Assessing Scientific Creativity:

Conceptual Analyses of Assessment Complexities

Dean Keith Simonton

University of California, Davis

Citation:

Simonton, D. K. (2012). Assessing scientific creativity: Conceptual analyses of assessment complexities. Commissioned paper, The Science of Science and Innovation Policy Conference, National Academy of Sciences.

Assessing Scientific Creativity:

Conceptual Analyses of Assessment Complexities

Too often creativity is seen as a relatively simple phenomenon. That is, researchers will speak of creative thought as if it were a single cognitive process located in a person's head (Simonton and Damian, in Press). This simplification is immediately apparent whenever someone refers to the creative *process* rather than *processes*—using the singular rather than the plural. Oversimplification is also evident when creativity is treated solely as a psychological phenomenon, ignoring the disciplinary context in which creativity most often emerges (Csikszentmihályi, 1999). Furthermore, the tendency toward oversimplification is frequently accentuated when discussing *scientific* creativity. In theory, creative individuals in the sciences enjoy special access to the "scientific method" that somehow does all of the real work. Hypotheses follow naturally and logically from prevailing theories and empirical findings, and testing those hypotheses entail the application of well-established techniques that were often mastered in graduate school if not before (e.g., high school science labs). Better yet, because science is postulated to be "objective" rather than "subjective," a scientist's individual creativity is assumed equivalent to how fellow scientists will judge his or her creativity. A fact is a fact, and logic is logic.

To be sure, empirical research suggests that scientific creativity must be more complex than this idealized view (Feist, 2006; Simonton, 2004). Philosophers of science would also object to this simplistic conception (McGrew, Alspector-Kelly, and Alihoff, 2009; but see Simon, 1973). Nonetheless, I argue here that even rather basic conceptual analyses of scientific creativity must reveal essential complexities. These properties help us appreciate why scientific creativity is by no means easy. There is good reason why a very large percentage of scientific publications are never cited in subsequent research (Redner, 1998). There is even more justification for believing that breakthrough discoveries must be extremely rare. Exceptional creativity is necessarily a scarce commodity.

My conceptual analysis consists of three main parts. First, I will provide a definition of the *personal* creativity associated with a given scientific idea, and then work out the consequences of this definition. Second, I offer a definition of *consensual* creativity assigned to the same idea, and then derive further consequences by comparing this definition with the previous definition. Third, I discuss how the foregoing conceptual analyses radically understate the actual complexities. I close with a discussion concerning some implications regarding both empirical research and science policy. In this last section I attempt to move from the abstract to the concrete.

PERSONAL SCIENTIFIC CREATIVITY

To simplify the analysis, I concentrate on the single scientist trying to evaluate an idea's creativity. By an "idea," I just mean any conjecture, hypothesis, explanation, or solution with respect to a scientific question or problem. Note, too, that this point I will ignore the collaborative nature of much scientific creativity, especially in the natural sciences.

Definition

Because researchers often disagree in their definitions of creativity (Plucker, Beghetto, and Dow, 2004; Runco and Jaeger, 2012), I have recently argued for a quantitative, multiplicative, and three-criterion definition (Simonton, 2011a, 2012a, 2012c). Unlike the most common definitions, this one acknowledges the real complexity of the phenomenon. To be specific, the *personal* creativity of a given scientific idea should be defined as follows:

$$c = (1 - p)u(1 - v), \text{ where } 0 \leq c \leq 1 \qquad (1)$$

Here c is a decimal fraction ranging from $c = 0$, for no creativity whatsoever, to $c = 1$, for the highest possible creativity. A useful way of conceiving c is that it represents the subjective probability that the individual scientist will deem that idea creative. Naturally, this definition has no meaning without also defining the three parameters, namely, p, u, and v. Here are their definitions:

1. The parameter p is the *initial probability* that the idea will be generated by the given scientist, where $0 \leq p \leq 1$. The term "probability" is here used in the psychological sense of "response strength" rather than in the abstract mathematical ("frequentist") sense. If $p = 0$, then the idea is not immediately available, but is presumed to be accessible after an incubation period requiring a suitable priming stimulus or stimuli (Hélie and Ron, 2010; Seifert, Meyer, Davidson, Patalano, and Yaniv, 1995). In any event, the factor $(1 - p)$ in Equation 1 then represents the idea's initial *originality* in that scientist's mind. High probability ideas have low originality.

2. The utility parameter u is the *final probability* that the scientist will eventually judge it useful, where $0 \leq u \leq 1$. Although an idea's utility is potentially a continuous variable, in many instances u reduces to a dichotomous 0-1 variable. For instance, when James Watson endeavored to discover the DNA code given the four nucleic acids, he used molecular models to inspect four possible arrangements, but only one set of these pairings (viz. adenine-

thymine and guanine-cytocine) yielded a perfect utility value (Watson, 1968). The others had zero usefulness, including his preferred "like-with-like" solution.

3. The parameter v is the scientist's *prior knowledge* of u, where again $0 \leq v \leq 1$. If $v = 0$, then the scientist is ignorant of whether or not the idea will be useful without first conducting a generation and test of that idea, but when $v = 1$ the individual already knows the value of u without the need for confirmation. The idea is obvious. In authentic algorithmic problem solving, for instance, $v = 1$, whereas in heuristic problem solving $v \ll 1$ (cf. Amabile, 1996; Simonton, 2011b). Whenever the v falls somewhere between 0 and 1, the idea constitutes a "hunch" likely based on tacit knowledge that cannot yet be articulated (the specific value indicating variable "feeling of knowing" states; cf. Bowers, Regehr, Balthazard, and Parker, 1990). When prior knowledge falls in this middle range, discovering that the idea is useful can still elicit some degree of surprise. In fact, we can take $(1 - v)$ as a direct measure of an idea's subjective *surprisingness*.

Equation 1 might look exotic, but it actually embodies a formal version of comparable three-criterion definitions. For example, Boden (2004) specified that a creative idea must be novel, valuable, and surprising. Likewise, to acquire legal protection the US Patent Office requires that an invention be novel, useful, and nonobvious (http://www.uspto.gov/inventors/patents.jsp). Finally, the above definition parallels Amabile's (1996) assertion that "a product or response will be judged as creative to the extent that" it is (a) novel, (b) "appropriate, useful, correct or valuable response to the task at hand," and (c) "the task is heuristic rather than algorithmic" (p. 35). Notice that the phrase "to the extent that" suggests that creativity and its defining components are quantitative rather than qualitative attributes of an idea, a suggestion also captured in Equation 1.

Consequences

Equation 1 might look simple, but its implications are definitely not so. These repercussions can be grouped under two heads: (a) multiplicative rather than additive integration and (b) applying three criteria rather than two.

Multiplicative Rather than Additive Integration

Equation 1 combines the three parameters by multiplying them together to yield a single number representing the scientist's assessment of an idea's creativity. That is, creativity is the joint product of originality, utility, and surprisingness. Given that the three factors $(1 - p)$, u, and $(1 - v)$ are all positive decimal fractions, it only takes one of these factors to equal 0 for $c = 0$. Put differently, each criterion exerts veto power over the others.

An idea cannot be creative if it has a probability of unity or has zero utility or is completely obvious. Thus, if I decide to make a basic wheel, I will not consider myself creative because I know that this highly useful invention has already been invented. In a different fashion, if I designed a safe made completely of soap bubbles, I would not consider it creative because the safe could not really protect—or even hold—its contents. Finally, if I solved an ordinary second-order differential equation using the quadratic formula, I would not consider the result creative because I know beforehand that the formula is guaranteed to provide me with the correct roots even if I have never seen them before. By comparison, if we averaged the three factors—which represents a guise of additive integration—then the reinvented wheel, the soap bubble safe, and the algorithmic solution to the differential equation would still be creative, albeit not perfectly so (see Simonton, in Press, for relevant Monte Carlo simulations). For instance, the soap bubble safe would get points for originality no matter how useless—so useless that it cannot keep anything safe!

The multiplicative integration of the three factors has another implication that is less obvious but no less crucial to our understanding of scientific creativity. Except when the values of all three factors are unity, their product will be much smaller, and can even approach zero. To give a numerical example, if $(1 - p) = 1$, $u = 1$, and $(1 - v) = 1$, then $c = 1$. But, if $(1 - p) = 0.9$, $u = 0.9$, and $(1 - v) = 0.9$, then $c \approx 0.7$, and if $(1 - p) = 0.5$, $u = 0.5$, and $(1 - v) = 0.5$, then $c \approx 0.1$. Hence, highly creative ideas are far rarer than are highly improbable, highly useful, or highly surprising ideas taken separately. If we looked at a representative sample of scientific ideas and separately assessed them on p, u, v, and c, we would find the distribution of c to exhibit a conspicuous positive skew even when the distributions of p, u, and v are symmetrical (which is unlikely anyway). Again, this repercussion does not appear were we to average the three values, in which case highly creative ideas would be far more commonplace. In the above example, we get cs of 1, 0.9, and 0.5, respectively, where the last two values are much larger than in the multiplicative integration. Indeed, because of the Central Limit Theorem (Simonton, 2008), the distribution of c in a large sample of ideas would closely approximate the normal distribution (Simonton, in Press).

Indirect support can be found for the multiplicative integration at the level of the individual scientist: Self-citations exhibit extremely skewed distributions just like citations in general (Bartneck and Kokkelmans, 2011; Davarpanah and Amel, 2009). Hence, from the scientist's personal perspective, highly creative ideas must be rather rare, making up a small proportion of his or her total ideational output. The bulk of the time a scientist generates ideas that are publishable but neither exciting nor citable.

Applying Three Criteria Rather than Two

Many creativity researchers use a two-criterion definition that omits the third criterion (Runco and Jaeger, 2012). That is, a creative idea must only be (a) novel or original and (b) useful, appropriate, or valuable. Even so, I have argued that this definition does not capture the full complexity of the phenomenon (Simonton, 2012c; see also Huber, 2000). Instead, it is necessary to add the third criterion, whether we call it surprising (Boden, 2004), nonobvious (http://www.uspto.gov/inventors/patents.jsp), or heuristic rather than algorithmic (Amabile, 1996). This third criterion also echoes Perkin's (2000) contrast between *reasonable* problems that "can be reasoned out step by step to home in on the solutions" and *unreasonable* problems that "do not lend themselves to step-by-step thinking. One has to sneak up on them" (p. 22). Unreasonable problems are inherently more creative than reasonable problems because they require an "aha!" or eureka experience. In classic Gestalt theory, such problems demand a drastic restructuring of the problem, an insight that views the problem in an utterly different manner (Köhler, 1925; Wertheimer, 1945/1982).

Adding this third criterion has several important consequences. The first and perhaps most obvious is that it makes highly creative ideas even more extraordinary. Because the factor $(1 - v)$ in Equation 1 will be a small-valued decimal fraction, the resulting product yielding c will be rendered appreciably smaller than had this factor been omitted. Of necessity, scientific ideas that are original, useful, and surprising will always be less common than ideas that are just original and useful. It is for this reason that post hoc explanations are always inferior to theory-based predictions even when the explanations are otherwise identical. A historic example is the contrast between Lorentz–Fitzgerald contraction formula and the exact same formula derived from Einstein's special theory of relativity. Where the former required some often implausible ad hoc hypotheses, the latter followed directly from the relativistic nature of space and time under moving (but non-accelerating) frames of reference.

A second implication of adding the third criterion is the more restricted role it implies for the operation of domain-specific expertise (Simonton, 2012c). This repercussion is made explicit in how the United States Patent Office applies the "nonobvious" criterion. Obviousness is not gauged by a naive person but rather by someone who has "ordinary skill in the art" (http://www.uspto.gov/web/offices/pac/mpep/documents/2100_2141_03.htm; see also Sawyer, 2008). In other words, if anybody with the requisite domain-specific expertise could generate the same idea by a step-by-step, even algorithmic method, then it cannot receive patent protection. The invention must be a true innovation, not a mere adaptation (cf. Kirton, 1976; Weber, 1992). Once more, this contrast is witnessed in the difference between the Lorentz-Fitzgerald contraction and the predictions of the special theory of relativity. The former was an ad hoc adaptation of classical physics whereas the latter was a revolutionary innovation of what was to become a core foundation of modern physics. Speaking more generally, major advances in science entail the creation of a new expertise rather than the application of an already existing expertise (Simonton, 2012b). Other well-known examples include Galileo's of telescopic astronomy, Leeuwenhoek's microscopic biology, and Mendel's quantitative genetics. None of these innovations could have been anticipated by contemporary experts.

If a highly creative scientific idea cannot come directly from domain-specific expertise, then where do creative ideas come from? The answer is suggested in Campbell's (1960) blind-variation and selective-retention (BVSR) theory of creativity and "other knowledge processes" (p. 380). Although now over a half-century old, BVSR has been recently updated with more empirical and mathematical treatments (e.g., Simonton, 2010, 2011a, 2011b, 2012a). According to the modernized version, the "sightedness" of an idea is defined using the same parameters used to define its creativity. Specifically, sightedness $s = puv$, where $0 \leq s \leq 1$ (cf. Simonton, 2012a). In words, an idea's sightedness is the joint product of its initial probability, ultimate utility, and prior knowledge of that utility: Highly sighted ideas are highly probable, highly useful, and highly obvious. Because blindness is just the opposite of sightedness, the former is readily defined as $b = 1 - s$. Given these definitions, it logically follows that as $s \rightarrow 1$ (or as $b \rightarrow 0$), $c \rightarrow 0$, that is, highly sighted ideas cannot be highly creative. In addition, as $s \rightarrow 1$, BVSR becomes increasingly unnecessary. Trial-and-error is not required when the ideas are routine, reproductive, or algorithmic. BVSR comes into play only when the utilities are initially unknown.

One might easily commit the fallacy of turning this inference around by asserting that for any idea it holds that as $s \rightarrow 0$ (or as $b \rightarrow 1$), $c \rightarrow 1$, abut this conclusion is as necessarily false as the previous conclusion is necessarily true. After all, whenever $u = 0$, then $s = c = 0$. The idea is totally useless, but the scientist does not know that without first engaging in a generation and test. Expressed differently, whereas highly sighted ideas cannot be highly creative, highly blind ideas can be either creative or noncreative or any level between (i.e., the expected variance in ideational creativity increases with blindness). Blindness is largely a measure of the scientist's initial ignorance regarding the actual utility values. Only when a blind idea turns out to be highly useful will it then be deemed highly creative. Blindness is a necessary but not sufficient condition for creativity.

CONSENSUAL SCIENTIFIC CREATIVITY

The preceding analysis dealt with "little-c creativity" as distinguished from "Big-C Creativity" (Csikszentmihályi, 1998; cf. Kaufman and Beghetto, 2009; Luckenbach, 1986). In particular, p, u, v, and c all represent the scientist's subjective judgment of an idea's relevant properties. Although little-c creativity is no doubt psychologically important to the scientist who came up with an idea—not the least when it results from a eureka experience—it may be far less important to other scientists, who might decide to ignore the idea altogether. In contrast, if scientific colleagues cannot possibly ignore the idea but instead incorporate it in their own work, then the idea can be said to display Big-C Creativity. This segregation closely parallels Boden's (2004) distinction between "P-creative" and "H-creative" ideas—the psychologically creative versus the historically creative. The only real difference is that while little-c and P-creative ideas are largely equivalent, only the "biggest" of the Big-C ideas are also H-creative. When an idea not only attracts numerous citations in the scientific literature but also earns its creator the Nobel Prize, then the transition has been fully executed.

But it is now time to define Big-C creativity and work out the consequences of that definition.

Definition

A scientist's creative idea must be evaluated by colleagues in the same field before it can be considered creative in the larger sense (Csikszentmihályi, 1999; Simonton, 2010). These colleagues (a) provide feedback on drafts of unpublished papers, (b) serve as peer reviewers to appraise grant proposals and submitted manuscripts, (c) write extramural letters during appointment and promotion actions, (d) choose whether or not to mention a given idea in their own publications, and (e) take part in the selection of those colleagues most deserving of scientific awards and honors. Empirical research suggests that $250 \leq n \leq 600$, where n denotes the size of the field (Wray, 2010; cf. Price, 1963). Because n is thus very large, I can assume that the scientist who creates the idea is included in the field even though that creator more precisely has only n - 1 colleagues. This assumption simplifies the forthcoming definition without inserting any serious alterations in their implications.

An idea's *consensual* creativity can now be defined as

$$C = 1/n \, \Sigma \, c_i, \text{ where } i = 1, 2, 3, ..., n \text{ and } 0 \leq C \leq 1 \qquad (2)$$

In words, Big-C creativity is the average of the separate creativity assessments by all n members of the field, including the assessment of the scientist who originated the idea. However, like any statistical mean, the individual scores can vary around the central tendency. It is accordingly necessary to define the variance $\sigma^2 = 1/n \, \Sigma \, (c_i - C)^2$, where $0 \leq \sigma^2 \leq 1$. This variance is as significant as the mean in understanding the relation between little-c and Big-C creativities. That significance comes from the fact that fields vary greatly in the consensus they exhibit regarding contributions made their disciplines (Cole, 1983; Klavans and Boyack, 2009; Simonton, 2009). Thus, physical sciences display higher consensus than do the biological sciences, which in their turn show higher consensus than found in the social sciences. High- versus low-consensus fields also differ on many other noteworthy features (Simonton, 2004b). For instance, in high-consensus fields (a) knowledge at the research frontier becomes obsolete at a faster rate (McDowell, 1982), (b) researchers are less prone to confirm their hypotheses (Fanelli, 2010), (c) basic concepts tend to be defined more precisely (Schachter, Christenfeld, Ravina, and Bilous, 1991), (d) research is more likely to be anticipated by other members in the field (Hagstrom, 1974), (e) undergraduate training is characterized by longer chains of prerequisite courses (Ashar and Shapiro, 1990), and (f) laws have a higher relative frequency relative to theories in introductory textbooks (Roeckelein, 1997).

In any case, according to my conceptual analysis, *high-consensus* fields are those in which $\sigma^2 \approx 0$ whereas *low-consensus* fields are those in which $\sigma^2 \approx 1$. Where field-member evaluation variance is minimal in the former, the variance is maximal in the latter. In most scientific disciplines, the level of consensus likely falls somewhere between these two extremes. In fact, a field with maximal evaluation variance could hardly be called a "field" because the consensus would be perfectly absent. Everybody contributing ideas to the discipline just does his or her "own thing" and then self-identifies it as creative. The latter event is more characteristic of some avant-garde arts than any bona fide science.

Consequences

The complexities of scientific creativity become most conspicuous when little-c and Big-C creativity are directly compared. In these comparisons I assume that scientist $i = 1$ was the idea's originator, so that the contrast of interest is really between c_1 and C. Once more, because the field is sufficiently large, it does not matter that c_1 was among the n values used in computing C. Any bias in the part-whole comparisons is not only minimal, but also the bias progressively reduces to zero as n increases or as σ^2 decreases.

The comparisons are easiest in high-consensus fields because we can then confidently affirm that $C \approx c_1$. The person $i = 1$ and the other n - 1 members of the field will strongly agree in their assessments. This ideal outcome should be the norm in high-consensus fields like physics where citations tend to concentrate on a relatively small number of journal articles as well as a select group of authors (Cole, 1983). Nothing more needs to be said about this ideal case.

In stark contrast, if scientist i =1 contributes an idea in a low-consensus field, then C and c_1 may not necessarily match, and when they do not match they may differ from each other in varying degrees. Consider the following three cases:

Case 1: If $C > c_1$, then the scientist yielded an idea that had a bigger impact on the field than he or she could have anticipated. The idea was objectively more original, more useful, and/or more surprising than subjectively appreciated. From a personal perspective, this would be a very pleasant outcome. At the same time, this case might be the rarest of the three because self-deception alone would probably induce a more favorable evaluation. That said, sometimes in the history of science new scientific ideas were kept secret rather than being widely and quickly disseminated. In such periods, a single individual might discover some new algorithmic technique that provided a general solution to a particular class of problems. Once the discovery was made, the solution looked mundane and uncreative at the personal level but strikingly original and surprising at the consensual level. A classic example is Niccolò Tartaglia's 16th-century solutions to cubic equations that enabled him to win a mathematical competition without having to engage in any serious problem solving. Not until Gerolamo Cardano revealed Tartaglia's secret formula did the unsurprising nature of his solutions become well known.

Case 2: If $C < c_1$, then what the target scientist thinks is highly original, useful, and surprising is much less so to his or her colleagues. Even highly eminent scientists can sometimes produce work that they personally believe is underappreciated by other members of the field (Arkin, 2011). The lack of a strong consensus makes it too easy for an idea to be viewed as more creative by its originator than by other field members.

Case 3: If $C \approx c_1$, Big-C and little-c assessments of the idea's creativity remain roughly equivalent because the scientist i = 1 is now somehow "typical" of others in the field. If the n collective creativity evaluations tend to cluster around a modal value, then this near equivalence would still be somewhat common even in fields that are not high consensus. The central tendency could still lie in the middle of the distribution and about two thirds of the creativity judgments might remain within a standard deviation (i.e., σ, the square root of the variance).

A vital lesson from the foregoing analysis is that an idea's personal little-c creativity (c_1) for the scientist and its consensual Big-C creativity (C) for the scientist's field will seldom coincide except under two very special circumstances. First, the scientist i = 1 must contribute the idea to a high-consensus field. Second, the scientist must count as a "typical" representative of the colleagues composing a low-consensus field. The latter circumstance would be less common than the former because of the expectation of appreciable dispersion around the central tendency of the distribution. As the evaluation variance becomes sufficiently large, that is, as $\sigma^2 \rightarrow 1$, then a very large proportion of the field would arrive at the value $c_i = 0$ ($i \neq 1$), the lower bound for the creativity appraisal, even if c_1 were large. The individual's great creative idea would then fall well short of universal acclaim among fellow members of the field. Indeed, in low consensus fields with a large n, it is not impossible for $C \approx 0$ while $c_1 \approx 1$. Scientist $i = 1$ is a statistical outlier whose idea has almost no takers whatsoever. Alfred Wegener's continental drift theory may well illustrate an idea having this lowly disciplinary status shortly after its publication.

NECESSARY COMPLICATIONS

The preceding definitions and their implications establish that the assessment of the creativity of any scientific idea is by no means simple. And yet, my conceptual analysis greatly oversimplifies the phenomenon! Consider the following four potential complications.

First, at the very outset I arbitrarily separated personal and consensual creativity assessments. The individual scientist first comes up with an idea judged personally creative and then he or she presents that idea to fellow field members to determine the consensual assessment of its creativity. This isolation of the personal from the consensual completely ignores the obvious fact that much scientific creativity occurs in collaborative groups, especially in research laboratories (see, e.g., Dunbar, 1995). Thus, even if the idea might have emerged in the head of a single scientist, that idea might be subjected to immediate assessment during the next lab meeting. Only after the idea survives that initial collective appraisal may it eventually be offered to the field as a whole for a broader evaluation. In this case, the issue is not just whether the scientist is representative of the field but also whether his or her research group is representative. Consequently, the assessments must operate at three levels rather than two—individual scientist, collaborative group, and disciplinary field. Three-level analyses are always more complex than two-level treatments.

Second, when defining consensual creativity, I focused solely on the means and variances of the creativity measure, ignoring its three components. Nevertheless, it is obvious that originality, utility, and surprisingness have their separate means and variances, adding an additional six statistics to the conceptual analysis. These additional statistics introduce new complications in the individual-versus-field assessments, complications that intensify all the more in low-consensus scientific fields. For instance, the field and individual might disagree not only about the assessed creativity of a given idea but also disagree about the basis of the disagreement. In other words, the two sets of assessments might differ regarding originality, utility, and surprisingness as well as the overall creativity. This lack of correspondence is aggravated by the fact that $C = 1/n\ \Sigma\ c_i \neq (1 - P)U\ (1 - V)$, where $P = 1/n\ \Sigma\ p_i$, $U = 1/n\ \Sigma\ u_i$, and $V = 1/n\ \Sigma\ v_i$. That is, the average of the cs across all n members of the field is not usually equal to the multiplicative product of the averages of the three factors.

Third, Equation 1 implicitly applies the same weights to the three criteria. Even so, criteria may be more important than others are. For example, the equation might be revised as

$$c = (1 - p)u(1 - v)w_{(1-p)}w_u w_{(1-v)},$$

where the three appended decimal-fraction weights allow for the possibility that in scientific creativity utility is more important than originality or surprise. Yet if the weights are permitted to differ for personal creativity, they might also be allowed to differ for consensual creativity. Members of the scientist's field could very well assign different weights than does the individual. For instance, surprisingness might be undervalued owing to the hindsight bias (Fischhoff, 2007; see also Simonton, 2012b). In retrospect, it is often far easier to see an idea as obvious when it was far less so at the time of its origination.

Fourth and last, the personal and consensual assessments defined so far have been presumed to be static rather than dynamic. At a given time, the scientist decides to offer a creative idea for field evaluation. Within a relatively short interval, the idea's creativity becomes fixed at both personal and consensual levels. Yet it is clear that the assessments might change over the years, a change that would be particularly conspicuous for the field, which has the luxury of reevaluating its judgments long after the scientist has passed away. A clear illustration is Wegener's theory, a theory that did not become mainstream science until long after his death. In any case, such changes can be accommodated by including appropriate subscripts. Thus, the consensual creativity of continental drift can be defined as C_t, where $t = 0$ when the idea was first published, and $C_t \approx 0$ then as well, even though $c_1 \approx 1$ for Wegener himself.

I had better stop here. My conceptual analyses of the complexities in the assessment of scientific creativity may have already become too complex for proper evaluation. Still, just as a thought experiment, imagine how complicated must be an exhaustive analysis that incorporates the above four complications! Only after a complete treatment will we ever understand why making highly creative contributions to any scientific discipline is very, very hard. Highly creative scientific ideas are not only rare at the individual level, but are even more so at the field level, and particularly in low-consensus disciplines. Is there any wonder why scientific genius is so exceptional?

FURTHER DISCUSSION

I will be the first to admit that the forgoing conceptual analyses were highly abstract. Indeed, it was deliberately so. I strongly believe that the empirical research on scientific creativity is riddled with confusion precisely because researchers do not agree on what they mean by saying some idea, person, or product is “creative.” Too often concepts that should be kept distinct are conflated into a single concept that thereby becomes more inclusive, but at the same time becomes more ambiguous and even self-contradictory. For example, I would argue that the common use of “novel” as the first criterion of creativity actually combines originality with surprise, two subjective constructs that can often be separated even if they often go together. An interesting example in the history of technology is the Pelton water wheel, a device invented by two independent inventors, only one of whom sought patent protection (Constant, 1978). Although the inventions were identical, and thus equally original and useful, one inventor had a less obvious starting point, one that required a serendipitous event to get to the goal, and thus found his creation more surprising than did the other inventor.

Even worse is the commonplace confusion of personal and consensual creativity. For instance, self-report measures of creativity will often ask respondents to record achievements that represent a mixture of little-c and Big-C creativity, where the former is just assumed without demonstration to dovetail into the latter (e.g., Carson, Peterson, and Higgins, 2005). The implicit assumption is that if personal creativity is high enough, then it meshes with the lower levels of consensual creativity. This assumption cannot be true in either (a) high consensus disciplines where the two judgments will be highly convergent or (b) low consensus disciplines where the two judgments will be highly divergent, with a huge gap between the two.

These conceptual confusions have likely consequences. For example, two separate and extensive literature reviews reveal that the cognitive neurosciences have failed to make any significant headway toward understanding creativity (Dietrich and Kanso, 2010; Sawyer, 2011). Yet can it really be otherwise? If different researchers use distinct measures that do not agree on what is actually being measured, how can they be observing the same processes in the first place? In simple terms, the brain of a little-c creator involved in everyday creativity may not operate the same as the brain as a Big-C creator engaged in historic creativity. As seen in my formal analyses, the latter is not produced from the former simply by increasing the font size.

Anyhow, I would like to end this paper by getting more concrete. I will accomplish this two ways. First, I briefly discuss how the conceptual analysis presented here might influence empirical research on scientific creativity. Second, I apply the analysis to an important issue in science policy, namely, how to optimize the effectiveness of peer review.

Empirical Research

Even though I have treated all of the personal parameters as logical entities, they obviously can be subjected to empirical analysis. Certainly active researchers, during the course of their investigations, can be asked to rate their ideas (as they emerge) on originality, utility, surprisingness, and overall creativity. Such ratings can be used to determine the actual weights assigned to the three criteria. In addition, these ratings can be continued intermittently to learn how these subjective judgments change over time. Although it would be more difficult, consensual creativity can also be studied via surveys to see how the assessments operate at the field level. Do colleagues assign different weights to the criteria in making judgments of the creativity of scientific ideas? What are the correlations between personal and consensual assessments and how do those correlations differ across high- versus low-consensus fields? To carry out studies of this nature, it would probably be better to shift the unit of analysis from the idea to the product. For example, empirical investigations submitted for publication could be easily examined this way, the author's self-assessments compared with the referees' evaluations.

Naturally, it would be equally valuable to determine how these personal and consensual judgments relate to the paper's actual impact, as determined by citations (see, e.g., Shadish, 1989). Surprisingly, relatively little research has been devoted to assessing the predictors of citation impact at the product level (for review, see Simonton, 2002, chap. 5). The bulk of the literature has focused on a scientist's impact aggregated across all publications (Ruscio, Seaman, D'Oriano, Stremlo, and Mahalchik, in Press). With this change level of analysis comes a change in predictors, such as identifying the personality and developmental variables associated with producing high-impact research programs (e.g., Feist, 1993, 1994; Helmreich, Spence, Beane, Lucker, and Matthews, 1980; Rodgers and Maranto, 1989; Simonton, 1992, 2000). Such work, while certainly important, tells us virtually nothing about why specific journal articles exert a truly extraordinary influence on the field—the so-called "citation classics" of the discipline. This research also does not help us understand why even great scientists can produce research that leaves little or no impression on their field.

Science Policy

I have demonstrated that the assessment of Big-C creativity in science is necessarily more precarious in low-consensus disciplines. In the latter, many ideas that are deemed highly creative at the personal level will not be so received at the consensual level. Hence, "neglected genius" will become increasingly common as $\sigma^2 \rightarrow 1$. Mendel in biology and Wegener in geology provide prototypical examples. In contrast, in fields like physics and chemistry, such neglect is not very common, and when it does occur, it does not last very long. Einstein may have had to wait until 1921 to receive the Nobel Prize for work published in 1905, when he was 26 years old, but he was already recognized as a rising star in theoretical physics before his 30th year (Hoffmann, 1972). That modest temporal delay is a far cry from having to wait for posthumous recognition, as was the case for Mendel and Wegener. In a sense, low-consensus fields are less efficient than high-consensus fields, an inefficiency that shows up in the slower rates by which creative scientists can contribute to the domain's accumulation of knowledge.

Needless to say, the contrast between high- and low-consensus fields becomes especially problematic when it comes to peer review. Without reliably positive peer evaluations, creative scientists cannot dependably publish in top-tier journals, win necessary grant funding from private and public foundations, obtain appointments and promotions in prestigious research institutions, or win early recognition by means of professional honors and awards. In addition, scientists in low-consensus fields will often find themselves at a disadvantage relative to scientists in high-consensus fields. The former will find their careers more capricious than the latter, with a more uncomfortable disjunction between self-assessed merit and its corresponding disciplinary acknowledgement. For example, in low-consensus fields, the peer review process may be just "a little better than a dice role" (Lindsey, 1988; see also Cicchetti, 1991; Cole, Cole, and Simon, 1981; Marsh and Ball, 1989). Indeed, if research depended on measures with reliabilities as low as often found for peer reviewing in even in top-tier journals, then that would research would be deemed unpublishable in those same journals. (see, e.g., Scott, 1974). Moreover, the disadvantages operate at both individual and field levels. The "soft" sciences certainly enjoy less esteem than do the "hard" sciences (Smith, Best, Stubbs, Johnston, and Archibald, 2000). This differential in scientific status is reflected in the Nobel Prizes, which practically ignore contributions outside of the natural sciences (the award for economics being a separate "add on" that was not part of the original bequest). Accordingly, the most creative scientist working in a social science will rate lower in scientific acclaim than will the most creative scientist working in a natural science.

The question then arises: What can be done to increase the consensus in low-consensus fields and thereby improve the effectiveness of peer review in such disciplines? To address this issue properly, it is first mandatory to identify the source of the disagreements among different members of the same field. Three main sources are possible.

First, the scientists who create in low-consensus fields might have different personal characteristics that make them less likely to agree to some disciplinary consensus than holds for scientists creating in high-consensus fields. For example, in Roe's (1953) study of 64 eminent scientists, the social scientists (psychologists and anthropologists) were found to be less factual, more emotional, and more rebellious than the physical scientists (physicists and chemists). Similarly, Chambers (1964) found that creative psychologists were more bohemian and unconventional than creative chemists (see also Cattell and Drevdahl, 1955). Based on these and other findings, it has been argued that creators in the social sciences have personalities that fall somewhere between those in the natural sciences and those in the arts and humanities, a placement that corresponds to the comparative amount consensus typical of these different disciplines (Simonton, 2009). Artists appear particularly indisposed toward conformity (Feist, 1998). This difference between scientists and artists is apparent in the highly personal, even idiosyncratic nature of much artistic creativity. As a result, multiple discovery and invention is far less common in the arts than holds in the sciences (Simonton, 2010). Scientific multiples presume a consensus on the central problems and methods to solve those problems.

Second, it could be that education and training is less systematic and standardized in the low-consensus fields. For example, I mentioned earlier how high-consensus fields feature undergraduate programs with longer chains of prerequisites (Ashar and Shapiro, 1990). In most social science majors, the only requirement for taking an upper-division course is often nothing more than a general substantive course and (perhaps) an introductory methods course. Furthermore, except for some loosely defined "breadth requirement," most of the upper-division courses have the effective status of "electives." In striking contrast, in most natural science majors, each course has precise prerequisites that require that courses be taken in a long series lasting a year or more. Moreover, the number of true electives is much smaller, meaning that graduates in the major have taken pretty much the same core courses even at the upper-division level, a shared educational foundation that would immensely facilitate disciplinary consensus. To illustrate from my own university and discipline, it is possible for two students to graduate with a bachelors in psychology despite having taken only one upper-division course in common—a single one-quarter introduction to psychobiology class! Otherwise, a graduate's transcript can lack courses that should be considered basic to psychological understanding, such as learning, memory, abnormal, or social.

Third and last, it is conceivable that the phenomena that define the primary focus of scientific inquiry vary immensely in complexity or ambiguity and that this variation directly influences both (a) the type of person attracted to the field and (b) the intellectual structure of the domain itself. If true, agreement among peer reviewers cannot be improved simply by "personnel selection" and "curriculum development." In particular, the social sciences cannot display higher consensus simply by attracting more would-be physicists with lower tolerance of ambiguity and by devising more standardized curricula with longer chains of prerequisite courses and fewer electives. In fact, attempts to do so might render the resulting discipline less comprehensive and more simplistic. In support of this conjecture, Skinnerian behaviorism once displayed a level of consensus more typical of the hard sciences (Cole, 1983), yet many psychologists would argue that behaviorism only accomplished this feat by ignoring huge number of important cognitive and neurological phenomena (Gardner, 1987). The idea that the different sciences exhibit inherent differences in complexity actually harks back to Auguste Comte (1839-42/1855), whose hierarchy of astronomy, physics, chemistry, biology, and sociology represented an historical progression, each successive science depending on the development of the preceding science. Although he may have got the details wrong, Comte probably captured an important truth (cf. Cole, 1983). A fully developed biology depends on chemistry, just as a fully developed chemistry depends on physics. Moreover, each successive science introduces new complexities—so-called "emergent properties"—that cannot be reduced to the prior science. With simplifying reductionism ruled out of court, the low-consensus disciplines will likely always remain so relative to the high-consensus disciplines.

If the third explanation is the most correct, then it is still possible to improve peer review. A cardinal principle of psychometric theory is that the reliability of a test increases with the number of items making up that test. Stated more generally, increasing the number of measurements allows the "random errors" to cancel out, shrinking the error bars on the combined assessment. Hence, if the goal is to obtain the same judgmental consensus in the social sciences as found in the natural sciences, a greater number of independent evaluators are necessary to attain the same benchmark. Preferably, too, the evaluators should represent the heterogeneity in the field. Unsurprisingly, this solution is sometimes adopted, however inadvertently or intuitively. More referees may be asked to evaluate a manuscript submitted to a psychology journal than are requested to judge a manuscript submitted to a physics journal. This accommodation is far superior to simply writing off the low-consensus sciences as too “soft” to warrant scientific support and recognition. After all, the theoretical and empirical questions addressed in low-consensus fields are just as worthy of answers as the questions raised in high-consensus fields. Science inquiry should never shy away from phenomena simply because they are complex and difficult.

REFERENCES

Amabile, T.M. (1996). *Creativity in context: Update to the social psychology of creativity*. Boulder, CO: Westview.

Arkin, R.M. (2011). *Most underappreciated: 50 prominent social psychologists describe their most unloved work*. New York: Oxford University Press.

Ashar, H., and Shapiro, J.Z. (1990). Are retrenchment decisions rational? The role of information in times of budgetary stress. *Journal of Higher Education, 61,* 123–141.

Bartneck, C., and Kokkelmans, S. (2011). Detecting h-index manipulation through self-citation analysis. *Scientometrics*, 87, 85-98.

Boden, M.A. (2004). *The creative mind: Myths and mechanisms* (2nd ed.). New York: Routledge.

Bowers, K.S., Regehr, G. and Balthazard, C. (1990). Intuition in the context of discovery. *Cognitive Psychology*, *22*, 72-110.

Campbell, D.T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review*, *67*, 380-400.

Carson, S., Peterson, J.B., and Higgins, D.M. (2005). Reliability, validity, and factor structure of the Creative Achievement Questionnaire. *Creativity Research Journal*, *17*, 37-50.

Cattell, R.B., and Drevdahl, J.E. (1955). A comparison of the personality profile (16 P. F.) of eminent researchers with that of eminent teachers and administrators, and of the general population. *British Journal of Psychology*, *46*, 248-261.

Chambers, J.A. (1964). Relating personality and biographical factors to scientific creativity. *Psychological Monographs: General and Applied*, *78* (7, Whole No. 584).

Cicchetti, D.V. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences*, *14*, 119-186.

Cole, S. (1983). The hierarchy of the sciences? *American Journal of Sociology*, *89*, 111-139.

Cole, S., Cole, J.R., and Simon, G. A. (1981, November 20). Chance and consensus in peer review. *Science*, *214*, 881-886.

Comte, A. (1855). *The positive philosophy of Auguste Comte* (H. Martineau, Trans.). New York: Blanchard. (Original work published 1839-1842)

Constant, E.W., II (1978). On the diversity of co-evolution of technological multiples: Steam turbines and Pelton water wheels. *Social Studies of Science, 8*, 183-210.

Csikszentmihályi, M. (1998). Reflections on the field. *Roeper Review: A Journal on Gifted Education*, *21*, 80-81.

Csikszentmihályi, M. (1999). Implications of a systems perspective for the study of creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 313-338).Cambridge: Cambridge University Press.

Davarpanah, M.R., and Amel, F. (2009). Author self-citation pattern in science. *Library Review*, *58*, 301-309.

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

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R.J. Sternberg and J.E. Davidson (Eds.), *The nature of insight* (pp. 365-396). Cambridge, MA: MIT Press.

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS ONE 5*(4): e10068. doi:10.1371/journal.pone.0010068.

Feist, G. J. (1993). A structural model of scientific eminence. *Psychological Science, 4*, 366-371.

Feist, G. J. (1994). Personality and working style predictors of integrative complexity: A study of scientists' thinking about research and teaching. *Journal of Personality and Social Psychology*, *67*, 474-484.

Feist, G. J. (2006). *The psychology of science and the origins of the scientific mind*. New Haven, CT: Yale University Press.

Fischhoff, B. (2007). An early history of hindsight research. *Social Cognition*, *25*, 10-13.

Gardner, H. (1987). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.

Hagstrom, W.O. (1974). Competition in science. *American Sociological Review*, *39*, 1-18.

Hélie, S., and Ron, S. (2010). Incubation, insight, and creative problem solving: A unified theory and a connectionist model. *Psychological Review*, *117*, 994-1024.

Helmreich, R.L., Spence, J.T., Beane, W.E., Lucker, G.W., and Matthews, K.A. (1980). Making it in academic psychology: Demographic and personality correlates of attainment. *Journal of Personality and Social Psychology*, *39*, 896-908.

Hoffmann, B. (1972). *Albert Einstein: Creator and rebel*. New York: Plume.

Huber, J.C. (2000). A statistical analysis of special cases of creativity. *Journal of Creative Behavior*, *34*, 203-225.

Kaufman, J.C., and Beghetto, R.A. (2009). Beyond big and little: The four c model of creativity. *Review of General Psychology*, *13*, 1-13.

Kirton, M.J. (1976). Adaptors and innovators: A description and measure. *Journal of Applied Psychology*, *61*, 622-629

Klavans, R., and Boyack, K.W. (2009). Toward a consensus map of science. *Journal of the American Society for Information Science and Technology*, *60*, 455-476.

Köhler, W. (1925). *The mentality of apes* (E. Winter, Trans.). New York: Harcourt, Brace.

Lindsey, D. (1988). Assessing precision in the manuscript review process: A little better than a dice roll. *Scientometrics*, *14*, 75-82.

Luckenbach, T.A. (1986). Encouraging "little C" and "Big C" creativity. *Research Management, 29*, 9–10.

Marsh, H.W., and Ball, S. (1989). The peer review process used to evaluate manuscripts submitted to academic journals: Interjudgmental reliability. *Journal of Experimental Education*, *57*, 151-169.

McDowell, J.M. (1982). Obsolescence of knowledge and career publication profiles: Some evidence of differences among fields in costs of interrupted careers. *American Economic Review*, *72*, 752-768.

McGrew, T., Alspector-Kelly, M., and Allhoff, F. (Eds.). (2009). *Philosophy of science: An historical anthology*. Oxford, UK: Wiley-Blackwell.

Perkins, D.N. (2000). *The eureka effect: The art and logic of breakthrough thinking*. New York: Norton

Plucker, J.A., Beghetto, R.A., and Dow, G.T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist*, *39*, 83-96.

Price, D. (1963). *Little science, big science*. New York: Columbia University Press.

Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B*, *4*, 131-134.

Rodgers, R.C., and Maranto, C.L. (1989). Causal models of publishing productivity in psychology. *Journal of Applied Psychology*, *74*, 636-649.

Roe, A. (1953). *The making of a scientist*. New York: Dodd, Mead.

Roeckelein, J.E. (1997). Psychology among the sciences: Comparisons of numbers of theories and laws cited in textbooks. *Psychological Reports*, *80*, 131-141.

Runco, M., and Jaeger, G.J. (2012). The standard definition of creativity. *Creativity Research Journal*, *21*, 92-96.

Ruscio, J., Seaman, F., D'Oriano, C., Stremlo, E., and Mahalchik, K. (in Press). Measuring scholarly impact using modern citation-based indices. *Measurement: Interdisciplinary Research and Perspectives*.

Sawyer, R. K. (2008). Creativity, innovation, and nonobviousness. *Lewis & Clark Law Review*, *12*, 461-487.

Sawyer, R. K. (2011). The cognitive neuroscience of creativity: A critical review. *Creativity Research Journal*, *23*, 137-154.

Schachter, S., Christenfeld, N., Ravina, B., and Bilous, R. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, *60*, 362-367.

Scott, W.A. (1974). Interreferee agreement on some characteristics of manuscripts submitted to the *Journal of Personality and Social Psychology*. *American Psychologist*, *29*, 698-702.

Seifert, C.M., Meyer, D.E., Davidson, N., Patalano, A.L., and Yaniv, I. (1995). Demystification of cognitive insight: Opportunistic assimilation and the prepared-mind perspective. In R.J. Sternberg and J.E. Davidson (Eds.), *The nature of insight* (pp. 65-124). Cambridge, MA: MIT Press.

Shadish, W.R., Jr. (1989). The perception and evaluation of quality in science. In B. Gholson, W. R. Shadish, Jr., R.A. Neimeyer, and A C. Houts (Eds.), *The psychology of science: Contributions to metascience* (pp. 383-426). Cambridge: Cambridge University Press.

Simon, H.A. (1973). Does scientific discovery have a logic? *Philosophy of Science, 40*, 471-480.

Simonton, D.K. (1992). Leaders of American psychology, 1879-1967: Career development, creative output, and professional achievement. *Journal of Personality and Social Psychology*, *62*, 5-17.

Simonton, D.K. (2000). Methodological and theoretical orientation and the long-term disciplinary impact of 54 eminent psychologists. *Review of General Psychology*, *4*, 13-24.

Simonton, D.K. (2002). *Great psychologists and their times: Scientific insights into psychology's history*. Washington, DC: American Psychological Association.

Simonton, D.K. (2003). Scientific creativity as constrained stochastic behavior: The integration of product, process, and person perspectives. *Psychological Bulletin*, *129*, 475-494.

Simonton, D.K. (2004a). *Creativity in science: Chance, logic, genius, and zeitgeist*. Cambridge, England: Cambridge University Press.

Simonton, D.K. (2004b). Psychology's status as a scientific discipline: Its empirical placement within an implicit hierarchy of the sciences. *Review of General Psychology*, *8*, 59-67.

Simonton, D.K. (2008). Distribution, normal. In W. A. Darity, Jr. (Ed.), *International encyclopedia of the social sciences* (2nd ed., Vol. 2, 415-417). Detroit: Macmillan Reference USA.

Simonton, D.K. (2009). Varieties of (scientific) creativity: A hierarchical model of disposition, development, and achievement. *Perspectives on Psychological Science*, *4*, 441-452.

Simonton, D.K. (2010). Creativity as blind-variation and selective-retention: Constrained combinatorial models of exceptional creativity. *Physics of Life Reviews*, *7*, 156-179.

Simonton, D.K. (2011a). Creativity and discovery as blind variation and selective retention: Multiple-variant definitions and blind-sighted integration. *Psychology of Aesthetics, Creativity, and the Arts*, *5*, 222-228.

Simonton, D.K. (2011b). Creativity and discovery as *blind* variation: Campbell's (1960) BVSR model after the half-century mark. *Review of General Psychology*, *15*, 158-174.

Simonton, D.K. (2012a). Creativity, problem solving, and solution set sightedness: Radically reformulating BVSR. *Journal of Creative Behavior*, *46*, 48–65.

Simonton, D.K. (2012b). Foresight, insight, oversight, and hindsight in scientific discovery: How sighted were Galileo's telescopic sightings? *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication. doi: 10.1037/a0027058

Simonton, D.K. (2012c). Taking the US Patent Office creativity criteria seriously: A quantitative three-criterion definition and its implications. *Creativity Research Journal*, *24*, 97-106.

Simonton, D.K. (in Press). Combinatorial creativity and sightedness: Monte Carlo simulations using three-criterion definitions. *International Journal of Creativity and Problem Solving*.

Simonton, D.K., and Damian, R.I. (in Press). Creativity. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology*. New York: Oxford University Press.

Smith, L.D., Best, L.A., Stubbs, D.A., Johnston, J., and Archibald, A.B. (2000). Scientific graphs and the hierarchy of the sciences. *Social Studies of Science*, *30*, 73-94.

Watson, J.D. (1968). *The double helix: A personal account of the discovery of the structure of DNA*. New York: Atheneum.

Weber, R.J. (1992). *Forks, phonographs, and hot air balloons: A field guide to inventive thinking*. New York: Oxford University Press.

Wertheimer, M. (1982). *Productive thinking* (M. Wertheimer, Ed.). Chicago: University of Chicago Press. (Original work published 1945)

Wray, K. B. (2010). Rethinking the size of scientific specialties: correcting Price's estimate. *Scientometrics*, *83*, 471-476.