CREATIVITY RESEARCH JOURNAL, *21*(1), 64–71, 2009

ISSN: 1040-0419 print/1532-6934 online
DOI: 10.1080/10400410802633517

A Semantic Network Approach to the Creativity Quotient (CQ)

Terry Bossomaier
Centre for Research in Complex Systems, Charles Sturt University, and Centre for the Mind

Mike Harré, Anthony Knittel, and Allan Snyder
Centre for the Mind

The Creativity Quotient (CQ) is a novel metric building on ideational fluency that accounts for both the number of novel ideas (ideation) and the number of distinct categories (fluency) these ideas fall into. Categories are, however, difficult to define unambiguously and objectively. We propose that the principal contribution of this article is an entirely algorithmic approach based on concept networks, and an information metric defined thereon. It requires only measures of the similarity between concepts, which may come from databases such as Wordnet, Wikipedia, Google, or corpus analysis tools. In the special case of strong, unique categories it reduces directly to CQ.

There are many ways one might approach the measurement of creativity, of which one of the oldest is ideational fluency (Thurstone, 1924). In an ideation fluency test, a subject gives as many uses as they can think of for some everyday object, such as a banana or a piece of paper. But, should each response be scored equally or should certain answers be weighted more than others (Getzels & Jackson, 1962; Guilford, 1959; Hocevar, 1979; Kaufman, 1981; Obonsawin, Crawford, Page, Chalmers, Cochrane, & Low, 2002; Runco & Mraz, 1992; Ward, 1969)? The development of scoring procedures for creativity tests presents an unusual problem (Getzels & Jackson, 1962; Guilford, 1959). It is intuitive that the suggested uses should not be weighted equally, as is often done (Bryan & Leszcz, 2000; Obonsawin et al., 2002). In particular, those uses offered in distinctly different categories should be weighted more than those that fall in the same category (Getzels & Jackson, 1962; Guilford, 1959).

This article argues that understanding conceptual networks is important for measuring fluency. Within this framework, not all responses are equal, as some have stronger connnections to other responses. Given that one use of a banana is to eat it, feeding it to a chimpanzee does not seem that original. Using it as a source of yellow dye is less obvious. The first use is an everyday use, familiar to many people already. The others exhibit replacement creativity, where a component in a system is replaced by something, as where Salvador Dali replaced the handset of a telephone with a lobster (Goldenberg, Mazursky, & Solomon, 1999). The replacement might indicate how information is organized within a person's mind. If so, could this imply general principles of organization that hold for sets of people?

Snyder, Mitchell, Bossomaier, and Pallier (2004) advanced a new information metric of creativity called the Creativity Quotient (CQ), which accounted for both ideation fluency and flexibility; however it requires a definition of categories (flexibility) of use. Although this is quite intuitive, it is somewhat difficult to define categories in any truly objective, or unique, way. For example, when considering uses for a piece of paper, a paper napkin could be categorized as a folding use, similar to a paper airplane or origami. But it

This project was partly funded by ARC Discovery Grant DP0211847 to Snyder and Bossomaier. Thanks to Toby Hawker for programming support provided under this grant.

Correspondence should be sent to Terry Bossomaier, Centre for Research in Complex Systems, University of Sydney, Australia. E-mail: tbossomaier@csu.edu.au

could also be put into the category of wiping things, as in blotting paper or kitchen towel. Although this need not seriously detract from the usefulness of CQ in any practical application, it is valuable to explore the possibility of developing a wholly objective measure across subject cohorts.

Categories are of interest to neuropsychologists concerned with cortical organization of concepts. Stroke victims may exhibit very specific deficits; for example they may lose the names of vegetables although the names of fruits remain intact, or they may lose their memory for faces, yet be able to deduce who somebody is from a mixture of other information. Categories for common inanimate and animate objects clearly do exist, but they are often hard to definitively localize within the brain. For example, the Facial Fusiform Area (FFA), so-called because it seemed, for a decade, to be the site of recognition of faces, is now known to code for other things too, while other areas carry memory about faces as well (Haxby et al., 2001). Equally surprising, the FFA responds to face contexts, even where there is no face or the face is blurred beyond recognition (Cox, Meyers, & Sinha, 2004). The most recent evidence moves back towards the superceded notion of a grandmother cell—that there is a cell that responds to, and only to, a very specific idea, such as that of one's grandmother. Recently, cells responsive to very specific subcategories, such as Bill Clinton or Halle Berry (Quiroga, Reddy, Keriman, Kock, & Fried, 2005) have been discovered.

Choosing categories is even more difficult if hierarchies as in, say, the taxonomy of animal species, are admitted. In the paper-use ideation fluency test, most subjects offer writing, painting, drawing, and sometimes typing/printing as distinct uses. Do these all belong in the category of surface marking or absorption? But drawing could easily be placed into a separate category (map, sketch, diagram, chart, plan, etc.).

Even more complex is the way categories form dynamically in the frontal cortex. So as images of a cat are morphed into a dog, cells in the visual cortex and subsequent temporal areas respond to these hybrids in a graduated way (Freedman, Riesenhuber, Poggio, & Miller, 2001). But in the frontal lobes, the categories are distinct. Even for 40% cat, cat neurons do not fire. Yet such distinct categories are dynamic and the same cells can rapidly change to categories of lion and wolf versus cat and Chihuahua. The alternative, presented in this article, to determining categories subjectively is to derive them from concept networks, where each concept forms a node in a graph.

The next section examines the CQ creativity quotient and related network analysis and approximation.

THE CREATIVITY QUOTIENT (CQ)

Snyder, Mitchell, Bossomaier, and Pallier (2004) derived a CQ metric that accounted for both ideation fluency (number of ideas) and flexibility (number of categories). This measure allocates the responses from an ideation fluency test into a number of categories, where the number of elements in each category is meaningful. The CQ metric, Q, is given in equation (1), where N is the number of categories and n_j is the number of elements in each category.

$$Q = \sum_{j=1}^{N} \log_2(n_j + 1) \tag{1}$$

Equation 1 can be understood intuitively or derived mathematically as shown in (Snyder, Mitchell, Bossomaier, et al., 2004). Equation 1 is exact given a hypothetical unknown set of independant categories. It is also a limiting case of the network model described below.

NETWORK METRICS

A subsuming approach is to begin with a network, which describes the strength of relationships between all concepts. So, for example, the connection between lion and tiger is strong, the connection between lion and tree much weaker (although they still share the property of being living things). There are many approaches that can be used to obtain the connection strengths between each pair of concepts.[1] Through creating a hierarchy in which each concept is linked below to the concept with which it has the strongest relationship, an effective set of categories emerges, which may result in deeper intuition into the meaning of the CQ metric. The nature of this relationship is discussed in more detail below.

An explosion of interest in networks has uncovered a range of properties that seem to be highly relevant to cognition. Concept networks give us a framework for understanding fluency and creativity in terms of paths and path likelihoods between concepts.

Language, of course, is full of categories, and a rich source of theoretical analysis. The edges of the graph between nodes (concepts) have a weight that can be a similarity or a distance between the nodes. The edge may also have a direction. But the relationship between word and concept networks is not isomorphic. Words have multiple meanings (polysemy and homonymy)

[1]At present, different methodologies for determining the distance between concepts do not give identical results. This is a very active area of research, driven by the explosive growth in search technologies, and some convergence is likely.

and individual concepts may be represented by multiple words (synonymy). As such, there is considerable variance in measures of word similarity, as demonstrated by Seco, Veale, and Hayes (2004).

FROM CATEGORIES TO HIERARCHIES AND GENERAL NETWORKS

It makes intuitive sense to think of the existing concept structure of the mind as a network, as searching and thinking seem to follow 'links' from one idea to another. Abundant evidence points to there being delays in retrieving information that somehow relates to brain organization. Thus, for example, the names of individual dogs (Corgi, Dalmatian, etc.) can be retrieved faster, and with lower initial latency, than examples from a higher order category, such as animals. At first sight, the process of following links might seem to be slow, but recent work on small world (Watts, 1999) and scale-free networks (Barabási, 2002) has shown that the number of links required to move from any node to any other node may be very small, e.g., 19 for the hundreds of millions of Web pages (Albert, Jeong, & Barabási, 1999).

To determine the rarity of a response given the responses that have already been made, the probability of a use being given, and its associated rarity, can be defined as a chain of conditional probabilities (Russell & Norvig, 1995).

Consider a charcoal sketch as a use of a piece of paper. One hypothetical way one might come up with such an idea is through thinking about burning paper to make charcoal and then applications of charcoal. Once one has got to charcoal, other uses might easily spring to mind, such as use as a toxin absorber. But a charcoal sketch could be reached through

$$\mathrm{p(sketch)} = \mathrm{p(sketch|charcoal)p(charcoal|burn)p(burn)} \quad (2)$$

where the conditional probability with every other use that has been given is taken into account, allowing the interrelation of concepts under multiple contexts to be included. Figure 1 shows an illustrative category tree for two classes of bird. The parrots share some properties (e.g., nut-cracking/plant-eating beak) that the birds of prey do not have. But they, too, have distinctive shared properties (e.g., flesh-tearing beaks). Thus, there is a higher level of mutual information, or concept mutuality, among the parrots or falcons than amongst a mixture of the two. The ideal categories would have none of these dependencies between categories and there would be no mutual information between them. Another way of expressing this is through the multi-information, I_m, which is shared information across the whole set of concepts or responses. So, the birds in the above example, share properties of wings, feathers, and so on. This has a precise definition in communication theory, given in Appendix A. Good responses will have low concept mutualities.

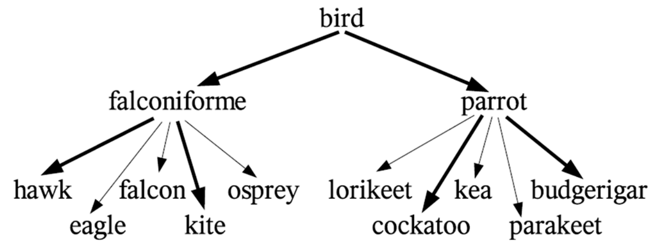

FIGURE 1 A conceptual tree.

Given the concept mutality, it is possible to estimate the multi-information which is the shared information across the entire set. This is obtained by calculating the maximum spanning tree, across the network of mutual information values between responses. The best creativity potential values will minimize the multi-information. This leads to a more general measure, denoted Q_{eff} and derived in the appendix, given by

$$\mathrm{Q_{eff}} = \mathrm{N} - \mathrm{I_m} \quad (3)$$

In the ideal case where the responses do fall into groups (effectively the categories) and where there is no mutuality between members of different groups, then the multi-information becomes the sum of the joint mutuality of each group, and $Q_{eff} = Q$.

In summary, to calculate CQ without recourse to heuristic category definition, the procedure is as follows:

1. Compute the concept mutuality between every pair of responses;
2. compute the multi-information of the response set using the maximum spanning tree;
3. use Equation 3 to calculate Q_{eff}.

CALCULATING CONCEPT CONNECTION STRENGTHS

There are, in broad terms, two ways of determining linguistic networks. One is to use human subjects to determine linkages directly. Compendia, such as thesauri, are a realization of this approach where human analysis of language usage creates word relationship data. Wordnet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990) is a compendium of human judgments about similarities between concepts, or syn-sets organized as a set of trees with a few cross connections.

The other approach is to take large bodies of text and use statistical methods to determine word relationships, known as latent semantic analysis (LSA; Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990). There is great interest in this domain, because it is language independent, fast, cheap, quasi-objective, and underlies new directions in search technologies. A recent approach to LSA is to use the Web as the data source, with Google (or similar search engine) as a method for determining strength of connections between concepts. The normalized number of conjoint hits serves as a measure of similarity (Cilibrasi & Vitanyi, 2004).

LSA methods are plagued by problems of polysemy and homonymy—words with multiple meanings. Some words given as uses for paper (e.g., *plan*, *cut*, *score*) may be interpreted in several ways that are not obvious from the wording of the response itself: planning tasks versus a (floor) plan, etc.; using the paper to cut something else versus cutting of the paper; using the paper as a score sheet versus weakening it along a line. So there might be a spurious connection between computers and potatoes via the links computer–chip (semiconductor or food)–potato.

Because, for this article, the source of concept mutuality measure is immaterial, Wordnet was selected where the polysemy/homonymy problem has already been addressed. Concepts appear as *syn-sets*, and any given word may appear multiple times. For simplification, referred to henceforth as concepts (which are represented in the figures by single words).

Seco et al. (2004) provided a way of getting similarities between any two concepts using Wordnet. The degree of relatedness of two concepts is derived from how many concepts they share. So tiger and lion share all the features of cats and are, thus, closer than lion and wolf, which share only the properties of large carnivorous mammals. By tracing the probability in traversing the hierarchy from the most recent common ancestor of both concepts being considered down to each concept, a value for the information content of moving between them is obtained. The information content for a concept comes directly from the number of hyponyms, the more hyponyms the lower the information content I_i of node i is, i.e.,

$$I_i = -log\frac{h_i}{w} \tag{4}$$

where h_i is the number of hyponyms of concept i, and w is the total number of concepts in Wordnet. This constant acts as a normalizing factor changing the count of the number of hyponyms that a concept, i, has, to the probability of a concept in the minimum information tree being subordinate to i. Thus, to say that one has spotted a big cat conveys less information than to say that the cat was a lion as a cat has more hyponyms than does a lion. However, because of multiple inheritance, the information content of each concept cannot be summed to get a total information for a set of concepts. Thus, jaguar and leopard are both large carnivorous cats but having carnivore as the common hypernym is arbitrary. One could equally well envisage *leopard* and *jaguar* as hyponyms of spotted animal, along with fawns and dalmatians.

ILLUSTRATIVE EXAMPLES

Existing datasets of ideation fluency test responses, obtained in a pilot study that preceded (Snyder, Mitchell, Ellwood, et al., 2004), were used to illustrate the application of the theoretical framework. The main data set was collected from a group of 62 subjects in written form, each response consists of a word or phrase that has been reduced for analysis, described as follows.

Data Analysis

The data were analyzed using software written in Perl 5.8.2 (Wall & Schwartz, 1991), C++ (Stroustrup, 1986) and Java 1.4.2 (Arnold & Gosling, 1998), running on a personal computer (Macintosh Quad G5; Apple Computer, Cupertino, CA).

Significant obstacles in the analysis of the responses is that a common concept may be given by a number of respondents using different words, known as synonymy, and that a given word may be interpreted in a number of ways, known as homonymy or polysemy, depending on how related the concepts are (Ullman, 1970). For example, the responses *set fire to*, *burn*, and *burning* may have the same intended semantic meaning, but were presented in different lexical forms. To identify the frequency of

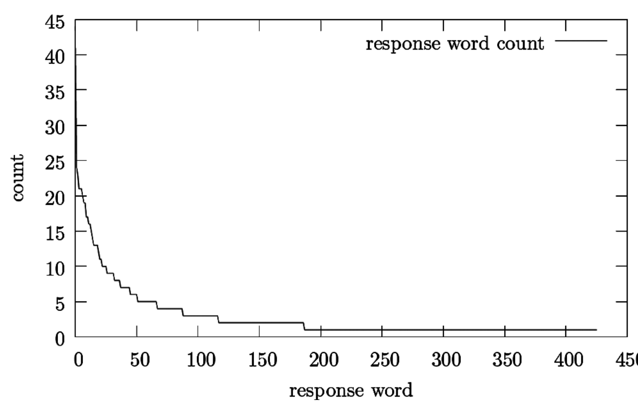

FIGURE 2 Histogram of paper uses. Note that more than half of responses occur once only.

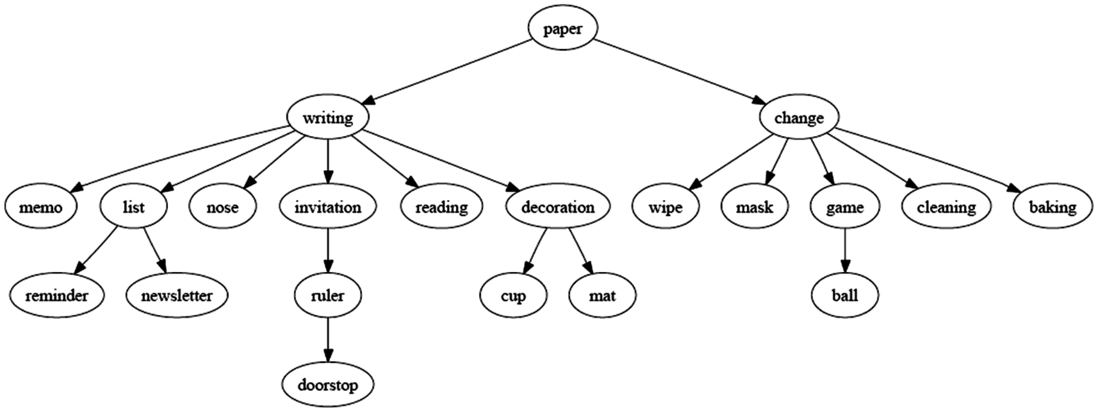

FIGURE 3 The minimal spanning tree (MST) of a random response set (8.0).

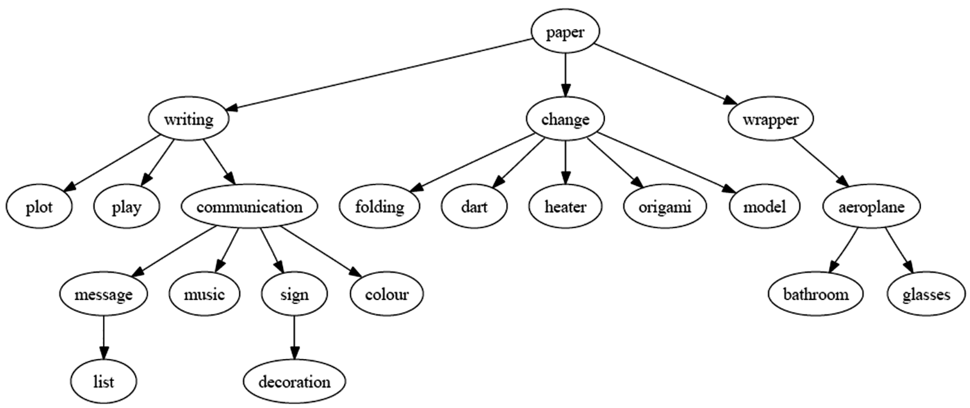

FIGURE 4 The MST of a random response set with low CQ value (7.4).

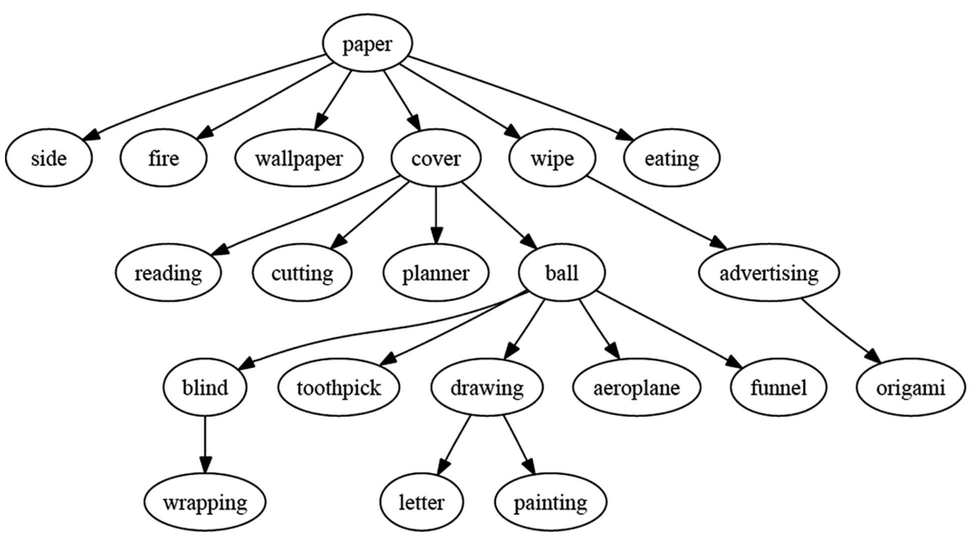

FIGURE 5 The MST of a random response set with low CQ value (12.3).

co-occurance of responses, such synonymous responses were identified.

An additional factor that affects identifying co-occurance of responses is distinguishing the point of separation between concepts that are to be considered equivalent, and those that are similar but considered different. For example, the responses *drawing* and *sketching* have common semantic elements but may be considered distinct, depending on the degree of similarity chosen to consider responses as separate.

Determining the degree of similarity between any two concepts and the degree of similarity necessary to distinguish concepts is subjective and generally problematic. To avoid these issues, words were identified as equivalent only if they have a common lexical root, for example *plane* and *airplanes* are considered the same, whereas *ignite* and *burn* are considered separate. This approach does not require subjective assessment of similarity, and commonality between words for the purposes of assessing the CQ value can be obtained from the external corpus, in this case Wordnet.

Responses which clearly had no association as a use of 'paper' have been removed. This filtering process requires a somewhat subjective judgement; in order to minimize subjectivity responses are assessed simply on whether the assessor can identify any reasonable interpretation of the response as a use for 'paper,' removing any need to assess a degree of appropriateness for each response.

Creativity and Fluency Trees

Figure 2 shows a histogram of the uses of paper generated by the subjects. Examples of uses can be seen from the spanning trees of response sets given in Figures 3–5. It is immediately clear that a small number of uses dominate the responses. In fact, more than half of uses appear once only in the set of 62 subjects. Thus, without an external taxonomy or corpus, collecting statistically meaningful results for these smaller uses would require a huge number of subjects. The very large fraction of unique responses introduces significant sources of error if each such response is allocated to a category by a test administrator. The approach of this article removes such variability.

Figures 3–5 show a typical response network with the connections between nodes weighted by the Wordnet data and the associated minimal spanning tree (MST). These networks were generated by evaluating the similarity between each pair of uses in the set, and finding the minimum spanning tree of the resulting clique graph. The response sets in Figure 3 were produced by randomly selecting words from the list of uses presented in the sample data, where the probability of each use being chosen is given by the frequency distribution in Figure 2. Each response set has 20 words, excluding *paper*, a typical number drawn from a wide range of response sizes. The Q_{eff} values for the three diagrams are 8.0, 7.4, and 12.3. The response sets in Figures 3–5 were chosen out of 100 random sets, produced with an equal probability of each word in the use list being chosen. The theoretical minimum value of a set of 20 words (excluding *paper*) is 4.3, and the maximum 20. These values represent a logarithmic scale, such that comparison of values within this range allows a vast degree of variation. Examining these trees, it is clear that the choice of categories is actually pretty difficult. The metric is very sensitive to the number of categories, making this a particularly important issue.

REFERENCES

Albert, R., Jeong, H., & Barabási, A.-L. (1999). Diameter of the World Wide Web. *Nature*, *401*, 130–131.

Arnold, K., & Gosling, J. (1998). *The Java programming language* (2nd ed.) New York: ACM Press/Addison-Wesley Publishing.

Barabási, A.-L. (2002). *Linked.* Cambridge, MA: Perseus.

Bryan, J., & Leszcz, M. A. (2000). Measures of fluency as predictors of incidental memory among older adults. *Psychology and Aging*, *15*, 483–489.

Cilibrasi, R., & Vitanyi, P. M. B. (2004, December). *Automatic meaning discovery using Google*. ArXiv Computer Science e-prints. Retrieved August 20, 2006 from http://arixiv.org/abs/cs/0412098

Cox, D., Meyers, E., & Sinha, P. (2004). Contextually evoked object-specific responses in human visual cortex. *Science*, *304*(5667), 115–117.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407.

Freedman, D., Riesenhuber, M., Poggio, T., & Miller, E. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*(5502), 312–316.

Getzels, J. W., & Jackson, P. W. (1962). *Creativity and intelligence*. New York: John Wiley.

Goldenberg, J., Mazursky, D., & Solomon, S. (1999). Creative sparks. *Science*, *285*(5433), 1495–1496.

Guilford, J. (1959). *Personality*. New York: McGraw-Hill.

Haxby, J., Gobbini, M., Furey, M., Ishai, A., Schouten, J., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2435–2430.

Hocevar, D. (1979). A comparison of statistical infrequency and subjective judgement as criteria in the measurement of originality. *Journal of Personality Assessment*, *43*, 297–299.

Kaufman, G. (1981). The functional significance of visual imagery inideational fluency performance. *Journal of Mental Imagery*, *5*, 115–120.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to Wordnet: An on-line lexical database. *International Journal of Lexicography*, *3*(4), 235–244.

Obonsawin, M. C., Crawford, J. R., Page, J., Chalmers, P., Cochrane, R., & Low, G. (2002). Performance on tests of frontal lobe function reflect general ability. *Neuropsychologia*, *40*, 970–977.

Quiroga, R. Q., Reddy, L., Keriman, G., Kock, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*, 1102–1107.
Runco, M. A., & Mraz, W. (1992). Scoring divergent thinking tests using total ideational output and a creativity index. *Educational & Psychological Measurement*, *52*, 213–221.
Russell, S. J., & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Upper Saddle River, NJ: Prentice-Hall.
Seco, N., Veale, T., & Hayes, J. (2004). An intrinsic information content metric for semantic similarity in Wordnet. In R. L. de Mántaras & L. Saitta (Eds.), *Proceedings of the 16th European Conference on Artificial Intelligence, ECAI 2004* (pp. 1089–1090). Amsterdam, The Netherlands: IOS Press.
Shannon, C. E., & Weaver, W. (1964). *The mathematical theory of communication*. Champaign, IL: University of Illinois Press.
Snyder, A., Mitchell, D., Bossomaier, T., & Pallier, G. (2004). The Creativity Quotient, an objective scoring of ideational fluency. *Creativity Research Journal*, *16*, 415–420.
Snyder, A., Mitchell, D., Ellwood, S., Yates, A., & Pallier, G. (2004). Nonconscious idea generation. *Psychological Reports*, *94*, 1325–1330.
Snyder, A. W., Laughlin, S. B., & Stavanga, A. G. (1977). Information capacity of eyes. *Vision Research*, *17*, 1163–1175.
Stroustrup, B. (1986). *The C++ programming language*. Boston: Addison-Wesley Longman.
Thurstone, L. (1924). *The nature of intelligence*. London: Kegan Paul, Trench, Trubner & Co.
Ullman, S. (1970). *Semantics: An introduction to the science of meaning*. Oxford, UK: Blackwell.
Wall, L., & Schwartz, R. L. (1991). *Programming Perl*. Sebastopol, CA: O'Reilly & Associates.
Ward, W. C. (1969). Rate and uniqueness in children's creative responding. *Child Development*, *40*, 869–878.
Watts, D. (1999). *Small worlds*. Princeton, NJ: Princeton University Press.

APPENDIX

The philosophy behind the CQ metric is one of estimating the number of possible ideas that might be created from a given set of responses, analogous to the number of pictures that can be created in the photoreceptor mosaic of the eye, given contamination by noise. An estimate of this number leads to a description length or information capacity of the response set (Snyder, Laughlin, & Stavanga, 1977; A. W. Snyder, Bossomaier, & Hughes, 1986).

To make this more precise necessitates some statistical model. The simplest approach is to consider each response as being drawn from a binary distribution where the response either occurs or does not occur, and to assume that both are equally likely.

If the responses were completely independent, each response would be in a separate category and the total number of possible response possibilities would be 2^N for N responses, or an information capacity of N. But the responses are not independent. Thus the information capacity falls below N and the quantity one needs to measure it is, according to Shannon (1964), the joint entropy of all the response sets, $H(x_i)$, where x_i are the response variables. To assess the independence requires an external model of the world and the relationship of concepts within it.

There are now two stages to the argument. First, we show that this joint entropy is the approximated by the CQ value Q. Then we derive the information theory measure. The last section describes formally the multi-information.

For a set of independent categories, X_c,

$$H(\{x_i\}) = \sum_c H(X_c) \tag{5}$$

because the joint probability distribution breaks up into a product of that for each category. Within the category, the maximum entropy is achieved when all the elements are independent (linked only by the properties of the category), i.e.,

$$H(\{x_i\}) = \sum_c H(X_c) = \sum_c N_c = Q \tag{6}$$

with N_c the number of responses in each category.

I_m the concept mutuality (the multi-information approximated by the pairwise sum) is the normalized mutual information between concepts. It is defined by

$$I_m = \sum_i H(x_i) - H(\{x_i\}) \tag{7}$$

But $\sum_i H(x_i)$ is simply given by N, the total number of responses. Thus, after rearranging

$$Q = N - I_m \tag{8}$$

Now, consider the two limits of just one category and N categories. In the first case, $Q = N$ exactly as for the direct measurement of CQ. In the second case, the mutual information between two concepts asymptotically approaches 1 as the concepts become identical. Thus, an upper bound to I_m must be $N - \log_2(N+1)$ where the correction term $\log_2(N+1)$ provides just enough information to label N distinct objects. Substituting this value for I_m gives $Q = \log_2(N)$, the CQ value for a single category response set.

The Determination of Multi-information

Given a set of N variables $\{x_i\}$, $i = 1, 2, \ldots, N$ where each variable is a vector $x_i = [x_i(1), \ldots, x_i(M_i)]$. Each element in vector x_i represents one of M_i discrete events, the vector itself is the sample space and the marginal probability distribution over x_i is $P_i(x_i)$, the joint distribution over two variables, $P_{i,j}(x_i, x_j)$ and the joint distribution over all variables $P(\{x_i\})$. In this case, each variable is a concept and events are occurrences of the concept in different contexts (such as bodies of text).

The entropy for a multivariate distribution of N variables is defined as (?, ?):

$$S(\{x_i\}) = -\sum_{\{x_i\}} P(\{x_i\})\log(P(\{x_i\})) \tag{9}$$

The multi-information is defined as (?, ?, ?):

$$I(\{x_i\}) = \sum_{x_i} S(x_i) - S(\{x_i\}) \tag{10}$$

$$= \sum_{\{x_i\}} P(\{x_i\})\log\left[\frac{P(\{x_i\})}{\prod_j P_j(x_j)}\right] \tag{11}$$

In the case where $\{x_i\} = \{x_1, x_2\}$ we recover the definition of mutual information (?, ?):

$$I(x_1, x_2) = \sum_{x_1,x_2} P_{1,2}(x_1, x_2)\log\left[\frac{P_{1,2}(x_1, x_2)}{P_1(x_1)P_2(x_2)}\right] \tag{12}$$

This article uses Seco's (2004) concept similarity metric on Wordnet as a proxy for $I(x_i, x_j)$.