dahl DW, Moreau CP (2002) The influence and value of analogical thinking during new product ideation. *J. Marketing Res.* 39(1):47–60.
- Diehl M, Stroebe W (1987) Productivity loss in brainstorming groups: Toward the solution of a riddle. *J. Personality and Soc. Psych.* 53(3):497-509.
- Eliashberg J, Hui SK, Zhang ZJ (2007) From Story Line to Box Office: A New Approach for Green- Lighting Movie Scripts. *Management Science*, 53(6):881–93.
- Girotra K, Terwiesch C, Ulrich KT (2010) Idea generation and the quality of the best idea. *Management Sci.* 56(4):591-605.
- Goldenberg J, Lehmann DR, and Mazursky D (2001) The Idea Itself and the Circumstances of Its Emergence as Predictors of New Product Success. *Management Science*, 47(1):69–84.
- Goldenberg J, Mazursky D, Solomon S (1999) Toward identifying the inventive templates of new products: a channeled ideation approach. *J. Marketing Res.* 36(2):200–210.
- Hayes AF (2009) Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs* 76(4):408-420.

- Kornish LJ, Ulrich KT (2014) The importance of the raw idea in innovation: Testing the sow's ear hypothesis," *J. Marketing Res.* 51(1):14-26.
- Meyers-Levy J, Tybout AM (1989) Schema congruity as a basis for product evaluation. *J. Consumer Res.* 16(1):39–54.
- Moore WL (1982) Concept testing. *J. Bus. Res.* 10(3):279-294.
- Moreau CP, Engeset MG (2016) The downstream consequences of problem-solving mindsets: How playing with LEGO influences creativity. *J. Marketing Res.* 53(1):18-30.
- Poetz MK, Schreier M (2012) The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas? *J. Prod. Innov. Manag.* 29(2):245–256.
- Preacher KJ, Hayes AF (2004) SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Res. Methods, Instruments, and Computers* 36(4):717-731.
- Rosenzweig S, Tellis GJ, Mazursky D (2015) Where Does Innovation Start: With Customers, Users, or Inventors? *Marketing Science Institute Working Paper Series 2015 Report No.* 15-108.
- Runco MA, Jaeger GJ (2012) The standard definition of creativity. *Creativity Res. J.* 24(1):92-96.
- Terwiesch C, Ulrich KT (2009) *Innovation Tournaments: Creating and Selecting Exceptional Opportunities* (Harvard Bus. Press, Boston, MA).

Toubia O, Netzer O (2017) Idea generation, creativity, and prototypicality. *Marketing Sci* 36(1):1-20.

Torrance, EP (1966) *Torrance Tests of Creative Thinking* (Personnel Press, Princeton, NJ).

Ulrich KT, Eppinger SD (2015) *Product Design and Development*, 6th ed. (McGraw-Hill Education, New York, NY).

Zhao X, Lynch JG, Chen Q. Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *J. Consumer Res.* 37(2):197-206.

APPENDIX A: ANALYSIS VARIATIONS

In each of Tables A-1 through A-4, we show variations on the analysis to supplement the simple results shown in Table 1 and in Table 5. Models (1) and (2) use the ideas themselves as the baseline semantic network, and Models (3), (4), and (5) use the text of the Google search results pages as the baseline semantic network.

Models (1) and (4) use the subset of controls in Toubia and Netzer's (2017) analysis that we felt had the strongest theoretical basis for inclusion. We include the mean and dispersion (coefficient of variation) of edge weights and node frequencies. Edge weights capture the strength of associations, which Toubia and Netzer argue is an essential step in innovation. Node frequencies capture how common the words are. Because familiarity can improve evaluation, we find a good reason to believe that frequency could drive evaluation. We omit the minimum and maximum values as controls, as the sum of minimum and maximum is strongly correlated with the average. Likewise, we omit the size of the semantic subnetwork as that is conceptually similar to the number of characters in the idea description.

Key Findings from Tables A-1 to A-4

Table A-1, Study 1a	Significance of distance for predicting creativity survives inclusion of controls for one of the baselines (the ideas) but not the other (Google results).
Table A-2, Study 1b	Significance of distance for predicting creativity does not survive the inclusion of controls for either baseline.
Table A-3, Study 6a	Distance does not predict creativity in four of the five models, and it is marginal in the fifth. Length is still a significant predictor of creativity in all five models. The coefficient of variation on the length in the frosting core ideas is 0.48 (mean length is 39.7 characters, standard deviation is 19.1).
Table A-4, Study 6b	Neither length nor distance predict creativity in any model. The coefficient of variation on the length in the household product core ideas is 0.17 (mean length is 95.5 characters, standard deviation is 16.6).

Table A-1: Frosting Ideas Analysis Variations (Study 1a)

The dependent variable is creativity.

	(1)	(2)	(3)	(4)	(5)
	Ideas	Ideas	Google Results	Google Results	Google Results
Constant	3.399*** (0.242)	3.01** (1.001)	3.651*** (0.087)	4.316*** (0.458)	-5.626 (16.601)
Distance from prototypical distribution	-1.759*** (0.390)	-2.165** (0.64)	-0.655** (0.184)	-0.005 (0.386)	0.047 (0.473)
Mean Edge Weight	1.806*** (0.463)	2.831* (1.155)		-0.209 (0.632)	1.883 (1.585)
CV Edge Weight	-0.003* (0.002)	-0.003 (0.002)		0.004 (0.003)	0.002 (0.004)
Min Edge Weight		1.698 (1.577)			-1.974. (1.051)
Max Edge Weight		0.448 (0.274)			-0.505 (0.687)
Mean Node Freq.	2.981*** (0.785)	8.782*** (2.132)		-6.345** (2.351)	-3.436 (2.763)
CV Node Freq.	-0.001 (0.001)	0.008* (0.004)		0.005 (0.004)	0.002 (0.009)
Min Node Freq.		-18.466* (8.194)			-40.486** (14.815)
Max Node Freq.		-1.616* (0.707)			-0.143 (0.807)
Mean Clustering Coeff.		-3.466* (1.34)			21.545 (37.834)
CV Clustering Coeff.		-0.006 (0.006)			-0.336 (0.413)
Min Clustering Coeff.		3.267 (1.984)			-17.817 (23.987)
Max Clustering Coeff.		0.665 (0.6)			11.312 (17.989)
Size of Semantic Subnetwork		0.006 (0.022)			-0.037 (0.029)
Number of Characters	0.001* (0.000)	0.001 (0.001)		0.001* (0.000)	0.001. (0.001)
Adj. R ²	0.33	0.41	0.12	0.24	0.28
N	94	94	89	89	89

. p<0.1, * p<0.05, ** p<0.01, *** p<0.001

When ideas themselves serve as the baseline network, we retain all 428 stems. When Google results serve as the baseline network, we retain the 783 stems that show up in five or more of the 50 documents. (There are 5996 stems in total.) In that analysis, of the 853 non-unique stems in the ideas, 620 of them (73%) are in the baseline semantic network. We see similar patterns if we retain only the 323 stems that show up in ten or more of the 50 documents.

Table A-2: Household Products Ideas Analysis Variations (Study 1b)

The dependent variable is creativity.

	(1)	(2)	(3)	(4)	(5)
	Ideas	Ideas	Google Results	Google Results	Google Results
Constant	3.060*** (0.605)	4.327 (2.65)	3.305*** (0.083)	1.143 (0.833)	-46.615 (29.393)
Distance from prototypical distribution	-2.010. (1.106)	-1.386 (1.252)	-1.084* (0.543)	0.339 (0.660)	1.601. (0.851)
Mean Edge Weight	1.036 (2.390)	2.378 (4.153)		7.496** (2.325)	9.948** (3.024)
CV Edge Weight	-0.004 (0.005)	0.007 (0.012)		0.010 (0.007)	-0.002 (0.01)
Min Edge Weight		16.75 (15.828)			-2.709 (2.419)
Max Edge Weight		0.051 (0.572)			0.597 (0.46)
Mean Node Freq.	9.029 (8.110)	4.247 (9.309)		-0.312 (2.357)	-2.084 (3.017)
CV Node Freq.	0.000 (0.005)	.016. (0.009)		-0.001 (0.004)	-0.008 (0.009)
Min Node Freq.		NA			-6.518 (16.538)
Max Node Freq.		-2.712 (2.454)			0.477 (0.749)
Mean Clustering Coeff.		-6.612 (4.117)			40.789. (24.234)
CV Clustering Coeff.		-0.007 (0.028)			-0.426 (0.358)
Min Clustering Coeff.		-2.575 (8.99)			6.465 (26.534)
Max Clustering Coeff.		NA			5.973 (9.468)
Size of Semantic Subnetwork		0.003 (0.007)			0.000 (0.01)
Number of Characters	0.000* (0.000)	0.000 (0.000)		0.000*** (0.000)	0.000 (0.000)
Adj. R ²	0.10	0.09	0.03	0.14	0.14
N	100	100	100	100	100

. p<0.1, * p<0.05, ** p<0.01, *** p<0.001

When ideas themselves serve as the baseline network, we retain all 1766 stems. When Google results serve as the baseline network, we retain the 857 stems that show up in five or more of the 50 documents. (There are 6354 stems in total.) In that analysis, of the 4000 non-unique stems in the ideas, 2101 of them (53%) are in the baseline semantic network. We see similar patterns if we retain only the 316 stems that show up in ten or more of the 50 documents.

Table A-3: Core Frosting Ideas Analysis Variations (Study 6a)

The dependent variable is creativity.

	(1)	(2)	(3)	(4)	(5)
	Original Ideas	Original Ideas	Google Results	Google Results	Google Results
Constant	2.359*** (0.356)	3.007*** (0.799)	3.319*** (0.178)	3.264*** (0.857)	-48.600* (24.272)
Distance from prototypical distribution	0.864. (0.447)	0.353 (0.565)	-0.094 (0.268)	0.150 (0.548)	-0.339 (0.585)
Mean Edge Weight	1.117. (0.631)	-0.287 (1.725)		-0.171 (0.670)	1.568 (2.525)
CV Edge Weight	0.002. (0.001)	0.000 (0.002)		-0.003 (0.003)	-0.001 (0.006)
Min Edge Weight		-1.025 (1.808)			-0.250 (1.822)
Max Edge Weight		0.702 (0.535)			-0.777 (1.351)
Mean Node Freq.	-0.978 (10.442)	8.655 (12.739)		-0.595 (3.912)	1.125 (4.975)
CV Node Freq.	0.001 (0.002)	0.004 (0.003)		0.003 (0.008)	0.007 (0.013)
Min Node Freq.		NA			NA
Max Node Freq.		-2.176 (1.65)			-0.781 (1.478)
Mean Clustering Coeff.		0.387 (1.614)			151.770. (78.306)
CV Clustering Coeff.		-0.004 (0.005)			-0.177 (0.920)
Min Clustering Coeff.		-4.226* (1.872)			-62.696 (63.076)
Max Clustering Coeff.		0.219 (0.873)			-32.729 (57.561)
Size of Semantic Subnetwork		-0.186* (0.086)			-0.104 (0.140)
Number of Characters	0.011** (0.004)	0.017** (0.006)		0.014** (0.004)	0.016** (0.005)
Adj. R ²	0.11	0.18	-0.01	0.08	0.09
N	94	94	83	83	83

. p<0.1, * p<0.05, ** p<0.01, *** p<0.001

When original ideas serve as the baseline network, we retain all 428 stems. When Google results serve as the baseline network, we retain the 783 stems that show up in five or more of the 50 documents. (There are 5996 stems in total.) In that analysis, of the 375 non-unique stems in the ideas, 286 of them (76%) are in the baseline semantic network.

Table A-4: Core Household Products Ideas Analysis Variations (Study 6b)

The dependent variable is creativity.

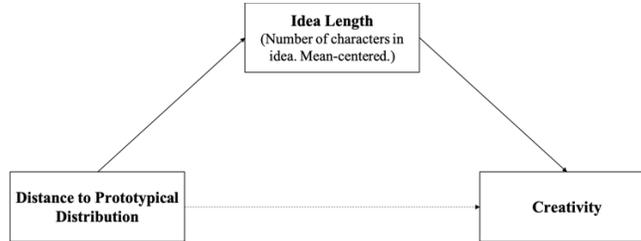
	(1)	(2)	(3)	(4)	(5)
	Original Ideas	Original Ideas	Google Results	Google Results	Google Results
Constant	3.723*** (0.320)	3.462*** (0.78)	3.228*** (0.104)	3.420*** (0.420)	-10.103 (14.247)
Distance from prototypical distribution	-0.401 (0.446)	-0.266 (0.699)	0.213 (0.249)	0.166 (0.280)	-0.569 (0.449)
Mean Edge Weight	0.530 (0.556)	-1.095 (1.589)		-0.107 (0.996)	2.604 (2.637)
CV Edge Weight	0.000 (0.002)	-0.004 (0.004)		-0.001 (0.002)	0.000 (0.004)
Min Edge Weight		-0.152 (2.453)			-1.190 (2.295)
Max Edge Weight		0.192 (0.398)			-1.362 (1.113)
Mean Node Freq.	0.126 (11.436)	7.637 (15.385)		0.105 (2.103)	-0.838 (2.704)
CV Node Freq.	-0.003 (0.002)	-0.003 (0.005)		-0.002 (0.004)	-0.009 (0.010)
Min Node Freq.		NA			-6.883 (7.235)
Max Node Freq.		-0.106 (2.621)			0.936 (1.075)
Mean Clustering Coeff.		0.723 (1.33)			19.141 (19.048)
CV Clustering Coeff.		-0.01 (0.009)			-0.022 (0.305)
Min Clustering Coeff.		-1.400 (1.342)			12.177 (25.568)
Max Clustering Coeff.		1.220 (0.768)			-14.673 (15.723)
Size of Semantic Subnetwork		-0.046 (0.037)			-0.081 (0.068)
Number of Characters	0.003 (0.003)	0.003 (0.003)		0.002 (0.003)	0.000 (0.003)
Adj. R ²	0.05	0.05	-0.00	-0.05	0.01
N	100	100	95	95	95

. p<0.1, * p<0.05, ** p<0.01, *** p<0.001

When original ideas serve as the baseline network, we retain all 1766 stems. When Google results serve as the baseline network, we retain the 857 stems that show up in five or more of the 50 documents. (There are 6354 stems in total.) In that analysis, of the 836 non-unique stems in the ideas, 426 of them (51%) are in the baseline semantic network.

APPENDIX B: MEDIATION ANALYSIS DETAILS

This appendix shows variations on the mediation analyses for the frosting and household products data sets. In **Model 1**, the distance to the prototypical distribution (“ks distance”) is the independent variable, the number of characters in the idea is the mediator, and creativity is the dependent variable.



Model 2 includes the set of controls we described in Appendix A as having the strongest theoretical basis for inclusion (average edge weight, coefficient of variation (CV) of edge weight, average node frequency, CV of node frequency). **Model 3** includes the full set of controls, as listed in the tables in Appendix A. All analyses performed using a bootstrap analysis using 5000 resamples.

Table B-1: Variations on the Mediation Analysis

	Model 1	Model 2	Model 3
	b_a*b	b_a*b	b_a*b
	(95% CI)	(95% CI)	(95% CI)
Frosting Products			
Baseline semantic network: 94 ideas	-0.33*	-0.11*	-0.14
	(-0.60 to -0.15)	(-0.46 to 0.10)	(-0.37 to -0.01)
Baseline semantic network: Google search results	-0.27**	-0.04**	-0.42
	(-0.50 to -0.11)	(-0.08 to 0.26)	(-0.84 to -0.16)
Household Products			
Baseline semantic network: 100 ideas	-0.78**	-0.21**	-1.05
	(-1.55 to -0.21)	(-0.89 to 0.29)	(-2.11 to -0.29)
Baseline semantic network: Google search results	-0.50**	-0.11**	-1.16
	(-0.97 to -0.18)	(-0.60 to 0.12)	(-2.27 to -0.57)

* Partial mediation, ** Full mediation

APPENDIX C: DETAILS OF ALL STUDIES AND SUPPLEMENTAL ANALYSES

Stimuli for Study 3

In this study, we used a random sample of 29 ideas from a set of 290 ideas that students generated as part of an innovation class. The prompt was to develop ideas for physical products for the college student market that would retail for \$50 or less. Below we show five of the original ideas and their two additional versions.

#	Version A: Original Idea	Version B: Idea Lengthened via Repetition	Version C: Idea Lengthened with Concrete but Unsurprising Details
1	This product involves using durable plastic to replicate a large sized cardboard box, with in-built slots that allow shelves to create separate compartments within. This will solve the problem of storing fragile items with everything else.	This product involves using durable plastic to replicate a large sized cardboard box, with in-built slots that allow shelves to create separate compartments within. This box will be made of plastic so it is durable and it will be pretty large. This compartment box will solve the problem of storing fragile items with everything else because the separate compartments will keep fragile items in place.	This product involves using durable plastic to replicate a large sized cardboard box, with in-built slots that allow shelves to create separate compartments within. The compartments will have removable pads that line them. This will solve the problem of storing fragile items with everything else. The box should also have a handle with an ergonomically designed grip to make transporting the box easy.
2	A device that allows you to rip a specified amount of toilet paper with one hand.	A device that allows you to rip a specified amount of toilet paper with one hand. This toilet paper device allows selecting the exact amount of paper you want, used easily with only one hand.	A device that allows you to rip a specified amount of toilet paper with one hand. The device contains a metal hood that hangs over the toilet paper roll, and the hood has a gently serrated edge.
3	Instead of 4 separate shakers containing garlic, oregano, red pepper, and parmesan cheese, there would be one large shaker with four separate compartments. The holes would be different sizes depending on the contents of that compartment.	Instead of 4 separate shakers containing garlic, oregano, red pepper, and parmesan cheese, there would be one large shaker with four separate compartments. The holes would be different sizes depending on the contents of that compartment so that garlic, oregano, red pepper, and parmesan don't get stuck in the holes. The shaker would be all-in-one and pretty large in size to hold all four of the toppings.	Instead of 4 separate shakers containing garlic, oregano, red pepper, and parmesan cheese, there would be one large shaker with four separate compartments. The holes would be different sizes depending on the contents of that compartment. On the outside of the shaker would be a label saying which topping was in which compartment. The lettering of the labels would have a red and green pizza-parlor design.
4	This hockey puck shaped, battery powered, LED flash light will be mounted to a shoulder strap of a backpack or any type of carrying bag. This will increase the safety when walking from school to home at night.	This hockey puck shaped, battery powered, LED flash light will be mounted to a shoulder strap of a backpack or any type of carrying bag. This flash light would have a bright LED light and would not need to be charged because it lights up using battery power. This will increase the safety when walking from school to home at night because you can hook it on to your belongings.	This hockey puck shaped, battery powered, LED flash light will be mounted to a shoulder strap of a backpack or any type of carrying bag. The mount would be a jaw-bone style clip that would be removable. You could move the light from bag to bag. Always having the light available will increase safety when walking from school to home at night, lighting up the way along the path.
5	Attach sensors to keys, iPod, telephone, PDA so that they can be easily located. Program a hand-held device to locate the sensors on the objects. The hand-held keychain determines how far the object is by speeding up the frequency of the beeps.	Attach sensors to keys, iPod, telephone, PDA so that they can be easily located. Keys, iPods, phones, and other devices or objects are often put down and need to be located. Program a hand-held device to locate the sensors on the objects. The hand-held keychain determines how far the object is by speeding up the frequency of the beeps.	Attach sensors to keys, iPod, telephone, PDA so that they can be easily located. Program a hand-held device to locate the sensors on the objects. The hand-held keychain determines how far the object is by speeding up the frequency of the beeps. The beeps would sound like musical notes, and they would get louder as they got more frequent.

In the first wave, respondent were randomly assigned to one of two blocks:

- Block 1: Compared Version A and Version B of each idea.
- Block 2: Compared Version A and Version C of each idea.

Ideas within a block were shown in random order. In both blocks, the respondents answered the following question about each pair of idea versions:

- The two versions describe the same idea, but the descriptions are not exactly the same. Considering the content of the descriptions, how big a DIFFERENCE do you see in the two versions? (4 point scale: No or essentially no difference, A small difference, A moderate difference, A large difference)

Twelve people completed the survey. Of them, nine passed the attention filter.

In the second wave, respondents saw one of three blocks:

- Block 1: One version of each idea, interleaved Versions A, B, C, A, etc.
- Block 2: One version of each idea, interleaved Versions B, C, A, B, etc.
- Block 3: One version of each idea, interleaved Versions C, A, B, C, etc.

Ideas within a block were shown in a random order.

Each idea was shown with these three questions:

1. How likely would you be to purchase a product based on this idea?
(5 point scale from Definitely would not purchase to Definitely would purchase).
2. How different is this idea from products on the market today?
(4 point scale: There is nothing else like it, There are other items that are something like this, There are other items that are just like this, There are a lot of other items that are just like this)
3. This idea is creative.
(5 point scale from Strongly agree to Strongly disagree)

At the end of a survey, we always used a set of attention filter questions. In this study, 180 people completed the survey and 153 of them passed the attention filter.

The attention filter questions mimicked the questions in the main block. However, instead of an idea description, the attention filter question showed this instruction (and then the same questions with the same options): “For all of the following questions about this idea, please select the second option: probably would not purchase, there are other items that are something like this, etc.”

We used this format for the attention filter in every study. We varied *which* options (top, second, etc.) we instructed the respondents to select.

Supplemental Analysis for Study 3: Rater Agreement

Table C-1: Cronbach’s Alpha by Block and Question, Study 3

The respondents are the “items” in the alpha calculation because we are interested in consistency of the raters. (The rows are the 29 ideas in a block.)

	Creativity	Purchase Intent	Uniqueness
Block 1 (N=50 responses)	0.79	0.86	0.93
Block 2 (N=50 responses)	0.80	0.84	0.91
Block 3 (N=52 responses)	0.79	0.86	0.91

Stimuli for Study 4

In this study, we used a different random sample of 29 ideas from the 290 ideas for products for college students. (See details in the description of Study 3.) Below we show five of the original ideas and the two additional versions from the studies.

#	Version A: Original Idea	Version B: Idea Lengthened via Repetition and Unsurprising Details	Version C: Idea Shortened
1	In most of the rooms, we have lamps. We also have photo frame, clocks and other stuff. Looking into clear night sky is so wonderful. If we can have some sort of lamps, which can address all these issues together, then it will be wonderful.	In most of the rooms, we have lamps. We also have photo frame, clocks and other stuff. Looking into clear night sky is so wonderful. If we can have some sort of lamps, which can address all these issues together, then it will be wonderful. It would be a stylish and multipurpose lamp that would enhance the decor in the room with patterns and decorations that are visually appealing. There are a lot of ways to accomplish this goal of combining useful room decorations like frames and clocks with lighting for a room. What room doesn't need more light and more decoration?	Stylish and multipurpose lamp
2	A spin sponge would provide a good way to reduce the amount of effort required to wash dishes. With interchangeable heads, there can be multiple attachments. A brush attachment can be used to clean tougher dish stains.	A spin sponge would provide a good way to reduce the amount of effort required to wash dishes. With interchangeable heads, there can be multiple attachments. A brush attachment can be used to clean tougher dish stains. This product would make dish cleaning faster, more efficient, and less messy. Your hands would spend less time in the soapy water, which would be better for your skin. Plus, the spinning action would be better than simple hand wiping at getting any dried food off of a plate or any greasy spots.	Spin sponge for washing dishes
3	A drinking glass that resembles a measuring cup, with tick marks for calories instead of ounces. Since students probably consume the most calories from alcohol,	A drinking glass that resembles a measuring cup, with tick marks for calories instead of ounces. Since students probably consume the most calories from alcohol, another possibility would be measuring different calorie levels specific to a type of drink. The tick marks would allow the user to be aware of	Calorie-measuring drinking glass

another possibility would be measuring different calorie levels specific to a type of drink.

and to control how many calories he or she is consuming from beverages. That type of information would aid self control in consumption of high calorie drinks, and help consumers make better decisions about the quantity and quality of energy that they put into their bodies. There is a direct relationship between how much people consume and how much exercise is required to burn off the intake. This calibrated cup makes that trade-off more explicit, so someone might wonder whether an extra drink is really worth the three miles of jogging or hour-long exercise class that would expend that much energy.

This earplug/alarm is targeted to people who sleep in noisy environments and want to block out the noise without worrying they won't hear their alarm. The user would be able to program when the alarm should go off.

This earplug/alarm is targeted to people who sleep in noisy environments and want to block out the noise without worrying they won't hear their alarm. The user would be able to program when the alarm should go off. And since the alarm will emit its sound from the earplug, there is no way that the alarm will blend in with background noise. This will help ensure that the alarm serves its purpose of waking up the sleeper. Having the alarm integrated with the earplug makes for the most efficient delivery possible for the noise of the alarm because the eardrum is very close by. This solution would drastically reduce the chance of sleeping through your alarm.

Earplug with built-in alarm

- 5 A bottle opener that has a compartment or attachable contraption to automatically hold bottle caps once opened. Theoretically, this would make the cleanup of a party much easier, because there would be far fewer beer caps strewn about.

A bottle opener that has a compartment or attachable contraption to automatically hold bottle caps once opened. Theoretically, this would make the cleanup of a party much easier, because there would be far fewer beer caps strewn about. As the bottle opener pries the cap off the bottle, the cap could be placed in the compartment. Disposing of all the caps collected in the compartment at once would be much easier than picking them up off the ground one by one or even throwing them out one at a time as you remove them.

Bottle opener with cap catcher

Respondents saw one of three blocks:

- Block 1: One version of each idea, interleaved Versions A, B, C, A, etc.
- Block 2: One version of each idea, interleaved Versions B, C, A, B, etc.
- Block 3: One version of each idea, interleaved Versions C, A, B, C, etc.

Ideas within a block were shown in a random order.

Each idea was shown with these four questions:

1. How much do you, personally, experience the need that this idea is designed to address?
(4 point scale: Not at all, Not none, but not much either, Somewhat, A great deal)
2. How likely would you be to purchase a product based on this idea?
(5 point scale from Definitely would not purchase to Definitely would purchase).
3. How different is this idea from products on the market today?
(4 point scale: There is nothing else like it, There are other items that are something like this, There are other items that are just like this, There are a lot of other items that are just like this)
4. This idea is creative.
(5 point scale from Strongly agree to Strongly disagree)

We used our standard attention filter questions. In this study, 204 people completed the survey and 173 of them passed the attention filter.

Table C-2: Results from Study 4

	Need	Purchase Intent	Creativity	Uniqueness
Version A (original)	1.94	2.19	3.30	3.00
Version B (longer)	2.07	2.33	3.47	3.05
Version C (shorter)	1.84	2.00	2.70	2.60
Paired t-test A vs. B p-value	p<0.0001	p<0.01	p<0.01	p=0.14
Paired t-test B vs. C p-value	p<0.0001	p<0.0001	p<0.0001	p<0.0001
Paired t-test A vs. C p-value	p<0.05	p<0.001	p<0.0001	p<0.0001

Stimuli for Study 5

Study 5 is exactly the same as Study 4 in the structure of blocks, the questions asked, and the attention filter. It used the same sample of 29 ideas (from the 290 products for the college students market) as in Study 3. In this study, Version C was a lengthened version that included a reference to a topic from a randomly selected Wikipedia page. Below we show five of the original ideas, the two additional versions from the studies, and the random Wikipedia page (with the full URL starting with <https://en.wikipedia.org/wiki/> before the page name).

#	Version A: Original Idea	Version B: Idea Lengthened via Repetition and Unsurprising Details	Version C: Idea Lengthened with Random Detail from Wikipedia	Random Wikipedia Page
1	This product involves using durable plastic to replicate a large sized cardboard box, with in-built slots that allow shelves to create separate compartments within. This will solve the problem of storing fragile items with everything else.	This product involves using durable plastic to replicate a large sized cardboard box, with in-built slots that allow shelves to create separate compartments within. This will solve the problem of storing fragile items with everything else. The compartments can be designed to be different sizes. And the box itself can also come in different sizes. The storage container could even come in different colors, and you might want to decorate it to match your room.	This product involves using durable plastic to replicate a large sized cardboard box, with in-built slots that allow shelves to create separate compartments within. This will solve the problem of storing fragile items with everything else. If you have papers from school (e.g., French class, history class, math class) or legal papers or whatever, but you also have things you don't want squished by the papers, this type of storage would be a big improvement.	Christine_Lagarde
2	A device that allows you to rip a specified amount of toilet paper with one hand.	A device that allows you to rip a specified amount of toilet paper with one hand. The device would fit on any design of a toilet paper dispenser.	A device that allows you to rip a specified amount of toilet paper with one hand. A product like this would come in handy not just in regular	Glenn_Haynes

		It would be convenient to use and attractive looking in the room.	bathrooms in homes, but also in other bathrooms, like those on boats.	
3	Instead of 4 separate shakers containing garlic, oregano, red pepper, and parmesan cheese, there would be one large shaker with four separate compartments. The holes would be different sizes depending on the contents of that compartment.	Instead of 4 separate shakers containing garlic, oregano, red pepper, and parmesan cheese, there would be one large shaker with four separate compartments. The holes would be different sizes depending on the contents of that compartment. This complete pizza seasoning shaker would be a great addition to any college dorm or household kitchen, allowing each person sharing a pizza to customize his or her own slices.	Instead of 4 separate shakers containing garlic, oregano, red pepper, and parmesan cheese, there would be one large shaker with four separate compartments. The holes would be different sizes depending on the contents of that compartment. This complete pizza seasoning shaker would give each person sharing a pizza a surgeon's precision in customizing his or her own slice to have the exact blend of flavors desired.	Medical-surgical_nursing
4	This hockey puck shaped, battery powered, LED flash light will be mounted to a shoulder strap of a backpack or any type of carrying bag. This will increase the safety when walking from school to home at night.	This hockey puck shaped, battery powered, LED flash light will be mounted to a shoulder strap of a backpack or any type of carrying bag. This will increase the safety when walking from school to home at night. The light bulb in an LED flashlight should have a long life, and the whole device will not use much battery power. LEDs are energy efficient, bright, and don't weigh a lot. That makes them ideal.	This hockey puck shaped, battery powered, LED flash light will be mounted to a shoulder strap of a backpack or any type of carrying bag. This will increase the safety when walking from school to home at night. If you are not a football player, you might feel that you could use a little extra protection when walking around after dark. This light would allow you to see around you and make you feel safer.	1900_Connecticut_Aggies_football_team
5	Attach sensors to keys, ipod, telephone, PDA so that then can be easily located. Program a hand-held device to locate the sensors on the objects. The hand-held keychain determines how far the object is by speeding up the frequency of the beeps.	Attach sensors to keys, ipod, telephone, PDA so that then can be easily located. Program a hand-held device to locate the sensors on the objects. The hand-held keychain determines how far the object is by speeding up the frequency of the beeps. The sensors would be attached to the objects with stickers. The stickers would come in different levels of adhesive. If you are worried about item theft, you may want a very strong adhesive so the sensor cannot be removed.	Attach sensors to keys, ipod, telephone, PDA so that then can be easily located. Program a hand-held device to locate the sensors on the objects. The hand-held keychain determines how far the object is by speeding up the frequency of the beeps. The sensors could help prevent theft. Prevention of theft is much better than having to turn to the court system to seek help from a judge and/or jury for stolen items. Even a Chief Justice might not be able to help you out.	Chief_Justice_of_Gibraltar

In this study, 205 people completed the survey and 180 of them passed the attention filter.

Table C-3: Results from Study 5

	Need	Purchase Intent	Creativity	Uniqueness
Version A (original)	2.04	2.23	3.03	2.67
Version B (longer)	2.11	2.31	3.19	2.75
Version C (Wikipedia)	2.03	2.22	3.13	2.77
Paired t-test A vs. B p-value	p=0.14	p=0.18	p<0.01	p=0.07
Paired t-test B vs. C p-value	p=0.07	p=0.13	p=0.22	p=0.65
Paired t-test A vs. C p-value	p=0.83	p=0.92	p=0.13	p<0.05

Stimuli for Study 6a

In Study 6a, we collected ratings for the core versions of the frosting ideas. In Appendix D, we present results from a study that shows that different people given a simple instruction could shorten (or “core”) ideas with high reliability of the ratings. Some of the original frosting ideas were already in “core” form. Those we did not change, so the same idea (e.g., espresso mocha frosting) appeared in both sets. We show a few examples of the original and core ideas below.

#	Original Version of the Idea	Core Version
1	A frosting package that has seasonal or birthday stencils and two or three small packets of colored sugar and sprinkles for easy decorating.	Frosting package with seasonal stencils and colored sugar and sprinkles.
2	Lavender cardamom frosting that is pale lavender in color and in the shape of a lavender flower.	Lavender-colored and flavored cardamom frosting that's shaped like a flower.
3	Special edition flavors that are around for a limited time such as a featured destination like Paris and the flavors could be inspired by desserts from the region like macaroons, nutella crepes, chocolate croissants, etc.	Special-edition destination-themed flavors.
4	You can combine all the frosting in a flowery bouquet all shaped like the flowers they represent and sell together.	Frosting in a flowery bouquet.
5	Designer frosting. A frosting with the option to sprinkle the designer confetti on it. For example, brown frosting with the LV (Louis Vuitton) confetti pieces in a caramel flavor	Frosting with designer brand confetti.
6	Espresso mocha frosting.	Espresso mocha frosting.

Each respondent was randomly assigned to either the original or core ideas and then saw 29 randomly selected ideas. Each respondent saw 29 randomly selected short ideas. Each idea was shown with these three questions:

1. How likely would you be to purchase a product based on this idea?
(5 point scale from Definitely would not purchase to Definitely would purchase).
2. How different is this idea from products on the market today?
(4 point scale: There is nothing else like it, There are other items that are something like this, There are other items that are just like this, There are a lot of other items that are just like this)
3. This idea is creative.
(5 point scale from Strongly agree to Strongly disagree)

In this study, 386 people completed the survey and 350 of them passed the attention filter.

Stimuli for Study 6b

In Study 6b, we collected ratings for the core versions of the household products ideas. In this study, we shortened all 100 of the household product ideas and collected ratings on the shortened versions. We show a few examples of the original and core ideas below.

#	Original Version of the Idea	Core Version
1	It's a problem that a two-liter soda bottle gets warm fast, especially when outside. I can't find any 2 liter koozies, with a detachable pour handle. Also, people taking bottles in and out of the fridge, as well as the pouring action, makes soda go flat fast and spills are a pain.	A koozie with a detachable handle for a two liter bottle to keep it cold and make it easy to pour.
2	This product is two twistable/latchable suction cups that have a thin bubble level between the cups. A customer would attach 1 suction cup and attach the 2nd using the built-in level. The suction cups could have hooks built onto the handles. The customer could be: - Home owners - Apartment owners - Any worker or resident with painted walls or a smooth surface	Two suction cups with hooks and a built-in level for hanging pictures straight.
3	Easily manage messy and excessive cord lengths. The problem of excessive cord lengths can be tackled easily by making a provision under the outlet with holes in it, through which bundles of excessive cord can be protruded, just by winding the excessive cord around your hand to make a nice bundle and then stick it through the nearest hole.	A power outlet that has a place to wind and store excess cord lengths.
4	Traditional coat hangers are triangular. To put a shirt on a traditional coat hanger, you typically have to either stretch the collar or pull the coat hanger up through the body of the shirt. Stretching sucks, and pulling up through the body of the shirt is time consuming and can be even more difficult to take off. Not to mention, thin coat hangers often leave "dimples" in the shoulders or your clothes. My improved coat hanger is an open triangle. Simply insert the open end of the coat hanger up into the sleeve of the garment and out through the collar. It's even easier to remove! This even works for pants. Simply fold pants in half and slide them onto the hanger. A small ball at the end of the open hanger prevents clothes from sliding off. And, they're one size fits all, slide on anything from your baby's onesie to your ugly Christmas sweater.	A clothes hanger with an open triangle shape to protect shoulders of shirts and make it easy to slide off pants.
5	The idea is a small plastic wristband device to clamp on a child or adult wrist which will have clasps on top to hold a small packet of pop-up tissues. With the bracelet and tissues on one wrist the tissues would pop up easily for a quick grab with the other hand when a cough or sneeze is coming on. Simple to put on and off, to use, and to change packets. Could come in different sizes and colors. Perfect for schoolchildren and teachers especially those that have colds and allergies. Anyone who does not have their hands free (ex. drivers) could find this safe and easy to use. Could call it KA-CHOO. This solution is for schools, students, teachers. Anyone with allergies. Health care workers and home health aides. Infant and child care workers. Craft artists and watercolor artists. Chefs, food production workers, and waitresses. Anyone whose hands are occupied at other tasks such as using a computer. Anyone interested in a more germ-free environment. Schools could buy this item in quantity in their school colors for each child.	Plastic wristband to hold a packet of pop-up tissues to have convenient availability when sneezing.

Each idea was shown the same questions as Study 6a, for the core frosting ideas (purchase intent, difference from products on market, and creativity).

In this study, 154 people completed the survey and 133 of them passed the attention filter.

Stimuli for Study 7a Frosting Associations

In a follow-up study on the core frosting ideas, we collected data on the match or suitability of attributes mentioned for the category. Here are some examples of the questions and the attributes.

1. For each of the listed FLAVORS, rate how suitable it is for a FROSTING. (5 point scale from Highly Suitable to Highly Unsuitable). Example flavors include cinnamon, pinto bean, almond, and espresso mocha.
2. For each of the listed APPEARANCE DESCRIPTIONS, rate how suitable it for a FROSTING. Example appearance descriptions include changes color, glow-in-the-dark, tie-dye.

In this study, 80 people completed the survey and of them 54 passed the attention filter.

Instructions and Stimuli for Study 7b College Student Products Decomposition

In this study, we used a random sample of 29 ideas from the 290 ideas for products for college students. (See details in the description of Study 3.) First, we had four people decompose the idea descriptions into two elements. Next, we collected ratings in two waves.

Here are some examples of the ideas.

#	Original Version of the Idea
1	A carrying device that holds quarters for doing laundry. It has two simple setting alarm clocks for monitor washer and dryer time simultaneously.
2	A convex screen which can be placed in front of the laptop to magnify the laptop screen. Similar to a projector, but a cheaper option for students.
3	A compact traveling kit that has small plastic bottles that are clearly labeled for typical toiletries (e.g. shampoo). Would be great for traveling.
4	A LED lamp with a flat vertical surface to stick Post-it notes on. This product would use a clamp to attach itself to a monitor of any thickness.
5	A hairbrush with some sort of removable layer that makes it easy to remove all the hair caught between the bristles efficiently in one quick step.

In the decomposition task, we structured the task as follows. First, we asked “Which of the descriptions below best applies to this idea?”

- This idea is a combination of two or more independent products.
- This idea is the addition of a feature or features to a base product.
- Neither of the descriptions above applies to this idea.

Depending on the answer to the first question, we asked the matching second question:

- You said this idea is a combination of two or more independent products. What products? (If there are more than two, just give two.)
- You said this idea is the addition of a feature or features to a base product. What are the base product and a feature? (If there are more than one, just give one of each.)
- You said this idea was NEITHER a combination of two or more independent products nor the addition of a feature or features to a base product. Please decompose the product into two components. Components can be any element of the idea: a feature, function, benefit, characteristic, setting, etc. What are two of the components of this idea? (If there are more than two, just give two.)

We had four people do the decomposition, and for every idea, we used the majority decomposition. Examples of the decompositions are show below.

#	Original Version of the Idea	Element	Element
1	A carrying device that holds quarters for doing laundry. It has two simple setting alarm clocks for monitor washer and dryer time simultaneously.	Alarm Clock	Carrying Device
2	A convex screen which can be placed in front of the laptop to magnify the laptop screen. Similar to a projector, but a cheaper option for students.	Convex Screen	Magnification
3	A compact traveling kit that has small plastic bottles that are clearly labeled for typical toiletries (e.g. shampoo). Would be great for traveling.	Traveling Kit	Plastic Bottles
4	A LED lamp with a flat vertical surface to stick Post-it notes on. This product would use a clamp to attach itself to a monitor of any thickness.	LED Lamp	Flat-Surface for Post-It Notes
5	A hairbrush with some sort of removable layer that makes it easy to remove all the hair caught between the bristles efficiently in one quick step.	Hairbrush	Removable Layer

We collected ratings in two waves. In the first wave, we showed both the idea descriptions and the decompositions. We asked the following three questions for each idea:

1. How good an idea do you think this is? (5 point scale from Very bad to Very good)
2. How unique is this idea? (3 point scale from Not at all unique to Very unique)

3. This idea is made up of two components: Alarm Clock and Carrying Device. How similar are the two parts (Alarm Clock and Carrying Device) to each other? (5 point scale from Very similar to Very different)

In this wave, 61 people completed the survey and of them 54 passed the attention filter.

In the second wave, we did not show the idea descriptions, just the decomposed elements. We asked the following two questions for each idea:

1. Using the first context that comes to mind, how similar are Alarm Clock and Carrying Device to each other? (5 point scale from Very similar to Very different)
2. How strong is the association between Alarm Clock and Carrying Device? (5 point scale from Very weak to Very strong)

In this wave, 57 people completed the survey and of them 40 passed the attention filter.

APPENDIX D: STUDY SHOWING RELIABILITY OF CORE IDEA EXTRACTION

Study 3 showed that a longer version of an idea that adds no additional information or contains concrete details is rated as more creative than a shorter version of the idea. This difference suggests that we should somehow account for the length of the idea when trying to learn which raw ideas are more creative than others. One way to do that is to work with the core idea. We recognize that such extraction requires subjective judgment. Therefore, we conducted a study (Study D-1) to show that these core ideas can be created by untrained assistants in a reliable way. We share the details of that study here.

Ideas. We randomly selected 30 ideas from the set of 100 household product ideas. We tasked four research assistants blind to the purpose of the study to extract core ideas. We instructed these four people, “in this task you will be asked to read the descriptions carefully and rewrite the idea, using the core components of the lengthy version, in 100 characters or less.” (The original 30 ideas in the sample averaged 713.0 characters, and ranged from 153 to 1580 characters.)

With the four versions of the core ideas, we collected creativity, uniqueness, and purchase intent ratings from Mechanical Turk workers. Each worker rated one block of 30 ideas, including exactly one of the four shortened versions for each idea, with each version randomly assigned to a block. We received 45-48 responses that passed the attention filter in each block.

Results. The ratings for the different versions of the same idea showed reasonable levels of agreement. Details of the agreement for each of the three measures are below.

Table D-1 shows the correlations between the average purchase intent ratings for the 30 ideas, across the versions created by four writers. The average correlation is 0.63, and Cronbach’s alpha is 0.87. All of the pairwise correlations are statistically significant (N=30,

p<0.01). Table D-2 shows the correlations for the creativity measures. The average correlation is 0.53, and Cronbach's alpha is 0.82. All of the pairwise correlations are statistically significant (N=30, p<0.05). Table D-3 shows the correlations for the uniqueness measures. The average correlation is 0.59, and Cronbach's alpha is 0.85. All of the pairwise correlations are statistically significant (N=30, p<0.01).

Table D-1: Purchase Intent					Table D-2: Creativity					Table D-3: Uniqueness				
	W 1	W 2	W 3	W 4		W 1	W 2	W 3	W 4		W 1	W 2	W 3	W 4
Writer 1	1.00	0.55	0.66	0.60	Writer 1	1.00	0.54	0.55	0.55	Writer 1	1.00	0.52	0.56	0.50
Writer 2	0.55	1.00	0.75	0.61	Writer 2	0.54	1.00	0.62	0.40	Writer 2	0.52	1.00	0.77	0.54
Writer 3	0.66	0.75	1.00	0.59	Writer 3	0.55	0.62	1.00	0.54	Writer 3	0.56	0.77	1.00	0.63
Writer 4	0.60	0.61	0.59	1.00	Writer 4	0.55	0.40	0.54	1.00	Writer 4	0.50	0.54	0.63	1.00

Discussion. In addition to the reasonable consistency in the ratings of the versions as shown above, we also found strong correspondence between the ratings of the shortened ideas and the original ideas. The average of the ratings of the four shortened versions is correlated 0.74 with the original purchase intent for the 30 ideas in this sample. The average of creativity is 0.48 with the long creativity. We found these levels of correspondence, especially for purchase intent, to be remarkable, given the differences in what was measured. The differences include the following: the original ideas were significantly longer than the shortened versions, some of the original ideas included images to illustrate the idea, and the original ratings occurred several years before the ratings of the shortened ideas and were from a different vendor who used a different compensation scheme for the raters.

Even though paring a raw idea down to its core involves subjective judgment, we found that untrained assistants could follow a simple instruction to do that in a way that yields reliable ratings of creativity, uniqueness, and purchase intent.

APPENDIX E: TALLY OF PAPERS WITH IDEA EVALUATION

Table E-1: All Papers in Keyword Search

Results from five journals 1999-2019 with results in Web of Science for these search terms: “new products,” “idea generation,” or “ideation.”

Title	Authors	Year	Journal	Creativity?
When Two Wrongs Make a Right: Using Conjunctive Enablers to Enhance Evaluations for Extremely Incongruent New Products	Noseworthy TJ, Murray KB, Di Muro F	2018	<i>J. Consumer Res.</i>	0
Design Crowdsourcing: The Impact on New Product Performance of Sourcing Design Solutions from the "Crowd"	Allen BJ, Chandrasekaran D, Basuroy S	2018	<i>J. Marketing</i>	1
The Impact of Idea Generation and Potential Appropriation on Entrepreneurship: An Experimental Study	Hooshangi S, Loewenstein G	2018	<i>Management Science</i>	0
Ideation-Execution Transition in Product Development: An Experimental Analysis	Kagan E, Leider S, Lovejoy WS	2018	<i>Management Science</i>	0
The Value of Marketing Crowdsourced New Products as Such: Evidence from Two Randomized Field Experiments	Nisikawa H, Schreier M, Fuchs C, Ogawa S	2017	<i>J. Marketing Res.</i>	0
Idea Generation, Creativity, and Prototypicality	Toubia O, Netzer O	2017	<i>Marketing Science</i>	1
Designed to Succeed: Dimensions of Product Design and Their Impact on Market Share	Jindal RP, Sarangee KR, Echambadi R, Lee S	2016	<i>J. Marketing</i>	0
The Effectiveness of Customer Participation in New Product Development: A Meta-Analysis	Chang W, Taylor SA	2016	<i>J. Marketing</i>	0

Lower Connectivity Is Better: The Effects of Network Structure on Redundancy of Ideas and Customer Innovativeness in Interdependent Ideation Tasks	Stephen AT, Zubcsek PP, Goldenberg J	2016	<i>J. Marketing Res.</i>	1
The Downstream Consequences of Problem-Solving Mindsets: How Playing with LEGO Influences Creativity	Moreau CP, Engeset MG	2016	<i>J. Marketing Res.</i>	1
Improving Online Idea Generation Platforms and Customizing the Task Structure on the Basis of Consumers' Domain-Specific Knowledge	Luo L, Toubia O	2015	<i>J. Marketing</i>	1
How Kinetic Property Shapes Novelty Perceptions	Kim J, Lakshmanan A	2015	<i>J. Marketing</i>	1
Core Versus Peripheral Innovations: The Effect of Innovation Locus on Consumer Adoption of New Products	Ma Z, Gill T, Jiang Y	2015	<i>J. Marketing Res.</i>	0
Remanufacturing, Third-Party Competition, and Consumers' Perceived Value of New Products	Agrawal VV, Atasu A, van Ittersum K	2015	<i>Management Science</i>	0
The Role of Arousal in Congruity-Based Product Evaluation	Noseworthy TJ, Di Muro F, Murray KB	2014	<i>J. Consumer Res.</i>	0
Optimal Visualization Aids and Temporal Framing for New Products	Zhao M, Dahl DW, Hoeffler S	2014	<i>J. Consumer Res.</i>	0
Consumer Adoption of New Products: Independent Versus Interdependent Self-Perspectives	Ma Z, Yang Z, Mourali M	2014	<i>J. Marketing</i>	1
Crowdsourcing New Product Ideas Under Consumer Learning	Huang Y, Vir Singh P, Srinivasan K	2014	<i>Management Science</i>	0

Integrating Problem Solvers from Analogous Markets in New Product Ideation	Franke N, Poetz M, Schreier M	2014	<i>Management Science</i>	1
The Importance of the Raw Idea in Innovation: Testing the Sow's Ear Hypothesis	Kornish LJ, Ulrich KT	2014	<i>J. Marketing Res.</i>	0
Crowdsourcing New Product Ideas over Time: An Analysis of the Dell IdeaStorm Community	Bayus BL	2013	<i>Management Science</i>	0
Get It? Got It. Good! Enhancing New Product Acceptance by Facilitating Resolution of Extreme Incongruity	Jhang JH, Grant SJ, Campbell MC	2012	<i>J. Marketing Res.</i>	1
Facilitating and Rewarding Creativity During New Product Development	Burroughs JE, Dahl DW, Moreau CP, Chattopadhyay A, Gorn GJ	2011	<i>J. Marketing</i>	1
Focus! Creative Success Is Enjoyed Through Restricted Choice	Sellier AL, Dahl DW	2011	<i>J. Marketing Res.</i>	1
The Circulation of Ideas in Firms and Markets	Hellman T, Perotti E	2011	<i>Management Science</i>	1
Opportunity Spaces in Innovation: Empirical Analysis of Large Samples of Ideas	Kornish LJ, Ulrich KT	2011	<i>Management Science</i>	0
The "Right" Consumers for Better Concepts: Identifying Consumers High in Emergent Nature to Develop New Product Concepts	Hoffman DL, Kopalle PK, Novak TP	2010	<i>J. Marketing Res.</i>	0
The Effect of Need for Uniqueness on Word of Mouth	Cheema A, Kaikati AM	2010	<i>J. Marketing Res.</i>	1
Idea Generation and the Quality of the Best Idea	Girotra, Terwiesch, & Ulrich	2010	<i>Management Science</i>	0

Enhancing the Motivational Affordance of Information Systems: The Effects of Real-Time Performance Feedback and Goal Setting in Group Collaboration Environments	Jung, Schneider, and Valacich	2010	<i>Management Science</i>	0
The Effects of Problem Structure and Team Diversity on Brainstorming Effectiveness	Kavadias S, Sommer SC	2009	<i>Management Science</i>	0
Thinking Inside the Box: Why Consumers Enjoy Constrained Creative Experiences	Dahl DW, Moreau CP	2007	<i>J. Marketing Res.</i>	1
Adaptive Idea Screening Using Consumers	Toubia O, Flores L	2007	<i>Marketing Science</i>	0
Idea Generation, Creativity, and Incentives	Toubia O	2006	<i>Marketing Science</i>	1
Designing the Solution: The Impact of Constraints on Consumers' Creativity	Moreau CP, Dahl DW	2005	<i>J. Consumer Res.</i>	1
Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures	Elberse A, Eliashberg J	2003	<i>Marketing Science</i>	0
The Influence and Value of Analogical Thinking During New Product Ideation	Dahl DW, Moreau CP	2002	<i>J. Marketing Res.</i>	1
Performance Assessment of the Lead User Idea-Generation Process for New Product Development	Lilien GL, Morrison, PD Searls K, Hippel EV	2002	<i>Management Science</i>	0
The Idea Itself and the Circumstances of its Emergence as Predictors of New Product Success	Goldenberg J, Lehman DR, Mazursky D	2001	<i>Management Science</i>	0
The Fundamental Templates of Quality Ads	Goldenberg J; Mazursky D; Solomon S	1999	<i>Marketing Science</i>	1
