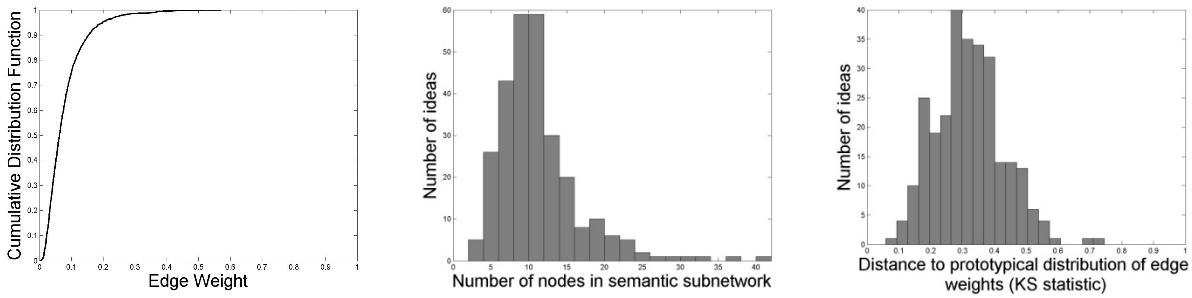
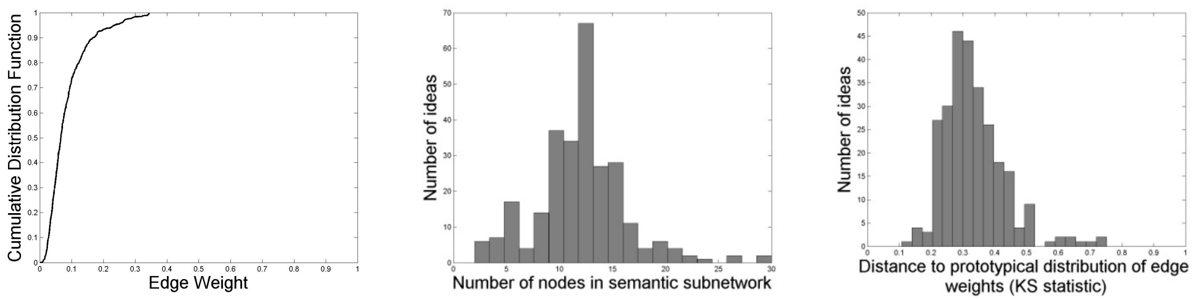
Web Appendices for “Idea Generation, Creativity, and Prototypicality”

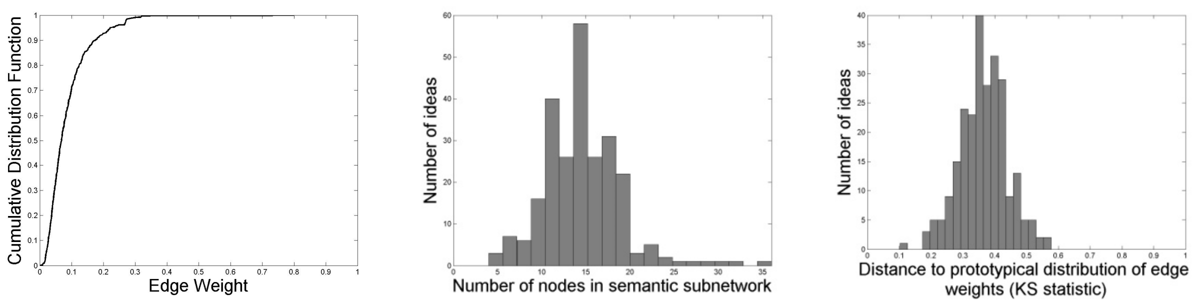
Web Appendix A: Descriptive Statistics for All Studies



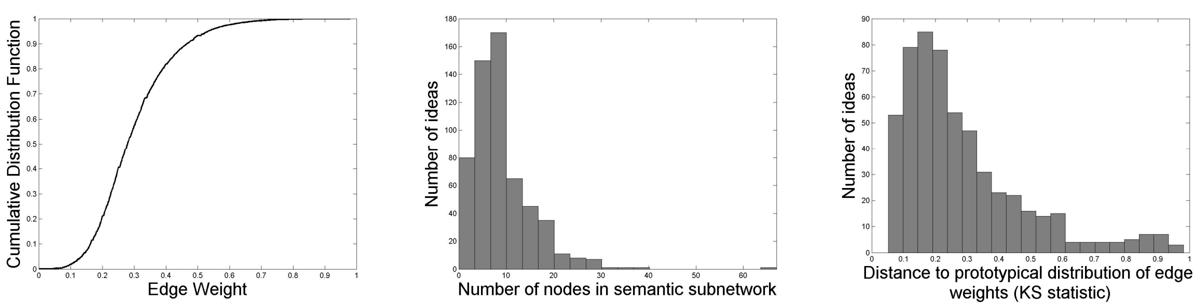
Study 1a



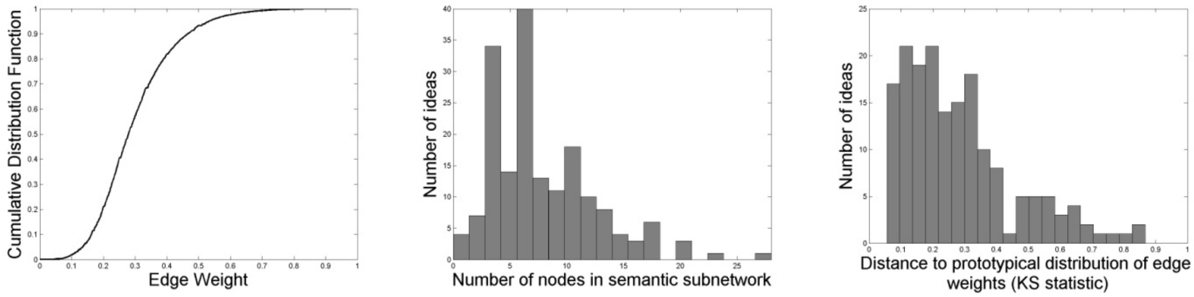
Study 1b



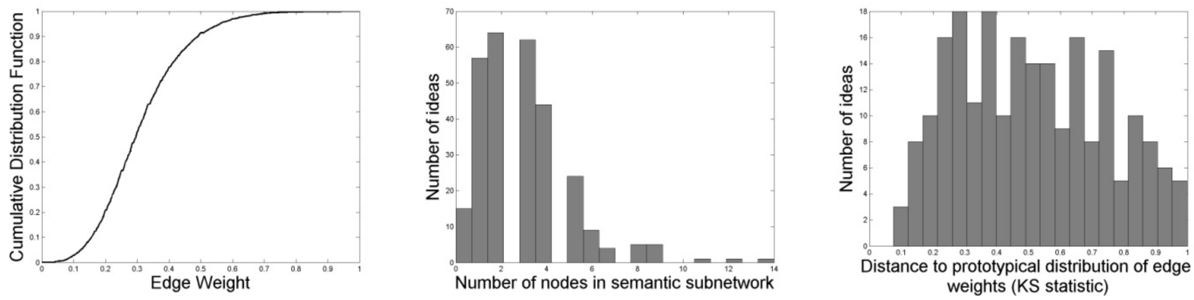
Study 1c



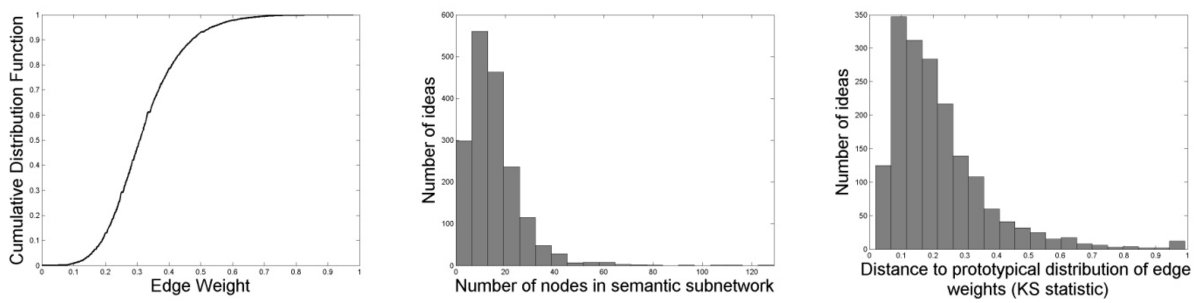
Study 2



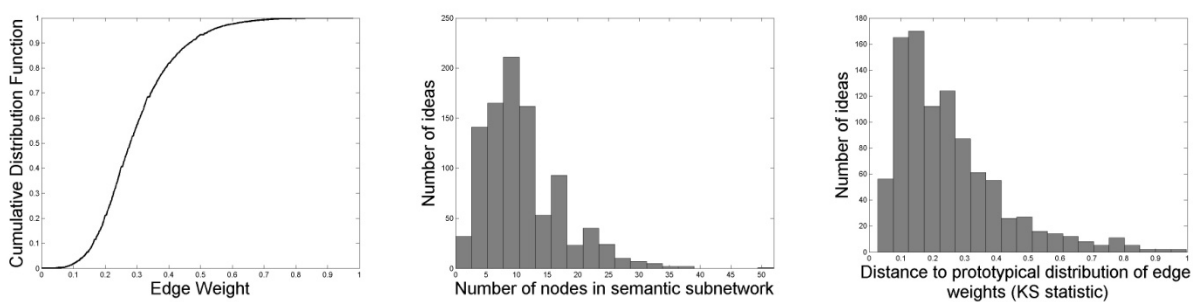
Study 3



Study 4



Study 5



Study 6

Left: prototypical distribution of edge weights. Center: distribution across ideas of the number of nodes in the semantic subnetwork. Right: distribution of prototypicality measure (Kolmogorov-Smirnov statistic) across ideas.

Web Appendix B: Receiver operating characteristic (ROC) curve for Study 4

We distinguish between precision and recall by creating a ROC curve as follows. We first rank ideas based on their distance to the prototypical edge weight distribution. Next, we compute the true positive rate and false positive rate that would result from predicting that the top *k* ideas would be selected by the company, where we vary *k* from 1 to the number of ideas. See Figure B1. We see that a classifier based on distance to the prototypical edge weight distribution performs adequately.

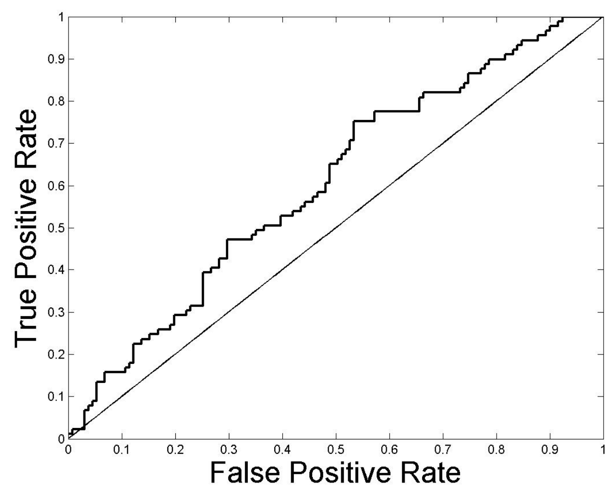


Figure B1. ROC curve for company evaluations - Study 4.

Web Appendix C: Robustness Checks

Fixed Effects for Word Stems

One limitation of our main regression specification is that it ignores the direct effect of the word stems present in an idea on its judged creativity. We test the robustness of our results when word stem fixed effects are introduced. In one series of regression, we introduce 20 fixed effects in each study corresponding to the 20 word stems (i.e., nodes in the baseline semantic network) that appear in the greatest number of ideas in that study. In another series of regressions, we increase the number of word stem fixed effects to 50. All other variables including the dependent variable and the covariates are similar to Table 2 in the manuscript. In both sets of regressions we still capture heterogeneity across participants with random effects. Results are reported in Tables C1 and C2. With 20 fixed effects, the effect of prototypicality remains statistically significant at $p<0.05$ in 4 out of the 8 studies, is marginally significant at $p<0.10$ in 3, and directionally consistent with our hypothesis in one. With 50 fixed effects, despite the great reduction in statistical power, the effect remains statistically significant at $p<0.05$ in 4 studies, is marginally significant at $p<0.10$ in 2 studies, and is of the expected sign in the remaining two studies. Under the null hypothesis that the effect is positive with probability 0.5 and negative with probability 0.5, this pattern of results would be observed with probability $0.5^8=0.004$.

		Study 1a	Study 1b	Study 1c	Study 2	Study 3	Study 4	Study 5	Study 6
	Distance to prototypical edge weight distribution	-3.279**	-1.125	-2.705**	-0.395**	-0.643*	-0.474*	-0.764**	-0.325*
Edge Weight	Average	-4.670	-26.227**	-5.836	0.337	-1.628	-0.812	-0.658	-0.289
	Coef. of var.	0.496	-0.044	0.332	0.000	-0.285	1.046*	-0.684**	0.158
	Min	-8.648	0.466	-5.328	-0.163	-0.205	2.091*	2.050**	0.237
	Max	-0.357	1.655	0.348	0.158	0.920	-0.856	1.184**	0.272
Node frequency	Average	-10.761*	-3.877	13.668	-1.595**	-0.761	-0.352	-3.916**	-1.437**
	Coef. of var.	1.652**	-0.459	2.489**	-0.393	-0.778	-0.066	-1.094**	-0.691
	Min	5.213	13.136**	15.493	0.344	0.688	0.313	-0.512	-0.135
	Max	-0.848	1.666	-11.975	0.581	0.997	-0.479	0.709**	0.800*
Node clustering coef.	Average	-0.729	0.336	-0.350	N/A	N/A	N/A	N/A	N/A
	Coef. of var.	-0.184	-2.949	-4.845	N/A	N/A	N/A	N/A	N/A
	Min	0.462	-1.806	-3.236**	N/A	N/A	N/A	N/A	N/A
	Max	-2.236	4.437	5.725	N/A	N/A	N/A	N/A	N/A
	Size of semantic subnetwork	-0.071**	-0.065**	-0.004	-0.001	0.004	0.027	-0.013**	-0.022**
	Number of characters/1000	1.496**	1.403**	1.126**	-0.032	0.623	0.677	-0.239**	1.279**
	# observations	276	271	251	555	173	220	1735	648
	# groups (authors)	178	177	167	300	61	163	703	391
	R^2 / χ^2	0.361	0.463	0.332	0.230	0.429	0.246	850.77	0.340

*: $p<0.1$, **: $p<0.05$.

Table C1. Including fixed effects for the 20 most frequent word stems in each study.

Note: We include fixed effects for the 20 most common word stems in each study. All other aspects of the regressions are similar to Table 2.

		Study 1a	Study 1b	Study 1c	Study 2	Study 3	Study 4	Study 5	Study 6
	Distance to prototypical edge weight distribution	-2.677**	-0.899	-3.163**	-0.353**	-0.585	-0.568*	-0.787**	-0.331*
Edge Weight	Average	-6.305	-26.583**	-3.136	0.440	-2.020	-0.701	-0.288	-0.704
	Coef. of var.	0.615	-0.160	2.014**	0.088	-0.027	0.979	-0.811**	0.248
	Min	-8.105	0.963	-6.283	-0.148	0.371	1.539	1.682**	0.493
	Max	0.603	3.145	-2.197	0.056	1.068	-0.832	1.320**	0.381
Node frequency	Average	-9.928	-1.228	15.593	-1.456**	-0.849	0.336	-3.465**	-1.405*
	Coef. of var.	1.210	0.384	3.082**	-0.255	-0.945	0.625	-0.731**	-0.695
	Min	7.044	12.285**	20.766*	0.225	0.181	1.418	-0.481	-0.157
	Max	-0.988	1.866	-18.779	0.487	1.238	-1.238	0.206	0.924*
Node clustering coef	Average	-0.536	-2.283	0.082	N/A	N/A	N/A	N/A	N/A
	Coef. of var.	0.218	-3.398	-8.198**	N/A	N/A	N/A	N/A	N/A
	Min	0.534	-1.238	-2.943**	N/A	N/A	N/A	N/A	N/A
	Max	-4.578*	1.193	8.204**	N/A	N/A	N/A	N/A	N/A
	Size of semantic subnetwork	-0.094**	-0.112**	-0.018	-0.003	0.026	0.007	-0.018**	-0.018*
	Number of characters/1000	1.662**	1.785**	1.078**	-0.052	0.830	0.976	-0.107	1.319**
	# observations	276	271	251	555	173	220	1735	648
	# groups (authors)	178	177	167	300	61	163	703	391
	R^2 / χ^2	0.475	0.529	0.351	0.290	0.592	0.361	1418.32	0.380

*: p<0.1, **: p<0.05.

Table C2. Including fixed effects for the 50 most frequent word stems in each study.

Note: We include fixed effects for the 50 most common word stems in each study. All other aspects of the regressions are similar to Table 2.

Using the Ideas Themselves to Create the Prototypical Distribution

We test the robustness of the results in Studies 2-6 when using the ideas submitted by participants to create the prototypical edge weight distribution, instead of using pages retrieved from Google. Table C3 reports the results. This analysis is the same as in Table 2, with the only difference that the prototypical edge weight distribution is defined as the average distribution among the ideas submitted in each study, instead of the pages retrieved from Google. (The baseline semantic network itself is still constructed using the Google results). We see that the coefficient corresponding to the prototypicality of the edge weight distribution is statistically significant in 4 of the 5 studies, and directionally consistent with our hypothesis in the fifth.

		Study 2	Study 3	Study 4	Study 5	Study 6
	Distance to prototypical edge weight distribution	-0.375**	-0.748**	-0.438**	-0.572**	-0.242
Edge Weight	Average	0.525	-0.769	-1.592	0.817	-0.840
	Coef. of var.	-0.100	-0.128	0.733	-0.569**	0.410
	Min	-0.457	-0.047	2.172*	2.030**	0.363
	Max	0.259	0.181	-0.359	0.720**	0.167
Node frequency	Average	-0.888*	0.600	-0.945	-4.072**	-0.788
	Coef. of var.	0.215	0.555	0.035	-0.879**	-0.199
	Min	0.462	0.292	0.801	-1.422**	0.261
	Max	0.141	-0.398	0.196	1.152**	0.514
	Size of semantic subnetwork	0.007	0.001	0.040	-0.002	-0.021**
	Number of characters/1000	-0.029	1.028**	0.922	-0.235**	1.299**
	# observations	555	173	251	1735	648
	# groups (authors)	300	61	167	703	391
	R^2 / χ^2	0.191	0.273	0.255	304.99	0.263

*: p<0.1, **: p<0.05.

Table C3. Prototypical distribution based on ideas themselves instead of Google results.

Kullback-Leibler Divergence Instead of Kolmogorov-Smirnov Statistic

Our analysis in the paper is based on the Kolmogorov-Smirnov statistic between the prototypical edge weight distribution and each idea's edge weight distribution. Another common measure of distance in the language processing community is the Kullback-Leibler divergence. However, the Kullback-Leibler divergence of a distribution Q from another distribution P is only defined if the support of P is a subset of the support of Q. In our main analysis the support of an idea's edge weight distribution may include 0, i.e., there may be combinations of word stems in the idea that never appear together in the training documents that were used to create the baseline semantic network. However, when the same set of documents that were used to create the baseline semantic network is also used to construct the prototypical edge weight distribution, by construction the support of the prototypical edge weight distribution does not include 0. Indeed, every combination of words that appear together in any pretest idea / Google result appears together in at least one of the training documents, since it appears at least in that same document. Therefore the Kullback-Leibler divergence may not be used throughout the paper.

However, this property does not hold when we use the ideas themselves to create the prototypical edge weight distribution in Studies 2-6, instead of the Google results. In that case, the support of each idea's edge weight distribution becomes a subset of the support of the prototypical distribution. Therefore, it is possible to use the Kullback-Leibler divergence of the prototypical edge weight distribution from each idea's edge weight distribution, instead of the Kolmogorov-Smirnov statistic. Results are reported in Table C4. We see that our hypothesis still holds when using this alternative measure of distance between distributions.

		Study 2	Study 3	Study 4	Study 5	Study 6
	Distance to prototypical edge weight distribution	-0.109**	-0.225**	-0.005	-0.082**	-0.117**
Edge Weight	Average	0.238	-1.886	-0.676	0.885	-1.143
	Coef. of var.	0.045	0.251	0.729	-0.494*	0.559*
	Min	-0.319	0.498	1.410	2.020**	0.628
	Max	0.104	0.100	-0.500	0.605**	0.008
Node frequency	Average	-0.391	1.251	-0.715	-3.538**	-0.320
	Coef. of var.	0.088	-0.042	0.146	-0.628**	-0.285
	Min	0.262	-0.431	0.569	-1.630**	0.207
	Max	0.071	-0.242	0.090	0.988**	0.398
	Size of semantic subnetwork	0.001	-0.018	0.053	-0.004	-0.029**
	Number of characters/1000	-0.034	0.999**	0.826	-0.194**	1.317**
	# observations	555	173	251	1735	648
	# groups (authors)	300	61	167	703	391
	R^2 / χ^2	0.197	0.274	0.053	296.45	0.271

*: p<0.1, **: p<0.05.

Table C4. Prototypical distribution based on ideas themselves instead of Google results. Using the Kullback-Leibler divergence instead of the Kolmogorov-Smirnov statistic.

Asymmetric Measure of Prototypicality

We compute a signed Kolmogorov-Smirnov statistic that captures whether the edge weight distribution corresponding to each idea is above (Kolmogorov-Smirnov>0) or below (Kolmogorov-Smirnov<0) the prototypical distribution at the point at which the two distributions are maximally distant. Let x^* be the point at which the absolute distance between an idea's cumulative distribution and the prototypical distribution is the largest. Instead of measuring the distance between the two distributions as the absolute value of the difference at x^*, we compute the signed difference between the idea's cumulative distribution and the prototypical distribution at x^*. We separate this distance between its positive part (max(distance,0)) and its negative part (max(0,-distance)). Ideas with non-zero positive parts (negative parts) are such that the idea's cumulative distribution is higher (lower) than the prototypical distribution at x^*, i.e. it puts more weight on smaller (larger) edge weights. We include both the positive and negative parts of the signed Kolmogorov-Smirnov statistic as two separate covariates. All other aspects are similar to the main analysis reported in Table 2. Results are available in Table C5. We find the same effect of prototypicality on judged creativity for ideas whose distributions are above the prototypical distribution (i.e., more small weights and therefore more novel combinations) as well as for ideas whose distributions are below the prototypical distribution (i.e., more large weights and therefore more familiar combinations), although the effect is stronger for the former type of ideas.

		Study 1a	Study 1b	Study 1c	Study 2	Study 3	Study 4	Study 5	Study 6
	Distance to prototypical edge weight distribution – positive part	-4.053**	-2.444**	-3.762**	-0.382**	-1.089**	-0.425*	-0.498**	-0.294
	Distance to prototypical edge weight distribution – negative part	-1.114	-2.280**	-1.003	-0.376*	-0.554	-0.276	-0.259	-0.582**
Edge Weight	Average	-13.913*	-13.167*	-13.209	0.620	-1.608	-1.477	0.929	-0.001
	Coef. of var.	0.937**	1.061*	0.832	-0.129	-0.103	0.647	-0.621**	0.270
	Min	-3.183	0.332	-3.458	-0.500	-0.155	1.804	1.837**	0.384
	Max	-0.966	-1.299	-0.545	0.268	0.104	-0.300	0.740**	0.137
Node frequency	Average	-10.888**	6.468	11.915**	-0.802	1.087	-0.985	-4.106**	-0.678
	Coef. of var.	1.441**	1.033	1.580*	0.175	0.618	-0.089	-0.964**	-0.224
	Min	5.233	10.550**	10.415	0.416	0.197	0.652	-1.858**	0.203
	Max	-0.254	2.580	-4.653*	0.136	-0.516	0.291	1.215**	0.491
Node clustering coef.	Average	1.825	-1.306	0.556	N/A	N/A	N/A	N/A	N/A
	Coef. of var.	0.664	-4.420**	-3.312	N/A	N/A	N/A	N/A	N/A
	Min	-0.078	-2.467**	-2.724**	N/A	N/A	N/A	N/A	N/A
	Max	-4.704**	5.410	3.640	N/A	N/A	N/A	N/A	N/A
	Size of semantic subnetwork	-0.038**	-0.013	0.013	0.007	0.000	0.033	-0.002	-0.022**
	Number of characters/1000	1.313**	1.161**	0.905**	-0.032	1.042**	0.940	-0.233**	1.299**
	# observations	276	271	251	555	173	220	1735	648
	# groups (authors)	178	177	167	300	61	163	703	391
	R^2 / χ^2	0.283	0.372	0.258	0.192	0.290	0.073	286.29	0.270

*: $p<0.1$, **: $p<0.05$.

Table C5. Distance split into positive and negative parts.

Alternative Measure of Edge Weight

While the Jaccard index is a common measure of scaled co-occurrence, it is biased against asymmetric data (Mika 2007). An alternative measure that alleviates this concern is the Salton Cosine (Salton and McGill 1983). The Salton cosine between two nodes *A* and *B* is:

$$S_{A,B} = \frac{|S_A \cap S_B|}{\sqrt{|S_A| \times |S_B|}}$$

Where S_A (S_B) is the set of documents in the training corpus that contain word stem *A* (*B*). Table C6 reports the results of an analysis similar to that reported in Table 2, with the only difference that edge weights are measured using the Salton cosine rather than the Jaccard index. We see that the same conclusions are reached using this alternative measure.

		Study 1a	Study 1b	Study 1c	Study 2	Study 3	Study 4	Study 5	Study 6
	Distance to prototypical edge weight distribution	-4.203**	-1.781**	-1.750	-0.258*	-0.811**	-0.515**	-0.252*	-0.379**
Edge Weight	Average	-8.781**	-9.733**	-3.035	0.345	-1.377	-1.443	0.434	-1.373
	Coef. of var.	0.884*	0.498	-0.385	-0.021	-0.921	1.245*	-1.756**	0.258
	Min	-2.538	0.050	-3.532	-0.275	-0.247	2.343**	1.118**	0.377
	Max	-0.385	-0.132	-0.033	0.350	0.668	-0.627	1.527**	0.524
Node frequency	Average	-13.893**	5.828	9.729*	-0.748	1.016	-1.055	-4.459**	-0.119
	Coef. of var.	1.300**	1.087	2.187**	-0.123	0.338	-0.031	-1.225**	-0.271
	Min	10.367	7.659**	13.217	0.110	0.129	0.745	-1.445**	0.076
	Max	-0.174	2.849	-5.247**	0.288	-0.290	0.361	1.245**	0.392
Node clustering coef.	Average	1.572	-2.819	0.330	N/A	N/A	N/A	N/A	N/A
	Coef. of var.	0.978	-1.883	-4.566	N/A	N/A	N/A	N/A	N/A
	Min	0.547	-0.145	-2.797	N/A	N/A	N/A	N/A	N/A
	Max	-4.795**	2.391	2.755	N/A	N/A	N/A	N/A	N/A
	Size of semantic subnetwork	-0.040**	-0.020	0.012	0.007*	-0.006	0.061	-0.002	-0.023**
	Number of characters/1000	1.260**	1.183**	1.007**	-0.038	0.967**	0.969	-0.238**	1.291**
	# observations	276	271	251	555	173	220	1735	648
	# groups (authors)	178	177	167	300	61	163	703	391
	R^2 / χ^2	0.288	0.353	0.206	0.186	0.284	0.102	346.64	0.271

*: p<0.1, **: p<0.05.

Table C6. Edge weights measured with Salton cosine instead of Jaccard index.

Alternative Number of Google results

Our main analysis in Studies 2-6 is based on the page source code of the top 50 search results from Google. We test whether a significant link between prototypicality and judged creativity may be found with an even smaller number of pages. We reduce the number of pages from 50 to 10 by increments of 10. In each case, when constructing the baseline semantic network we keep word stems that appear in at least 20% of the pages. We run one set of regressions for each possible number of pages, as the ones reported in Table 2. For ease of exposition, we only report the coefficient for the distance to the prototypical edge weight distribution. Results are shown in Table C7. We see that in all cases, the results are directionally consistent with our hypothesis. In some studies, the effect is significant with as few as 20 pages. In others, 50 pages are needed to obtain a significant effect. Our recommendation for researcher would be to extract at least 50 pages in order to obtain robust and significant effects.

	Study 2	Study 3	Study 4	Study 5	Study 6
top 50 results (Table 2)	-0.380**	-0.962**	-0.411*	-0.402**	-0.401**
top 40 results	-0.203	-1.033**	-0.595**	-0.410**	-0.272
top 30 results	-0.158	-0.545*	-0.376	-0.481**	-0.667**
top 20 results	-0.080	-0.644**	-0.310	0.123	-0.411**
top 10 results	-0.308*	0.020	0.120	-0.205	-0.295

*: p<0.1, **: p<0.05.

Table C7. Coefficient on distance to prototypical edge weight distribution – number of Google results varied between 10 and 50.

Alternative Specification

One possible concern with our main regression specification is the multicollinearity between our control variables, in particular the average, the minimum and the maximum of the edge weight, node frequency, and node clustering coefficient distributions. In particular, the average is likely to be highly correlated with the sum of the minimum and the maximum. We repeat our analysis from Table 2 without controlling for averages. Results are provided in Table C8. Our conclusions are largely unchanged.

		Study 1a	Study 1b	Study 1c	Study 2	Study 3	Study 4	Study 5	Study 6
	Distance to prototypical edge weight distribution	-2.391**	-2.487**	-3.116**	-0.422**	-0.869**	-0.322	-0.633**	-0.537**
Edge Weight	Coef. of var.	0.614	1.754**	1.188*	-0.215	-0.092	0.526	-1.123**	0.212
	Min	-5.865	2.205	14.424	-0.446	0.131	1.023	1.580**	0.065
	Max	-1.071*	-3.543**	-1.036	0.352*	0.125	-0.617	0.732**	-0.027
Node frequency	Coef. of var.	2.419**	0.517	0.488	0.239	0.403	0.227	-0.424	0.351
	Min	11.048	8.847**	8.454	0.284	0.267	0.432	-3.022**	0.325
	Max	-3.160**	4.666	-0.917	-0.165	-0.149	-0.350	-0.190	-0.037
Node clustering coef.	Coef. of var.	-0.350	-4.883**	-3.916	N/A	N/A	N/A	N/A	N/A
	Min	-0.069	-2.987**	-2.333**	N/A	N/A	N/A	N/A	N/A
	Max	-3.136	5.365	3.833	N/A	N/A	N/A	N/A	N/A
	Size of semantic subnetwork	-0.022	-0.002	-0.004	0.008*	-0.002	0.043	0.006	-0.019**
	Number of characters/1000	1.261**	1.118**	1.013**	-0.013	1.033**	0.913	-0.317**	1.288**
	# observations	276	271	251	555	173	220	1735	648
	# groups (authors)	178	177	167	300	61	163	703	391
	R^2 / χ^2	0.244	0.355	0.227	0.189	0.284	0.068	209.49	0.263

*: *p*<0.1, **: *p*<0.05.

Table C8. Removing averages from controls.

Synonyms

Another potential concern is that the existence of synonyms may impact our findings. In particular, the edge weight between stems may be affected by the particular choice of synonyms used to express a particular idea. For example, while synonyms are very much related in terms of language it is unlikely that synonyms would be mentioned together in a document, as they are often substitutes. To address this issue, we downloaded the WordNet corpus (Princeton University 2010) through the Natural Language Tool Kit (NLTK) module for Python. For each word stem in our data, we created a set of synonyms sets ("synsets") defined as the set of synonym sets to which at least one word in the stem belongs. For example, the stem "abil" is associated with the words “ability” and “abilities,” which together belong to two synonym sets. For each pair of word stems A and B, we compute a Jaccard index equal to the size of the intersection of their synsets divided by the union. We classify two word stems as synonyms if their Jaccard index is more than 2/3, which we found to be the smallest value that leads to a transitive synonym relationship, i.e., that allows us to define equivalence classes of synonyms. We find that the number of synonym word stems is modest: one pair in the baseline network of Studies 1a, 1b and 4, two pairs in Studies 2, 3 and 6, and three pairs in Study 5. Nevertheless, we merge each group of synonyms and repeat our analysis. See Table C9. Our conclusions are largely unchanged.

		Study 1a	Study 1b	Study 1c	Study 2	Study 3	Study 4	Study 5	Study 6
	Distance to prototypical edge weight distribution	-3.312**	-2.378**	-3.116**	-0.365**	-0.905**	-0.325	-0.355**	-0.250
Edge Weight	Average	-8.230	-12.452*	-7.645	0.557	-0.725	-0.750	2.110**	-0.804
	Coef. of var.	0.691	1.062*	0.560	-0.009	-0.270	0.610	-0.616**	0.514
	Min	-4.901	0.747	6.361	-0.297	0.000	1.619	1.796**	0.555
	Max	-0.816	-1.438	-0.351	0.229	0.296	-0.373	0.739**	0.227
Node frequency	Average	-10.761**	6.274	11.729**	-1.098**	0.646	-1.018	-4.468**	-1.257**
	Coef. of var.	1.538**	0.971	1.945**	-0.047	0.350	0.011	-0.821**	-0.715
	Min	14.042	10.260**	12.401	0.365	0.149	0.685	-1.692**	0.060
	Max	-0.413	2.808	-5.132**	0.409	-0.208	0.300	1.122**	1.037**
Node clustering coef.	Average	1.544	-1.440	0.539	N/A	N/A	N/A	N/A	N/A
	Coef. of var.	0.310	-4.592**	-3.440	N/A	N/A	N/A	N/A	N/A
	Min	-0.217	-2.537**	-2.716**	N/A	N/A	N/A	N/A	N/A
	Max	-4.312*	5.693	3.964	N/A	N/A	N/A	N/A	N/A
	Size of semantic subnetwork	-0.035**	-0.014	0.014	-0.036	-0.005	0.026	-0.001	-0.027**
	Number of characters/1000	1.324**	1.176**	0.914**	3.698	1.023**	1.123	-0.232**	1.354**
	# observations	276	271	251	555	173	220	1735	648
	# groups (authors)	178	177	167	300	61	163	703	391
	R^2 / χ^2	0.272	0.371	0.246	0.196	0.288	0.064	301.49	0.277

*: p<0.1, **: p<0.05.

Table C9. Grouping synonyms together.

Prototypicality vs. Compromise

Our main analysis focuses on the average creativity rating as the dependent variable. One might wonder whether ideas with an average edge weight distribution present a “compromise” that is judged as more creative on average, but that is not necessarily seen as creative by many judges. In order to address this concern, we repeat our analysis using the proportion of favorable creativity ratings as the dependent variable, instead of the average. In each study except Study 5 in which votes were binary, we run a binomial regression with a logistic link where the dependent variable is the proportion of creativity ratings of 4 or 5 out 5 of received by each idea. All other aspects of the regressions are similar to Table 2. Results are presented in Table C10. We see that a prototypical edge weight distribution is not only associated with higher average creativity ratings, it is also associated with a higher proportion of favorable creativity ratings.

		Study 1a	Study 1b	Study 1c	Study 2	Study 3	Study 4	Study 6
	Distance to prototypical edge weight distribution	-5.224**	-3.176**	-3.018**	-0.586**	-1.430**	-0.103	-0.459**
Edge Weight	Average	-16.730**	-16.233**	-1.873	1.117	-0.901	-0.830	-0.522
	Coef. of var.	1.059**	1.702**	1.202**	-0.104	-0.385	0.819**	0.587*
	Min	3.005	-2.140	11.535	-0.555	0.011	1.466	0.818
	Max	-0.968	-1.904	-0.434	0.206	0.076	-0.336	0.112
Node frequency	Average	-12.586**	7.154*	11.593**	-1.513**	1.648	0.043	-1.089*
	Coef. of var.	2.400**	1.192*	3.156**	0.484	1.455	1.557**	0.331
	Min	20.322**	16.858	8.215	0.729	0.757	1.542**	0.596
	Max	-1.528	4.941	-6.767**	0.315	-1.032	-1.410**	0.466
Node clustering coef.	Average	2.468*	-2.509	3.434**	N/A	N/A	N/A	N/A
	Coef. of var.	0.993	-5.396**	-9.678**	N/A	N/A	N/A	N/A
	Min	-0.213	-2.617**	-5.462**	N/A	N/A	N/A	N/A
	Max	-5.322**	10.241**	9.500**	N/A	N/A	N/A	N/A
	Size of semantic subnetwork	-0.051**	-0.038**	-0.017	0.012**	-0.005	0.079**	-0.034**
	Number of characters/1000	1.853**	2.231**	1.867**	0.039	1.761**	1.663**	2.204**
	# observations	276	271	251	555	173	220	648
	# groups (authors)	178	177	167	300	61	163	391
	χ^2	185.72	299.32	227.88	151.26	107.12	80.61	326.86

*: *p*<0.1, **: *p*<0.05.

Table C10. Using the proportion of creativity ratings of 4 or 5 (out of 5) as dependent variable.

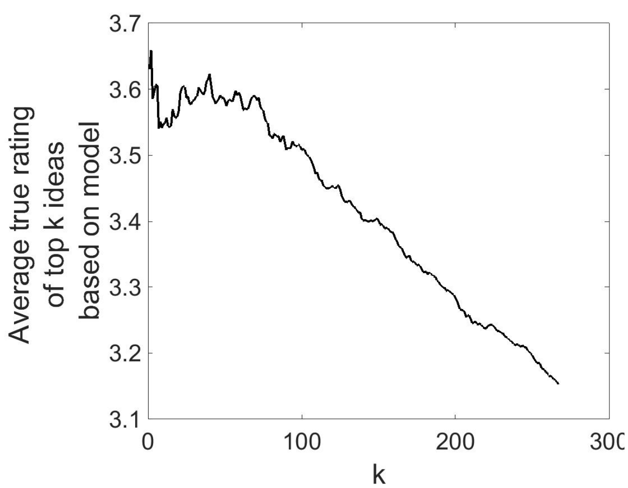
Web Appendix D: Alternative Measures of the Relationship between Prototypicality and Judged Creativity

In each study, we use the model reported in Table 2. For all studies except Study 5, we compute the rank correlation (across ideas) between the fitted (based on the regressions reported in Table 2), and observed creativity ratings. For Study 5, we compute the rank correlation between the fitted and predicted proportion of positive votes.

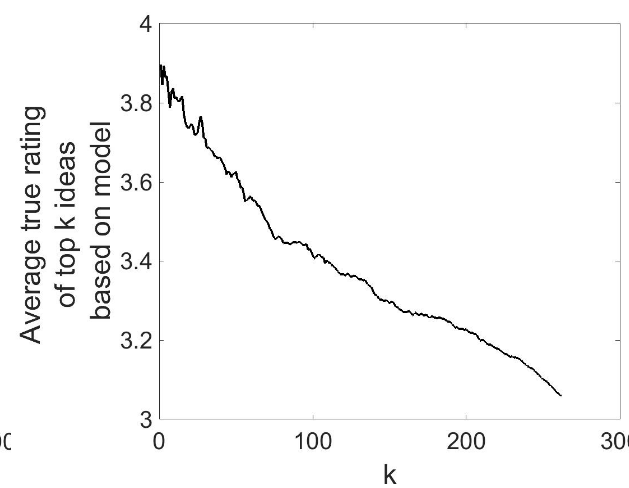
	Rank Correlation	p-value
Study 1a	0.541	<0.01
Study 1b	0.596	<0.01
Study 1c	0.519	<0.01
Study 2	0.411	<0.01
Study 3	0.531	<0.01
Study 4	0.288	<0.01
Study 5	0.103	<0.01
Study 6	0.498	<0.01

Table D1. Rank correlation between fitted and observed creativity ratings.

Additionally, in each study, we use the model reported in Table 2 to rank ideas based on their fitted creativity ratings. We take the top *k* ideas, where *k* varies between 10 and the number of ideas. We compute the average true creativity rating (or in Study 5, average true proportion of positive votes) for the top *k* ideas based on the model. See Figure D1.



Study 1a



Study 1b

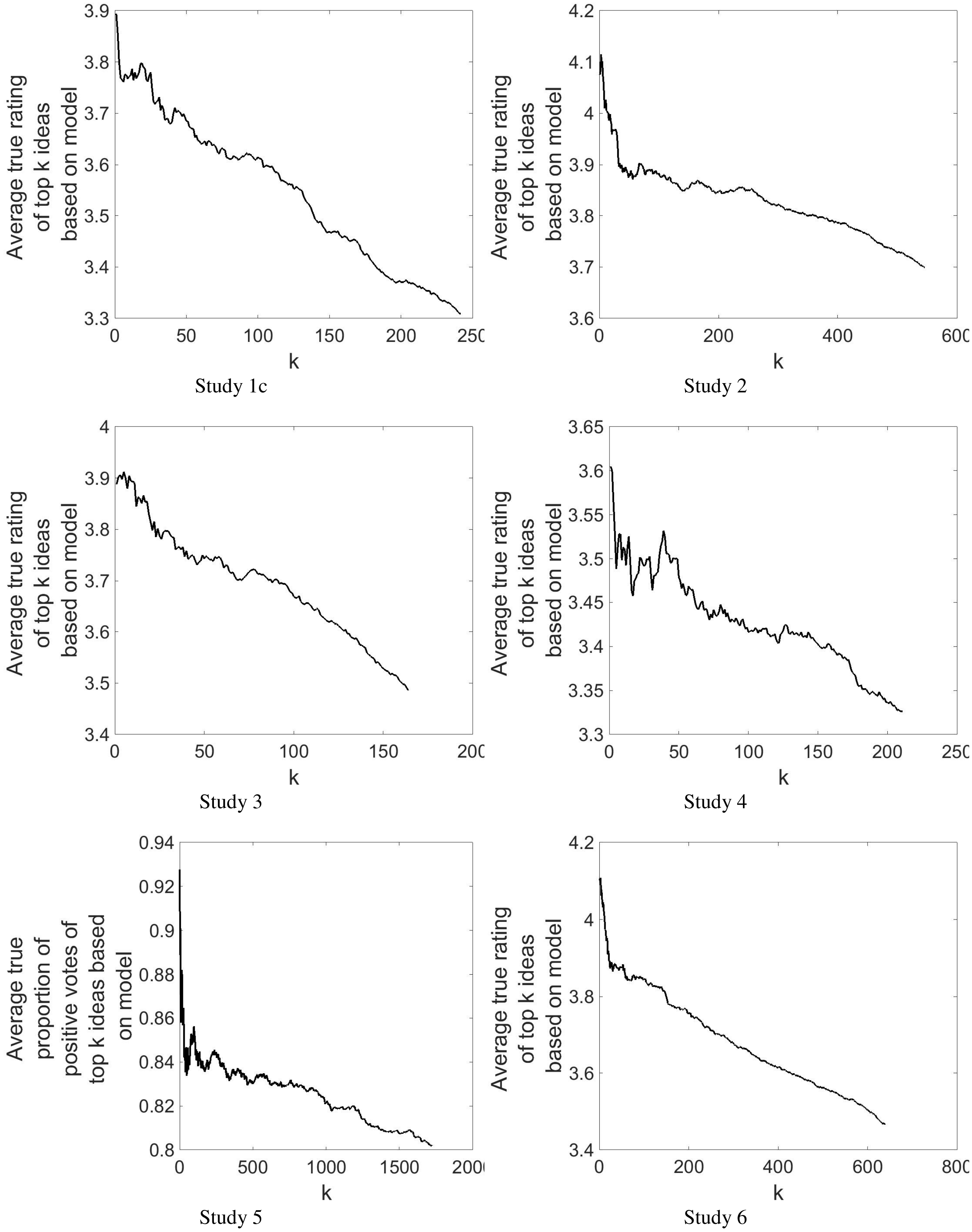


Figure D1. Average true rating of top ideas based on model.

Web Appendix E: Alternative Dependent Variables

In all studies (except Study 5), all ideas were rated on four dimensions: purchase interest, predicted popularity, writing quality, and creativity. The Cronbach α between these four measures ranges between 0.783 and 0.887 across studies, with an average across studies of 0.855. Therefore it is appropriate to construct an overall measure of idea quality by averaging these four dimensions. Table E1 reports the results of a series of regressions in which the dependent variable is this overall measure of idea quality. We see that the effect of prototypicality is statistically significant at $p<0.05$ in 6 out of 7 studies and directionally consistent with the hypothesis in the other. This confirms that ideas with more prototypical edge weight distributions tend to be judged more favorably overall, not just in terms of their creativity.

We also run a series of regressions with purchase interest, predicted popularity, and writing quality as dependent variables. See Tables E2-E4. We see that the coefficient for prototypicality is always directionally consistent with ideas with more prototypical edge weight distributions being judged more favorably (with one exception in Study 4 when the DV is writing quality). However, the results are not as strong and uniformly statistically significant as they are with judged creativity as the dependent variable.

Of particular interest is the series of regression with writing quality as the dependent variable. With this measure as the dependent variable, the coefficient for prototypicality is statistically significant at $p<0.05$ in 3 out of 7 studies, and directionally consistent with the hypothesis in 3 studies. This finding suggests a potential concern that the effect of prototypicality on judged creativity be mediated by writing quality, i.e., ideas with more prototypical edge weight distributions are judged as more creative only because they are better written. We address this concern by performing a series of mediation tests using the bootstrapping approach developed by Preacher and Hayes (2004). For each study, we draw 1,000 sets of observations, where each set is drawn with replacement from the data (the number of observations in each set is the same as in the data), and for each set we run one regression with writing quality as the dependent variable and another with judged creativity as the dependent variable, controlling for writing quality. The p-values for the mediation and direct effects are based on the proportions of times across all 1,000 sets in which regression coefficients and/or combinations thereof have specific signs. We find that writing quality is not a significant mediator for the effect of prototypicality on judged creativity in any of the studies (the *p*-values average 0.24 across studies). Therefore, ideas with more prototypical edge weight distributions are *not* judged as more creative only because they are better written.

We perform similar mediation analysis using purchase interest and predicted popularity as potential mediators. With one exception (purchase interest in Study 1c), none of the mediation tests are statistically significant (the p-values are 0.21 on average for purchase interest and 0.23 for predicted popularity).

		Study 1a	Study 1b	Study 1c	Study 2	Study 3	Study 4	Study 6
	Distance to prototypical edge weight distribution	-1.962**	-1.880**	-2.578**	-0.300**	-0.648**	-0.103	-0.310**
Edge Weight	Average	-6.778	-8.075	12.247*	0.819	0.433	-0.745	0.627
	Coef. of var.	0.488	0.340	0.997*	-0.108	-0.550	0.776**	-0.270
	Min	-11.401	-4.545	0.320	-0.541	-0.177	1.925**	-0.373
	Max	-0.471	-1.399	-0.520	0.270	0.462	-0.456	0.369*
Node frequency	Average	-6.878*	3.826	-0.551	-1.064**	-0.664	-0.631	-0.856*
	Coef. of var.	0.735	0.884	0.458	-0.095	0.525	-0.996*	0.377
	Min	14.058*	8.489**	7.999	0.098	0.061	-0.945	0.237
	Max	0.232	3.216	-2.168	0.333	-0.250	0.871	0.007
Node clustering coef.	Average	0.336	-0.832	-2.447	N/A	N/A	N/A	N/A
	Coef. of var.	-2.028*	-2.482**	-1.892	N/A	N/A	N/A	N/A
	Min	-0.776	-1.352*	-0.788	N/A	N/A	N/A	N/A
	Max	-3.131*	4.317	4.326	N/A	N/A	N/A	N/A
	Size of semantic subnetwork	-0.006	0.024*	0.025*	0.006	0.001	0.068**	-0.002
	Number of characters/1000	0.292	0.272	0.311	0.000	0.584	-1.154*	0.588**
	# observations	276	271	251	555	173	220	648
	# groups (authors)	178	177	167	300	61	163	391
	R^2	0.226	0.367	0.255	0.250	0.319	0.067	0.267

*: p<0.1, **: p<0.05.

Table E1. Using average idea quality rating as dependent variable.

Note: We replace the dependent variable with each idea's average idea quality rating (averaged across the 4 dimensions used in the evaluation). Study 5 is omitted because ideas were not evaluated on that dimension. All other aspects of the regressions are similar to Table 2.

		Study 1a	Study 1b	Study 1c	Study 2	Study 3	Study 4	Study 6
	Distance to prototypical edge weight distribution	-1.075	-1.822**	-3.135**	-0.299*	-0.669*	-0.410	-0.247
Edge Weight	Average	-3.567	1.560	31.219**	0.891	1.319	-0.366	1.347
	Coef. of var.	0.139	0.532	1.255*	0.041	-0.559	0.775	-0.251
	Min	-11.967	-3.544	-5.070	-0.186	-0.145	2.151	-0.436
	Max	-0.142	-3.042**	-0.253	0.193	0.540	-0.735	0.328
Node frequency	Average	-7.893	-0.398	-4.603	-1.479**	-1.849	0.677	-1.147*
	Coef. of var.	0.270	0.341	0.162	-0.306	0.519	-2.002**	0.601
	Min	16.057	4.779	7.378	-0.037	0.391	-2.262**	0.501
	Max	1.354	2.474	-1.677	0.554	-0.399	1.028	-0.244
Node clustering coef	Average	-0.270	-0.727	-5.349**	N/A	N/A	N/A	N/A
	Coef. of var.	-2.414	-2.144	1.658	N/A	N/A	N/A	N/A
	Min	-0.909	-0.785	0.707	N/A	N/A	N/A	N/A
	Max	-3.180	5.134	6.188	N/A	N/A	N/A	N/A
	Size of semantic subnetwork	0.007	0.052**	0.042**	0.004	-0.005	0.089*	0.011
	Number of characters/1000	-0.360	-0.572*	-0.316	0.022	0.276	-2.129**	0.089
	# observations	276	271	251	555	173	220	648
	# groups (authors)	178	177	167	300	61	163	391
	R^2	0.110	0.152	0.227	0.140	0.143	0.125	0.120

*: p<0.1, **: p<0.05.

Table E2. Using purchase interest rating as dependent variable.

Note: We replace the dependent variable with each idea's average purchase interest rating. Study 5 is omitted because ideas were not evaluated on that dimension. All other aspects of the regressions are similar to Table 2

		Study 1a	Study 1b	Study 1c	Study 2	Study 3	Study 4	Study 6
	Distance to prototypical edge weight distribution	-1.438*	-1.755**	-1.829	-0.306**	-0.808**	-0.024	-0.362**
Edge Weight	Average	-4.084	-3.544	19.775**	1.093*	2.423	-0.414	1.630**
	Coef. of var.	0.696	0.052	0.953	-0.010	-0.392	0.729*	-0.516*
	Min	-11.172	-8.491	-9.287	-0.379	-0.028	1.678*	-0.579
	Max	-0.766	-1.251	-0.491	0.116	0.120	-0.493	0.337
Node frequency	Average	-6.328	4.564	-5.100	-1.346**	-2.678**	0.009	-1.136**
	Coef. of var.	0.132	1.205	-0.109	-0.262	0.250	-0.998*	0.896**
	Min	17.307*	7.758*	9.437	0.141	0.465	-1.041	0.513
	Max	1.051	-0.190	-0.137	0.480	0.345	0.374	-0.299
Node clustering coef.	Average	-0.289	-0.669	-5.099**	N/A	N/A	N/A	N/A
	Coef. of var.	-1.999	-1.273	-2.256	N/A	N/A	N/A	N/A
	Min	-0.466	-0.682	0.454	N/A	N/A	N/A	N/A
	Max	-2.951	6.055	7.203*	N/A	N/A	N/A	N/A
	Size of semantic subnetwork	0.019	0.048*	0.029*	0.004	-0.006	0.100**	0.012
	Number of characters/1000	-0.582*	-0.403	-0.211	0.024	0.551	-2.423**	0.079
	# observations	276	271	251	555	173	220	648
	# groups (authors)	178	177	167	300	61	163	391
	R^2	0.130	0.173	0.183	0.156	0.238	0.110	0.174

*: p<0.1, **: p<0.05.

Table E3. Using predicted popularity rating as dependent variable.

Note: We replace the dependent variable with each idea's average predicted popularity rating. Study 5 is omitted because ideas were not evaluated on that dimension. All other aspects of the regressions are similar to Table 2.

		Study 1a	Study 1b	Study 1c	Study 2	Study 3	Study 4	Study 6
	Distance to prototypical edge weight distribution	-2.272**	-1.743**	-2.590**	-0.218	-0.355	0.065	-0.249
Edge Weight	Average	-10.501*	-17.357**	3.985	0.596	-1.439	-0.963	0.190
	Coef. of var.	0.440	-0.146	0.884	-0.276	-1.102**	0.777*	-0.607**
	Min	-16.447	-5.052	7.530	-1.041**	-0.775	2.073*	-0.807*
	Max	-0.278	-0.206	-0.792	0.486**	1.063**	-0.358	0.631**
Node frequency	Average	-3.969	4.231	-1.378	-0.663	1.045	-1.890*	-0.538
	Coef. of var.	1.208**	0.933	0.441	-0.050	0.810	-1.154	0.207
	Min	10.700	10.715**	5.637	-0.080	-0.705	-0.832	-0.235
	Max	-1.003	8.070**	-2.740	0.250	-0.350	1.756**	0.133
Node clustering coef.	Average	0.263	-0.330	-0.772	N/A	N/A	N/A	N/A
	Coef. of var.	-3.735**	-2.171	-1.552	N/A	N/A	N/A	N/A
	Min	-1.345*	-1.574	-0.855	N/A	N/A	N/A	N/A
	Max	-2.704	4.009	-1.221	N/A	N/A	N/A	N/A
	Size of semantic subnetwork	-0.013	0.013	0.022	0.008*	0.012	0.052	-0.009
	Number of characters/1000	0.675**	0.836**	0.836**	-0.014	0.350	-1.228	0.880**
	# observations	276	271	251	555	173	220	648
	# groups (authors)	178	177	167	300	61	163	391
	R^2	0.267	0.446	0.292	0.288	0.430	0.032	0.303

*: *p*<0.1, **: *p*<0.05.

Table E4. Using average writing quality rating as dependent variable.

Note: We replace the dependent variable with each idea’s average writing quality rating. Study 5 is omitted because ideas were not evaluated on that dimension. All other aspects of the regressions are similar to Table 2.

Web Appendix F: Boundary Conditions

Prototypicality Measured Based on Node Frequency

We replace the distance to the prototypical edge weight distribution with the distance to the prototypical node frequency distribution. All other aspects of the regressions are similar to Table 2.

		Study 1a	Study 1b	Study 1c	Study 2	Study 3	Study 4	Study 5	Study 6
	Distance to prototypical node frequency distribution	-0.631	-0.693	1.002	-0.290	-0.686	0.017	-0.270*	-0.183
Edge Weight	Average	3.470	-13.286*	-1.452	0.560	-0.025	-0.639	1.654**	-0.782
	Coef. of var.	-0.150	-0.168	-0.691	-0.059	0.142	0.725	-0.525*	0.446
	Min	-11.634	-16.392**	-14.783	-0.653	-0.288	1.392	1.638**	0.277
	Max	-0.240	0.125	0.938	0.300	0.297	-0.505	0.708**	0.209
Node frequency	Average	-8.031*	8.800*	8.588*	-0.771	0.263	-0.720	-4.298**	-0.800
	Coef. of var.	1.068*	1.317	2.017**	0.034	0.421	0.159	-0.939**	-0.342
	Min	8.018	11.538**	7.082	0.188	-0.006	0.565	-1.860**	0.083
	Max	-0.084	1.471	-4.822**	0.236	-0.368	0.082	1.259**	0.575
Node clustering coef.	Average	0.835	-1.387	2.174	N/A	N/A	N/A	N/A	N/A
	Coef. of var.	-0.685	-3.877**	-5.633**	N/A	N/A	N/A	N/A	N/A
	Min	-0.606	-2.172*	-3.610**	N/A	N/A	N/A	N/A	N/A
	Max	-3.321	4.151	3.369	N/A	N/A	N/A	N/A	N/A
	Size of semantic subnetwork	-0.032*	-0.013	0.018	0.006	-0.001	0.055	-0.002	-0.022**
	Number of characters	1.384**	1.157**	0.922**	-0.032	0.995**	0.804	-0.223**	1.301**
	# observations	276	271	251	555	173	220	1735	648
	# groups (authors)	178	177	167	300	61	163	703	391
	R^2 / χ^2	0.228	0.353	0.223	0.185	0.263	0.053	286.29	0.262

*: *p*<0.1, **: *p*<0.05.

Table F1. Prototypicality measure based on node frequency.

Prototypicality Measured Based on Clustering Coefficient

We replace the distance to the prototypical edge weight distribution with the distance to the prototypical clustering coefficient distribution. Studies 1a, 1b and 1c are the only ones in which the baseline semantic network had a low enough density to induce adequate variations in the clustering coefficient. All other aspects of the regressions are similar to Table 2.

		Study 1a	Study 1b	Study 1c
	Distance to prototypical clustering coef. Distribution	0.934**	0.314	0.263
Edge Weight	Average	2.359	-14.613**	0.262
	Coef. of var.	0.230	-0.188	-0.551
	Min	4.086	-15.752*	-11.414
	Max	-0.734	0.084	0.647
Node frequency	Average	-8.392*	8.355*	9.051*
	Coef. of var.	0.919	1.268	2.292**
	Min	8.835	10.537**	10.559
	Max	0.149	2.018	-5.196**
Node clustering coef.	Average	3.744*	0.127	2.845
	Coef. of var.	0.843	-3.100*	-3.697
	Min	-0.488	-2.109*	-3.027**
	Max	-3.677	4.111	3.427
	Size of semantic subnetwork	-0.035**	-0.014	0.011
	Number of characters/1000	1.350**	1.183**	0.997**
	# observations	276	271	251
	# parameters	178	177	167
	R^2	0.240	0.348	0.218

*: *p*<0.1, **: *p*<0.05.

Table F2. Prototypicality measure based on clustering coefficient.

Vector Space Representation

We compute the inverse document frequency of each word stem as $idf = \log(N/df)$ where N is the number of training documents (Pretest ideas or Google results) and df is the number of training documents that contain this word stem. Then for each document we create a vector with one component for each word stem, which is 0 if the stem does not appear in this document, and equal to its idf otherwise. We compute the average vector over the training documents (Pretest ideas or Google results). This average vector has entry $-p_w \log(p_w)$ for word stem w, where p_w is the proportion of training documents that contain word stem w. For each idea, we compute the Euclidean distance between the vector representing that idea and the average vector from the training dataset. Table F3 report the results of a series of regressions with the same dependent variables as in Table 2.

	Study 1a	Study 1b	Study 1c	Study 2	Study 3	Study 4	Study 5	Study 6
Distance to prototypical document	0.075**	0.044*	0.008	0.572**	0.562	0.280	0.386**	0.303*
Number of characters/1000	0.287	1.026**	1.037**	0.050	1.126**	1.182**	-0.373**	0.807**
# observations	276	271	251	555	173	220	1735	648
# groups (authors)	178	177	167	300	61	163	703	391
R^2 / χ^2	0.148	0.188	0.160	0.093	0.209	0.024	42.02	0.219

*: $p<0.1$, **: $p<0.05$.

Table F3. Prototypicality measure based on vector space representation.

Latent Dirichlet Allocation (LDA)

LDA (Blei et al., 2003; Tirunillai and Tellis, 2014) takes a "bag of words" approach, i.e., each document is characterized by a set of tokens. We define the following for each study:

d indexes documents in the training set (pretest ideas or Google results).

k indexes topics.

w indexes word stems (which are referred to as "words" here).

φ_k is a vector that contains the set of topic-word associations for topic *k*.

θ_d is a vector that contains the set of document-topic associations for document *d*.

w_i^d is the index of the word associated with token *i* in document *d*.

z_i^d is a latent variable equal to the index of the topic to which token *i* in document *d* belongs.

LDA assumes the following data generating process:

Prob(z_i^d) ~ *Multinomial*(θ_d)

Prob($w_i^d|z_i^d=k$) ~ *Multinomial*(φ_k)

We assume the following priors with the standard hyper parameters set to α=1 and β=0.01:

θ_d ~ *Dirichlet*(α)

φ_k ~ *Dirichlet*(β)

Given this specification, parameters may be estimated using Gibbs sampling, based on the posterior distributions of all variables, which are given in closed form as follows:

$$Prob(z_i^d = k|w_i^d, \{\varphi_k\}, \theta_d) = \frac{Prob\left(w_i^d \middle| z_i^d = k, \varphi_k\right) Prob(z_i^d = k|\theta_d)}{\sum_{k'} Prob\left(w_i^d \middle| z_i^d = k', \varphi_{k'}\right) Prob(z_i^d = k'|\theta_d)} = \frac{\varphi_k(w_i^d)\theta_d(k)}{\sum_{k'} \varphi_{k'}(w_i^d)\theta_d(k')} \quad \text{(F1)}$$

$$Prob\left(\theta_d \middle| \{z_i^d\}\right) = Dirichlet\left(\alpha + m_d^1, \alpha + m_d^2, \dots, \alpha + m_d^K\right) \quad \text{(F2)}$$

$$Prob\left(\varphi_k \middle| \{z_i^d\}, \{w_i^d\}\right) = Dirichlet\left(\beta + n_k^1, \beta + n_k^2, \dots, \beta + n_k^W\right) \quad \text{(F3)}$$

Where m_d^k is the number of tokens in document *d* that were assigned to topic *k*, and n_k^w is the number of tokens equal to word *w* that were assigned to topic *k*, across documents.

We estimate this model using MCMC with 5,000 iterations, using the first 1,000 as burn-in and saving one in 10 iterations thereafter. We estimate the model for each training dataset (pretest ideas or Google results) separately, with a number of topics varying from 1 to 30. We select the number of topics for each training dataset using the Deviance Information Criterion (Celeux et al., 2006), giving rise to a number of topics that varies between 12 and 22 across studies.

We estimate the topic distribution of ideas submitted in each study based on the model estimated on the training documents for that study. For each MCMC iteration that is saved, we draw a value of z_i^d for each token in each idea based on Equation F1 and a value of θ_d for each idea based on Equation F2. We use the posterior mean of the topic distribution as our point estimate, i.e., each idea is represented as a vector with dimensionality equal to the number of topics. For each idea we compute the distance to the prototypical document as the Euclidean distance between the vector representing that idea and the average vector from the training dataset (pretest ideas or Google results). Table F4 reports a series of regressions where the dependent variables are the same in Table 2.

	Study 1a	Study 1b	Study 1c	Study 2	Study 3	Study 4	Study 5	Study 6
Distance to prototypical document	-0.520	-3.328**	-2.443	3.251**	-3.954**	2.111	9.057**	-0.851
Number of characters/1000	0.997**	1.406**	1.057**	0.176**	1.699**	1.136**	-0.176**	1.049**
# observations	276	271	251	555	173	220	1735	648
# groups (authors)	178	177	167	300	61	163	703	391
R^2 / χ^2	0.117	0.226	0.169	0.056	0.196	0.036	206.19	0.215

*: *p*<0.1, **: *p*<0.05.

Table F4. Prototypicality measure based on vector space representation. Analysis based on topic modeling.

Misspecification of the Baseline Semantic Network

We examine whether the baseline semantic network (and the resulting prototypical edge weight distribution) need to be topic-specific or whether we can use a baseline semantic network from one topic to evaluate ideas generated on a different topic. We replicate the analysis from Table 2 for Studies 1a-1c, for each baseline semantic network and corresponding prototypical edge weight distribution. For space reasons, we only report the coefficient on the distance to the prototypical edge weight distribution. See Table F5. The diagonal terms are identical to Table 2; the off-diagonal terms correspond to situations where the baseline semantic network is misspecified. We see that the coefficient is statistically significant only when the baseline semantic network is correctly specified, i.e., when it is based on the same idea generation topic as the study itself.

	Study 1a	Study 1b	Study 1c
Baseline semantic network from Study 1a	-3.294**	-0.894	-1.594
Baseline semantic network from Study 1b	2.045*	-2.410**	-0.225
Baseline semantic network from Study 1c	0.705	-0.754	-3.116**

*: p<0.1, **: p<0.05.

Table F5. Coefficient on distance to prototypical edge weight distribution when baseline semantic network is misspecified.

Web Appendix G: Additional Analyses for Study 6

Before comparing the judged creativity of ideas across conditions, we compare the *number* of ideas. The number of original idea per participant is not statistically different across conditions (all p's>0.12 – Wilcoxon rank sum tests). This should be expected because the conditions differ only in how they encourage participants to modify their original ideas, not in the submission of original ideas. In terms of the average number of modifications per original ideas, the *Random Words* condition is not statistically different from the *Control* condition (p>0.22), but both the *Minimum Distance* and *Maximum Prototypicality* conditions are significantly larger than the *Control* condition at p<0.05. In other words, both the *Minimum Distance* and *Maximum Prototypicality* conditions increase participants' propensity to modify their ideas compared to a control that does not suggest words to participants or to a condition where words are selected randomly. The average number of modifications per original idea is marginally significantly higher in the *Minimum Distance* condition compared to the *Maximum Prototypicality* condition (p=0.08). This is not surprising, because the words offered to participants in the *Minimum Distance* condition were selected to be highly connected to the words currently in the idea, making it easy to modify ideas.

We now turn to the more central analysis of the *judged creativity* of the modified ideas. The *Minimum Distance* condition, despite giving rise to the largest number of modifications per idea, gives rise to modifications that do not improve the original idea: the average difference between the judged creativity of the last version of a modified idea and the original idea is not statistically different from 0 in this condition (p=0.25), and it is significantly lower than in each of the three other conditions (p<0.04, 0.07 and 0.02 relative to *Control*, *Random Words*, and *Maximum Prototypicality,* respectively). The other three conditions all give rise to statistically significant improvement in judged creativity for the modified idea over the original idea (all p's<0.01), and are not statistically different from one another (all p's>0.65).

On the one hand, recommending words using the maximum prototypicality criterion generated a relatively large number of modifications and significantly improved the judged creativity of the ideas. On the other hand, one may wonder why asking participants to improve their ideas without any word recommendation or using random words led to similar improvements in judged creativity, *conditional* on an idea being modified. To investigate this question we consider the change in the Kolmogorov-Smirnov statistic between the last modified version of an idea and its original version (see last column in Table G1). The results mirror those based on judged creativity. The *Minimum Distance* condition is not statistically different from 0 (p=0.51), it is significantly higher (i.e., closer to 0) than the three other conditions (all p's<0.01), which all give rise to differences in the Kolmogorov-Smirnov statistic that are significantly different from 0 (all p's<0.03) but not significantly different from one another (all p's>0.34). Moreover, the correlation (across conditions) between the change in the Kolmogorov-Smirnov statistic and the change in judged creativity is statistically significant (ρ=-0.318, p<0.01). This correlation is significant in the *Control* condition (ρ=-0.362, p<0.03) and the *Maximum Prototypicality* condition (ρ=-0.327, p<0.02) and marginally significant in the *Random Words* condition (ρ=-0.252, p<0.09). However it is not significant in the *Minimum Distance* condition (ρ=-0.121, p=0.35). This suggests that people tend to improve the prototypicality of their ideas when attempting to improve their ideas, whether this process is guided or spontaneous. The *Maximum Prototypicality* condition appears to have facilitated this process by recommending good candidate words, and the *Minimum Distance* condition appears to have inhibited it by only recommending words that were too closely related to those in the previous idea.

Finally, we explore asymmetries in the results using a signed Kolmogorov-Smirnov statistic. We separate the original ideas that were modified at least once between those with a positive Kolmogorov-Smirnov statistic (i.e., more small weights and therefore more novel combinations) vs. a negative Kolmogorov-Smirnov statistic (i.e., more large weights and therefore more familiar combinations). When focusing on original ideas with a positive Kolmogorov-Smirnov statistic only, the average change in the signed Kolmogorov-Smirnov statistic between the last modified version of an idea and its original version is significantly negative for all conditions except the *Control* condition (p>0.26), but the correlation between the decrease in the Kolmogorov-Smirnov statistic and the change in judged creativity is not statistically significant overall or in any condition (all p's>0.38). When focusing on original ideas with a negative Kolmogorov-Smirnov statistic, however, the pattern of results is more similar to the results obtained with the non-signed Kolmogorov-Smirnov statistic. The average change in the signed Kolmogorov-Smirnov statistic from the original to the last modified version of the idea is significantly positive (i.e., the Kolmogorov-Smirnov statistic is decreased) in all conditions and marginally significant in the *Minimum Distance* condition ($p<0.06$). The correlation between the change in the Kolmogorov-Smirnov statistic and the change in judged creativity is significant overall (ρ =-0.418, $p<0.01$), and it is significant in all conditions except the *Minimum Distance* condition ($p>0.31$). This suggests that the tool worked primarily by helping participants with ideas that were too familiar increase the novelty of their ideas, but that participants with ideas that were too novel were not able to increase familiarity in a meaningful way using the suggested words.

Condition	Average number of original ideas per participant	Average number of modifications per original idea	Average difference in judged creativity between last modification and original idea	Average difference in Kolmogorov-Smirnov statistic between last modification and original idea
Control	1.590	0.333	0.236	-0.106
Random Words	1.600	0.489	0.166	-0.084
Minimum Distance	1.874	0.838	0.032	-0.014
Maximum Prototypicality	1.643	0.604	0.220	-0.109

Table G1. Study 6 results.

References

Blei, DM, Ng, AY, and Jordan, MI. 2003. “Latent Dirichlet Allocation.” Journal of Machine Learning. 3. 993-1022.

Celeux, G, Forbes, F, Robert, CP, and Titterington, DM. 2006. “Deviance Information Criteria for Missing Data Models.” *Bayesian Analysis*. 1 (4). 651-673.

Mika, Peter. 2007. “Electronic Sources for Network Analysis.” In *Social Networks and the Semantic Web*. 51-63. Springer: New York, NY.

Preacher, K and Hayes, A. 2004. “SPSS and SAS Procedures for Estimating Indirect Effects in Simple Mediation Models.” *Behavior Research Methods, Instruments, and Computers*. 36(4), 717–731..

Princeton University. 2010. “About WordNet.” http:/wordnet.princeton.edu.

Salton, G and MJ McGill. 1983. *Introduction to Modern Information Retrieval*. McGraw-Hill: New York, NY.

Tirunillai, S and Tellis, GJ. 2014. Mining Marketing Meaning from Online Chatter: Brand Analysis of Big Data Using Latent Dirichlet Allocation. *Journal of Marketing Research*, 51(4), 463-479.