Social Epistemology
A Journal of Knowledge, Culture and Policy

ISSN: 0269-1728 (Print) 1464-5297 (Online) Journal homepage: https://www.tandfonline.com/loi/tsep20

Transformative science: a new index and the impact of non-funding, private funding, and public funding

Barrett R. Anderson & Gregory J. Feist

To cite this article: Barrett R. Anderson & Gregory J. Feist (2017) Transformative science: a new index and the impact of non-funding, private funding, and public funding, Social Epistemology, 31:2, 130-151, DOI: 10.1080/02691728.2016.1241321

To link to this article: https://doi.org/10.1080/02691728.2016.1241321

Published online: 21 Feb 2017.

SOCIAL EPISTEMOLOGY, 2017
VOL. 31, NO. 2, 130–151
http://dx.doi.org/10.1080/02691728.2016.1241321

Transformative science: a new index and the impact of non-funding, private funding, and public funding

Barrett R. Anderson[a,b] and Gregory J. Feist[a]

[a]Department of Psychology, San Jose State University, San Jose, CA, USA; [b]Department of Psychology, University of California, Santa Cruz, Santa Cruz, CA, USA

ABSTRACT
Understanding how impactful scientific articles were funded informs future funding decisions. The structural significance of articles is broken down into two submeasures: citation count and "generativity" (a novel measure defined as being highly cited and also leading to a comparatively large number of other highly cited work). Generativity is an attempt to provide a quantitative operationalization of transformativeness, a concept often used as a funding criterion despite not being a well-defined construct. This report identifies highly impactful and generative publications indexed in the subject area of psychology in the Web of Science in the year 2002. Publications that reported funding sources were found to be more generative than those that did not, and research that was privately funded was found to be more generative than publicly funded research. This analysis is exploratory, and hopefully contributes to a foundation for future empirical investigations into the structure and nature of transformative science that granting agencies would want to fund.

KEYWORDS
Bibliometrics; impact factor; metascience; psychology of science; citation analysis; scientometric

The question of transformative science – what is it and how do we know it in advance? – is central to not only the history of science but also to its future. Granting agencies exist to fund the best future science; science that ideally will change the field. As important as transformative science is, how it's defined and what criteria are used to assess it, are almost complete unknowns. The current study is an attempt to address these questions both theoretically and with a proposed novel form of assessing transformative science. But first, what is "transformative science?"

Transformative science

We propose that a historical and structural view provides one answer to the question of what is a "transformation" in science. The tree of knowledge provides the structure. A common metaphor of human knowledge co-opts the form of a great tree. The roots of the tree reach deep in the past, and its branches grow up into the distant future. As the tree grows up through time, the trunk splits into various branches, each of which defines a new field of knowledge. Rising up into the tree, each of these branches divides again and again. At first, these changes are easy to follow: natural science splitting off from philosophy, further divisions defining the early boundaries of physics, geology, astronomy, and biology. But as times goes on the complexity increases. Sometimes branches go nowhere (phrenology, astrology), sometimes they are very fruitful (natural selection, relativity), and sometimes a branch that has long been dormant begins to grow again (naturalistic decision-making). Branches that have for some time grown apart from each other may begin to grow together again in an unexpected way

CONTACT Barrett R. Anderson barander@ucsc.edu

(astrobiology, behavioral economics, psychobiotics). This complex, fruitful, and many-splendored Tree of Knowledge describes the history of science.

The tree also describes an ongoing conversation, where ideas combine and build on those that came before them. The history of science is no less a history of the individual personalities that contributed to it, but in a way that may be unique among human endeavors it is possible in science to separate the thought from the thinker. It is equally valid to describe the history of science as a history of ideas.[1] From this perspective, the body of the Tree of Knowledge is composed of various books, monographs, notes, theses, dissertations, articles, discussions, symposia, conversations, websites, and emails – all of the physical artifacts and ephemeral moments that the life of an idea will flow through. Karl Popper (1972) referred to these products of human knowledge as the "third (ontological) world."

Metasciences and cladistics

The whole of the tree is too much to take in at a glance. Any hope of understanding even a small portion of its structure requires a systematic approach. Depending on the specifics, such an approach might be part of one of four metasciences – the history, philosophy, psychology and sociology of science (e.g. Feist 2006; Gholson et al. 1989; Kuhn 1962; Merton 1973). Any study of the physical or electronic artifacts that form the body of the tree is a branch of bibilometrics or scientometrics. Recently these fields have also gone by the names informetrics, webometrics, or cybermetrics (Andrés 2009; De Bellis 2009). These names evidence the increasing technological complexity of scientific communication, but it would be a mistake to read this variation as reflecting a change in the fundamental subject of study. This subject, the transmission and measurement of scientific knowledge, remains the same.

Drawing from the analogical relationship between the Tree of Knowledge and the biological Tree of Life, a tree that describes the evolutionary relationships between species, the effort to characterize the structure of the tree can be described metaphorically as a form of cladistic analysis (Rieppel 2010). Cladistics is a method of classification that divides organisms into groups based on common ancestry, called clades. These clades are the branches of the Tree of Life, and a cladogram is a diagrammatic illustration of these relationships. By analogy, a cladistic analysis of the Tree of Knowledge would consider the transmission of concepts through communication rather than the transmission of genes through species.[2] An example of a cladogram of a small portion of the Tree of Knowledge is provided in Figure 1. Such an analysis will be beyond the scope of the present study, but the cladistic model provides the appropriate context in which to consider measures of structural significance, such as scientific transformations. These measures will allow us to identify those important nodes that either begin new branches or transform existing ones. Put another way, these measures allow us to identify those nodes that significantly impact the structure of the tree.

Defining "Transformative"A National Science Foundation (NSF) workshop on the meaning and implications of transformative research took place in March 2012 (Frodeman and Holbrook 2012). Indeed, inspired by a call from US Science Advisor John H. Marburger III, an entire new funding programme began in 2006 at NSF – Science of Science and Innovation Policy (SciSIP) – whose charge it was to fund science and innovation policy research. Such research aims to deliver empirical information to policy makers (e.g. politicians, funding agencies, and administrative scientists) in their effort to make more efficient and informative decisions about funding science, especially transformative and innovative science. Moreover, transformational research was added to the NSF merit review criteria in 2009, but similar concepts (research that is potentially transformative, high-risk, innovative, or that might in the most favorable cases lead to discoveries that extend to other fields of science) have been identified as important funding criteria for at least the last quarter century.

The definition of transformative research has generally been vague (to the point that defining the term was identified as a goal in the H.R. 5116–111th Congress: America COMPETES Reauthorization Act of 2010) but always implies the intensification of change in science. We do not think it would be

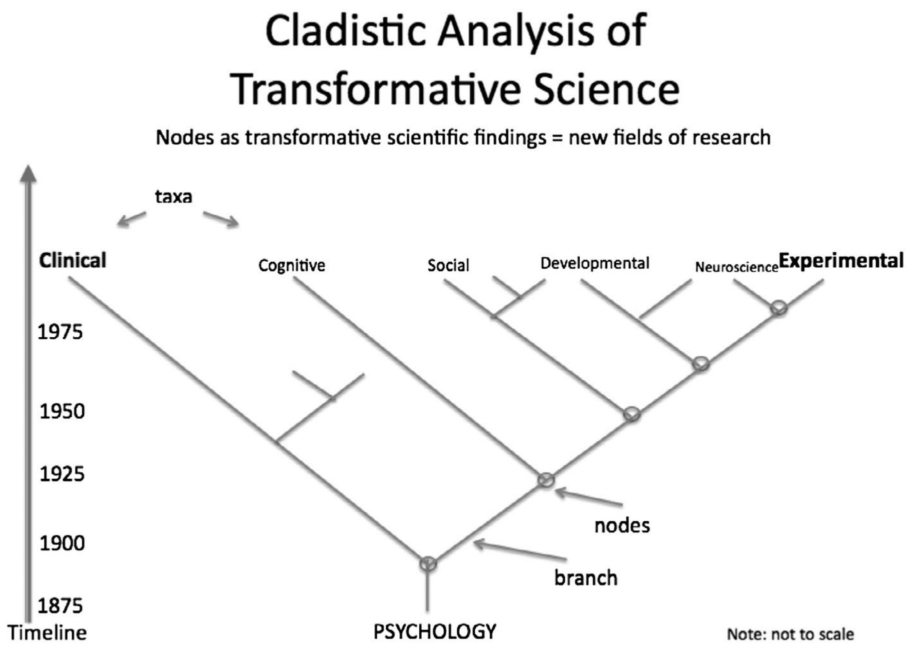

Figure 1. Example cladogram of psychology from 1875 to present.

controversial to contend that work that starts a new branch of science, or that fundamentally changes an existing one, would be considered transformative.

Transformativeness as a funding criterion was originally inspired by the concept of revolutionary science from Thomas Kuhn's *Structure of Scientific Revolutions* (1962), which discussed the role of paradigm shifts in scientific progress. In Kuhn's model, anomalies that emerge in the course of normal science eventually lead to a crisis, which can only be resolved by revolutionary science. Revolutionary science defines a new paradigm that incorporates the anomalies and provides a whole new set of questions for normal science to ask. According to Kuhn, this revolutionary science is a necessary consequence of the buildup of anomalies from normal "puzzle-solving" science. Using Kuhn's definition of revolutionary science, anything that promotes normal science also promotes transformative science. One cannot selectively promote transformative science, in Kuhn's model, but his definition is not the only one possible. There are other ways to conceptualize transformative science (as a disruptive innovation, or on a continuum with normal science), and some of these other perspectives imply it is possible to take a more interventionist role in its promotion.

Generativity

Regardless of the specifics of the definition, research that is transformative must necessarily be highly cited. No matter how potentially transformative a work might be in isolation, that actual transformation has to occur within the social activity of science. Scientists collaborate, forming teams throughout the process of designing experiments, conducting research, and presenting their findings. They constantly evaluate each other's work at conferences, in peer-reviewed papers, and in grant applications. All of these interactions provide the context for a scientific culture. To be influential, a potentially transformative idea has to successfully travel through this culture and take hold in the minds of the scientists who are participating in it.

It is our belief that many of the articles that cite a work of transformative science will also be highly cited themselves. The transformation of an entire field is more than a single event, and we suspect that research that is transformative will also be highly generative. Although it may be that not all generative research will be transformative, we hope that a new measure of generativity will provide a good approximation for objectively quantifying transformativeness. We are proceeding in the present study under the assumption that this will be true, with the caveat that at the time of writing a validation of generativity has not been conducted. Such a validation would require resources beyond those currently available.

Structural significance

The purpose of the present study is to identify research that has been structurally significant in the Tree of Knowledge (i.e. transformative and created new branches), and to describe how this research is being funded. Examining the funding of science in the recent past will give us a sense of how diligent we have been in our custodianship of the tree, with a special focus on those transformative moments of creativity in which new branches appear. Many believe that federal funding agencies, for example, are too conservative and tend to reject proposals that are too "out of the box", some of which go on to transform scientific knowledge. We intend to examine this claim in the current study. Understanding how science has been funded can help inform and improve future funding decisions. The impact of these decisions is broader than just on those who desire a good return on their investment in science – it also includes every person who lives in a world that can be transformed by the next big idea. Discussions that will lead to better choices about the near future of science necessarily begin with an understanding of the recent past, and these conversations should take place in as empirically grounded a context as possible.

Identifying structurally significant work is a substantial challenge. Even an expert may not be able to immediately identify important work without the benefit of historical context. For reasons of familiarity, and to keep the scope of the study in check, we focus on a small section of the recent past in the field of psychology. This also has the advantage of avoiding the effect of different standards across different disciplines (Bornmann et al. 2008). The window of time – in this case chosen to be 2002 – should not be so far back that the decisions that were made are no longer relevant to those being made today. Yet, the time frame should also not be so close to the present that the available data are too inconsistent or incomplete. Although this would appear to argue for only considering older work that already has a well-established place in the history of science, that advantage has to be weighed against the benefit of providing more current information. Presumably, information about work that is closer to the present day would be more relevant and useful to a contemporary decision-maker. For this reason, we will choose to rely on imperfect metrics to provide us with something akin to a first draft of the history of the funding transformative science.

Research that focuses on the value of science, and especially on creative productivity, tends to use metrics based on individual publications – the *least publishable unit* in science (Simonton 2004). Hence, the analysis is generally at the level of the individual scientist. In some cases the metric rises to the level of journal, institution, or even nation, especially among sociologists of science. The present study will remain focused on the level of individual publications. Starting at this lowest possible level avoids unnecessary computational complexity. This decision simplifies data collection and analysis. More importantly, the lower level of complexity prevents unnecessary confusion, providing the most straightforward example of the novel measure.

But we focus on more than just the individual publication. We also focus on publications that generate other publications, that is, they are structurally significant to the Tree of Knowledge. These publications must be impactful or generative. Information about references and citations will be necessary to operationalize these measures of structural significance, and that information is both less ambiguous and more easily traceable at the level of individual publications. A description of both types of structural significance under consideration follows:

(1) *Impactful publications* are those that have received a large number of citations. Many researchers built on the ideas that impactful publications communicated.

(2) A *generative publication* is one that leads to a new branching point in the Tree of Knowledge (see Figure 1). Identifying this specific structural impact requires a broader view than the individual publication. The simplest description of a generative publication has two requirements, (a) that the publication is itself highly cited, and (b) that a large number of those publications that cite the original publications are also themselves highly cited.

For purposes of this investigation we will be looking at the most structurally significant publications in the field of psychology in the year 2002. Specifically, we will be looking at publications that are more structurally significant than their peers, defining peers as other publications in the same field, in the same year. This focus on peer publications is important because the number of researchers varies between fields, as well as across time (Garfield 2006; Radicchi, Fortunato, and Castellano 2008). It is possible that even with our sample limited to a single field in a single year, more populated subfields will be overwhelmingly represented simply due to a greater number of publications. If it becomes clear that this is the case, then a more finely grained distinction between subfields will be called for, and any analysis will require further subdivision or some form of normalization.

Research questions

In the process of reviewing the most structurally significant publications for information regarding their funding sources, it is possible that several comparisons will present themselves. Two research questions are anticipated:

(1) First, is research that reports its funding source more likely to be structurally significant than research that does not? There may not always be a straightforward relationship between funding and quality, but it would be surprising to find anything other than an overall positive effect of support. Ideally, this comparison would be between funded and unfunded publications, but the funding status of publications that do not report their funding is necessarily ambiguous. Presumably any publications that do not report their funding sources, but are structurally significant, are worthy of further attention.

(2) Second, is privately funded research more likely to be structurally significant than publicly funded research? It may be that highly structurally significant science (both highly impactful and highly generative) will be less likely to be funded by federal sources than science with a medium structural significance but more likely than science with a low structural significance. That is, there may be a curvilinear relationship between structural significance and federal funding, with scientific research with a medium structural significance being more likely to be federally funded, compared to research with a high and low structural significance. Within the NIH, transformative research has been identified as "high risk, high reward research" (Austin 2008), although there is some dispute about whether those terms should be synonymous (Frodeman and Holbrook 2012).

Method

Design

The design of this study is an archival one, in which the published literature in the scientific databases was coded on two characteristics: structural significance and source of funding. Structural significance is broken down into two quantitative submeasures, times cited (impact) and generativity. Each article was coded for source of funding in three ways: funded vs. unfunded; public vs. private funding entity; and if funded, name of funding agency. During coding, an additional category for funding sources was added: domestic (U.S.) vs. international. These codings provide categorical independent variables. The design of the investigation is between subjects ANOVA, with "subjects" being research articles from different categories. The dependent variables are times cited and generativity, both of which are continuous. When it is necessary in our analysis to distinguish between the higher-level categories of funding sources, the public vs. private axis will be labeled Sector and the domestic (U.S.) vs. international axis will be labeled National Origin.

Procedure

Thomson ISI Web of Science has been the traditional source for citation data (Harzing and van der Wal 2008; Norris and Oppenheim 2007). Other potentially useful sources have emerged recently (Meho and Yang 2007), the most notable of which is Google Scholar. Although Google Scholar has several advantages, including free availability, high speed, and broad scope, it is in some ways less useful and less transparent than Thomson ISI. Google Scholar does not provide (a) the ability to sort results by citation count (b) the ability to export results, or (c) an application programming interface (API) which would allow a researcher to easily develop solutions to the previous limitations. Google Scholar also does not provide information about how its database is put together. Although this is an understandable omission for a proprietary tool, it makes it less useful for this type of study.

Other newer options, such as Altmetrics and Academia.edu, take a fundamentally different approach to measuring impact, placing additional weight on online interactions. Although many powerful analyses can take advantage of this new type of scientometric data (Bollen, Van de Sompel, Hagberg, Bettencourt, et al., 2009), neither of these options provides another source of citation data. The data collection portion of the study consisted of three phases:

Phase one consisted of collecting the top 10% (by citation count) of the records in the Thomson ISI Web of Science that match predetermined criteria. These four criteria are language (English), publication type (peer-reviewed article), date of publication (2002), and subject area (psychology[3]). This search resulted in 1774 records. Following this, we selected a sample consisting of one half of the top 10% of the entire collection (887 articles). To create this sample we sorted the records by citation count, randomly selected odds or evens (by coin flip), and included every other article from (and including) the starting point. Our intent here was to select a random sample in which the distribution of citation counts very closely or exactly matched the distribution of citation counts in the top 10%.

In *phase two*, we assigned each of the publications selected in the first phase two structural significance scores, namely impact and generativity. Impact is simply the raw citation count, which was already included in all records collected from the database. Generativity required more effort and was only assigned to records in the sample. Generative articles are those papers that (a) are highly cited (first order), and that (b) incite a next generation of research that itself becomes highly cited (second order). More concretely, generativity is a count of the number of high impact articles that cite a given high impact article. The steps to calculate a generativity score are outlined in Figure 2 and are:

(1) In the first step, a high impact threshold was defined. For the purpose of this measure, high impact articles were defined as any article in the top 10% by citation count of articles published in the same language, the same year and the same field (defined by Web of Science category).
(2) The second step was identifying those first order articles that are above the threshold defined in the first step. All of the first order articles (i.e. articles in the sample) necessarily met this threshold. Importantly, this means that only high impact articles (identified as A_1 and A_2 in the figure) will have any generativity score at all.
(3) The third step was to define a high impact threshold for the second order articles (the citing articles). In this case the peers are not the articles in the initial sample, but other articles that were published in the same language, year, and field. It is important here to note that the 88,691 s-level articles ranged across 147 of the 250 Web of Science Categories, and in many cases more than one category applied to a given article. Although conceptually an ideal generativity score would include thresholds for all 147 categories, in practice this proved impractical. Fortunately, restricting the analysis to categories that individually accounted for

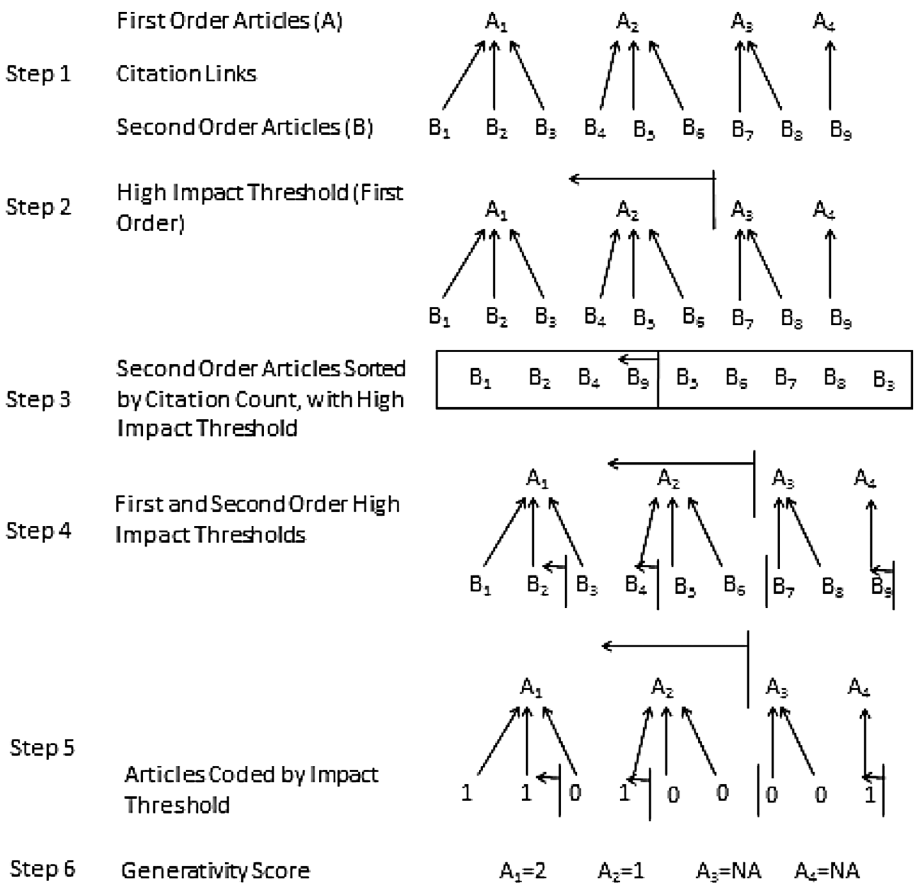

Figure 2. Steps to calculate generativity score.

at least 1% of the sample identified 13 categories (See Table 1) that together accounted for 80.98% of all second-order articles in the sample. (An initial draft version of the generativity score, generated only from articles in the Psychology category, correlated with the final combined generativity score based on all 13 categories, $r = .91$, $p < .001$.)

(4) The fourth step was identifying those second order articles that were above the thresholds defined in the third step.

(5) The fifth step was to convert second order articles to numerical values. Any article that was identified as above the threshold in the previous step (for any applicable category) should be counted as a one; any article below the threshold (for all applicable categories) can be counted as a zero.

(6) Finally, the numerical values from the previous step are summed for each article, resulting in a positive integer for each high impact article in the sample. This is the generativity score.

To provide a concrete example (with invented values), we will begin with the article A_1. We will assume that A_1 has 268 citations in the Web of Science. A_1 is in our sample and therefore is a first order article. Each of those 268 articles that cite A_1, and all of the other articles that cite articles that are in our sample, are second order articles ($B_1 – B_{max}$). We will assume that for the field of psychology in the year 2002 in the Web of Science that the articles in the top 10% by citation count have at least 50 citations. Since A_1 has a number of citations equal to or greater than 50, it does have a generativity score. Next, we generate thresholds based on all of the second order articles (this will need to be per year and

Table 1. generativity citation count thresholds for second-level articles.

Web of science category	Year 2002	2003	2004	2005	2006	2007	2008	2009
Behavioral Sciences	54	55	48	43	36	32	25	19
Business	48	44	41	36	29	23	17	12
Economics	48	44	41	36	29	23	17	12
Education & Educational Research	28	26	24	23	20	16	12	10
Family Studies	38	36	38	31	28	23	16	11
Neurosciences	74	69	64	58	48	42	33	25
Pediatrics	42	39	37	33	30	24	19	15
Pharmacology & Pharmacy	50	66	44	40	35	30	25	19
Psychiatry	74	69	66	58	48	42	32	23
Psychology	54	54	50	42	36	30	23	22
Public, Environmental & Occupational Health	52	51	46	41	35	29	23	16
Rehabilitation	38	34	32	31	26	22	17	13
Substance Abuse	52	47	49	40	33	30	23	17

per Web of Science category). The generativity score is the number of those 268 s-order articles that have citation counts above the appropriate threshold. Of the 268 articles that cite A_1 16 are in the top 10% of articles in their year and in at least one of the categories that they belong to. Therefore A_1 has a generativity score of 16.

As previously mentioned, generativity scores apply only to high impact articles. The case of a low impact article that is cited by a high impact article might be a case of latent potential, but it is also possible that the initial article was of only auxiliary utility (See Figure 3). Articles are cited for a variety of reasons (Bornmann and Daniel 2008), and not all citations are created equal.

In *phase three* each publication in the sample that was collected in phase one was briefly reviewed. This review served to identify whether a funding source had been reported, and to record the identity of that source. Funding information was gathered from the article itself. Individual funding sources were categorized as public, if they were a government funded agency, or private, if not. During this process a second category of interest emerged, domestic (U.S.) and international funding sources. Each funding source was also categorized on this criterion.

Results

Descriptive statistics of sample

The following figures characterize the entire sample, the top 10% of English-language articles published in Psychology in 2002 and indexed in the Web of Science. The sample contains 1774 articles from 265 journals. The top 10 journals by count of articles accounted for about a third of the sample (28.07%). More than half of the articles (50.45%) were from the top 30 journals.

Out of the half of the sample reviewed for funding source (887 articles), 290 (32.69%) did not list any funding source. Considering only those articles that did list funding sources, 63.71% listed a single source and 95.89% list 3 or fewer (see Table 2). Funding sources that accounted for more than one half of one percent of all funding sources listed are listed in Table 3. In total, this accounts for slightly more than one half (56.22%) of all funding sources. The NIH, including those organizations that operate under it, accounted for 29.92% of the total. Individual articles with more than one funding source are in some cases funded by a mix of public and private, or domestic and international sources (see Figures 3 and 4).

Data preparation

The highly skewed nature of citation data necessitated performing a log transformation before conducting inferential tests (see Figures 5 and 6). Following convention, base 10 was chosen because it is effective for normalizing skewed distributions of continuous numerical data (Osborne 2002). Visual inspection indicates that normalization of Generativity was successful (see Figures 6 and 7), whereas

		First Order Impact	
Second Order Impact		Highly Cited	Not HighlyCited
	Highly Cited	Transformative Science	Latent PotentialOR Auxiliary Contribution
	Not Highly Cited	False Start	Ordinary Science

Figure 3. Categories of impactful and generative science.

Table 2. Number and of funding sources per article.

Number of funding sources	Count of articles	Percentage
1	402	63.71
2	161	25.52
3	42	6.66
4	16	2.54
5	7	1.11
6	3	0.48
Sum	631	100.00

normalization of Times Cited was more questionable (see Figures 8 and 9). The raw values for Times Cited and for Generativity were strongly and positively correlated ($r = 0.87$, $p < 0.001$), as were their log transformations, Times Cited log 10 (TC log 10) and Generativity log 10 (G log 10)($r = 0.69$, $p < 0.001$; see Table 4 and Figures 10 and 11).

Generativity and journal ranking

Because Generativity is defined at the level of individual articles, it is possible to create a derivative measure at a higher level, such as researcher or journal. Simplified examples of such a ranking system, based on mean Generativity (see Table 5), or on the percentage[4] of Generative articles (see Table 6), are provided. It is important to note (1) that these rankings are based only on those journals that included at least one generative article, and (2) that the rankings are not weighted based on the number of articles published in each journal. A table of 2002 psychology journals ranked by impact factor (see Table 7) is included for comparison.

Relationship between generativity index and traditional impact metrics

To examine the relationship between generativity and traditional impact metrics, we conducted a regression analysis with generativity (log 10 transformed) as the dependent variable and mean citation count and impact factor as the predictor variables. This analysis was restricted to the top 20 ranked journals (by each measure). Out of the top 20 ranked journals by impact factor and by generativity, four overlapped. For ranking by impact factor and percentage of generative articles, six overlapped. Only the *Journal of Cognitive Neuroscience* overlapped across all three criteria. Within this sample generativity was significantly predicted by mean citation count, $b = -0.474$, $t(44) = 3.93$, $p < 0.001$, and impact factor, $b = -0.324$, $t(44) = -2.69$, $p < 0.01$. These variables accounted for approximately 35% of the variance in generativity, Adjusted $R^2 = 0.347$, $F(2,44) = 13.21$, $p < 0.001$. Interestingly, this indicates that more the cited papers in more obscure journals tended to be more generative.

Planned comparisons

Although the current research is exploratory, our inferential analysis was guided by two research questions: First, is research that reports its funding source more likely to be structurally significant than research that does not? Second, is privately funded research more likely to be structurally significant than publicly funded research? (This second question was simplified from our original intent, which was to determine if there is a curvilinear relationship between structural significance and federal funding.)

Table 3. Individual funding sources accounting for more than one half of one percent of the sample.

	Count	Percentage	Parent agency	Country	Public	Mean generativity	SD
The National Institute of Mental Health (NIMH)	128	13.73	NIH	US	Public	1.082	3.719
National Institute of Health (NIH)	62	6.65	NIH	US	Public	1.050	2.367
National Science Foundation (NSF)	54	5.79		US	Public	1.080	2.282
National institute on Drug Abuse (NIDA)	33	3.54	NIH	US	Public	1.056	1.678
Social Sciences and Humanities Research Council of Canada	23	2.47		Canada	Public	1.047	1.424
Medical Research Council (UK)	21	2.25		UK	Public	1.170	1.193
National Institute on Aging	18	1.93	NIH	US	Public	0.872	1.504
National Institute of Child Health and Human Development (NICHD)	18	1.93	NIH	US	Public	1.101	0.979
German Research Foundation (DFG)	17	1.82		Germany	Private	1.012	1.000
Natural Sciences and Engineering Research Council of Canada	15	1.61		Canada	Public	1.098	1.566
Wellcome Trust	15	1.61		UK	Private	1.312	1.076
National Institute on Alcohol Abuse and Alcoholism (NIAAA)	14	1.50	NIH	US	Public	0.986	1.049
WT Grant Foundation	12	1.29		US	Public	0.953	0.775
Centers for Disease Control and Prevention (CDC)	9	0.97	HHS	US	Public	1.213	0.331
The John D. and Catherine T. MacArthur Foundation	9	0.97		US	Private	1.017	0.800
Maternal and Child Health Bureau (MCHB)	9	0.97	HHS	US	Public	1.227	0.622
Australian Research Council	8	0.86		Australia	Public	0.917	0.639
Economic and Social Research Council (UK)	8	0.86		UK	Private	0.909	1.097
James S. McDonnell Foundation	8	0.86		US	Private	0.832	1.154
Netherlands Organisation for Scientific Research (NWO)	8	0.86		Netherlands	Public	1.128	0.245
United States Department of Veterans Affairs (VA)	7	0.75		US	Public	0.877	0.493
Canadian Institutes of Health Research (CIHR)	6	0.64		Canada	Public	1.114	0.770
National Institute of Neurological Disorders and Stroke (NINDS)	6	0.64	NIH	US	Public	1.054	0.686
Spencer Foundation	6	0.64		US	Private	1.243	0.576
Eli Lilly and Co.	5	0.54		US	Private	0.911	0.313
Royal Netherlands Academy of Arts and Sciences (KNAW)	5	0.54		Netherlands	Private	0.937	0.042
Total		56.22					

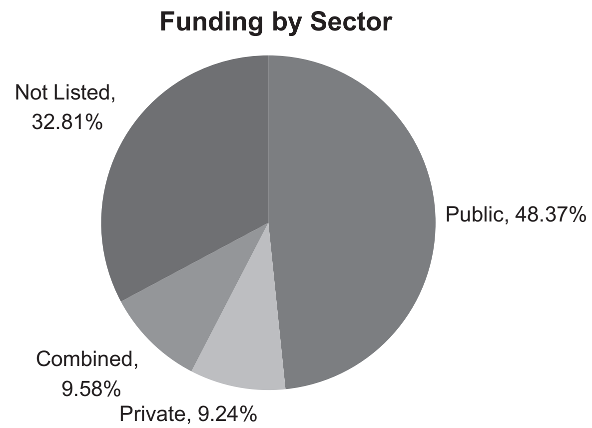

Figure 4. Article funding sources by sector.

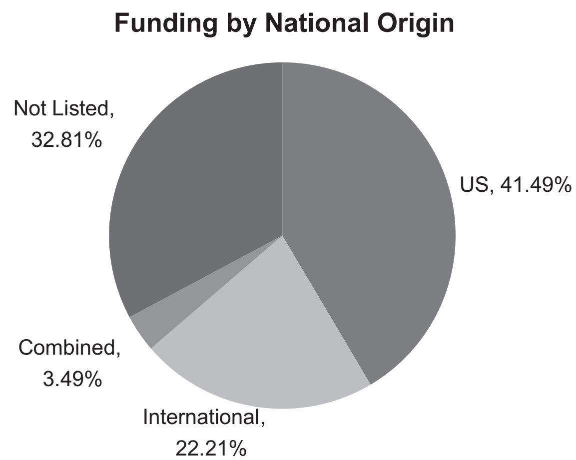

Figure 5. Article funding sources by national origin.

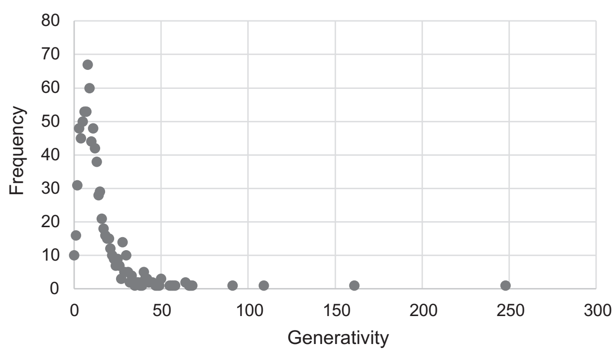

Figure 6. Distribution of generativity before normalization.

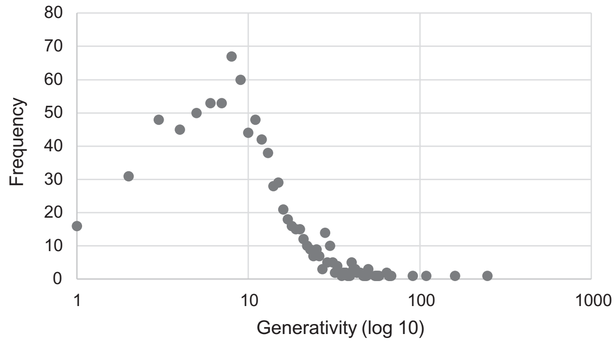

Figure 7. Distribution of generativity after normalization.

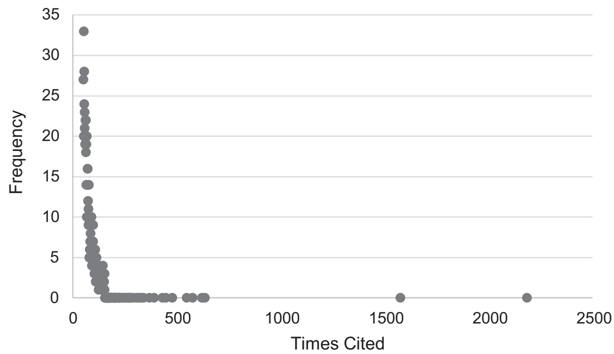

Figure 8. Distribution of times cited before normalization.

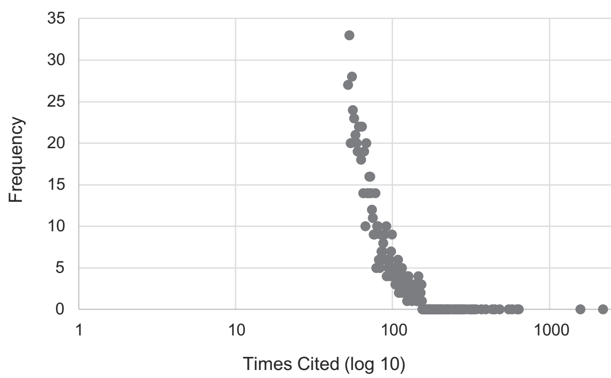

Figure 9. Distribution of times cited after normalization.

Table 4. Descriptive statistics for generativity and times cited.

Measure	Mean	Median	Mode	SD	Skewness	Kurtosis
Times cited	99.99	74.00	53.00	106.49	12.27	208.99
Generativity	13.14	10.00	8.00	14.49	7.19	92.92
TC log 10	1.93	1.87	1.72	0.20	1.78	5.39
G log 10	1.03	1.04	0.95	0.32	−0.12	0.96

Notes: TC log 10 = times cited log 10, G log 10 = generativity log 10.

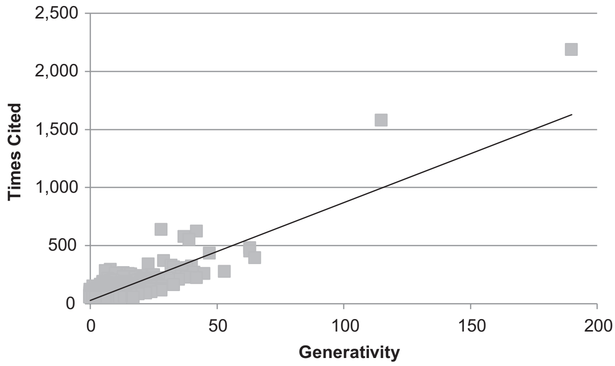

Figure 10. Correlation of generativity and times cited.

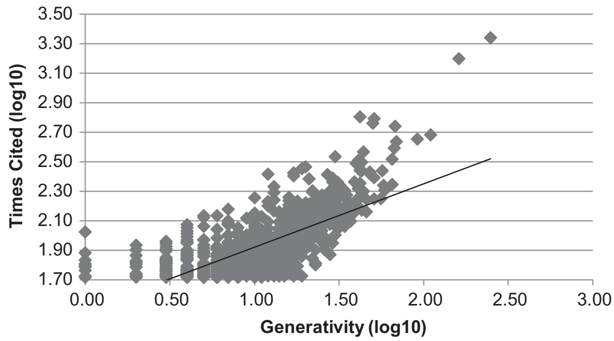

Figure 11. Correlation of generativity log 10 (G log 10) and times cited log 10 (TC log 10).

To answer the first question, whether research which reports its funding source more likely to be structurally significant than research which does not, we conducted an ANOVA with Generativity (G log10) as the DV and Funding Source (Reported, Not Reported) as the IV. We found that research that reported its funding source (M = 1.04, SD = 0.31) was more generative than research that does not (M = 1.00, SD = 0.33), F (1,885) = 3.94, $p < 0.05$. We repeated this analysis with Times Cited (TC log 10) as the DV. The difference was much smaller, and was not statistically significant (Reported: M = 1.93, SD = 0.21, Not Reported: M = 1.93, SD = 0.19, F (1,885) = 0.08, $p = 0.77$).

The second research question, whether privately funded research is more likely to be structurally significant than publicly funded research, lead us to conduct an ANOVA with G log 10 as the DV, and Funding Source (Not Reported, Public, Private, Combined) as the IV. This analysis indicated that there was a significant effect of Funding Source on G log 10, F (3,883) = 4.16, $p < 0.01$, partial $\eta^2 = 0.014$, see Figure 12. The same analysis, replacing the DV with TC log 10, did not indicate a significant effect (F (3,883) = 2.37, $p = 0.069$, partial $\eta^2 = .008$; see Figure 13).

Post hoc comparisons with G log 10 indicated that Generativity was greater for articles with a private funding source (M = 1.13, SD = 0.29) than for those with a public funding source (M = 1.02, SD = 0.33), $p < 0.05$, and greater for those with a private funding source than for those whose funding source was not listed (M = 1.00, SD = 0.32), $p < 0.01$.

Table 5. Journal ranking for psychology in 2002 by mean generativity.

Rank	Journal title	No. of articles	No. of generative articles	Mean citations	Mean generativity	% Generative	Impact factor
1	*Psychological Methods*	28	7	1028.86	116.00	0.39	1.315
2	*Perception*	108	5	103.20	41.00	0.28	1.314
3	*Journal of the Experimental Analysis of Behavior*	64	1	179.00	40.00	0.06	1.579
4	*Journal of Child Psychology and Psychiatry*	10	2	120.00	33.00	0.11	2.514
5	*Advances in Experimental Social Psychology*	7	5	201.60	29.00	0.28	4.7
5	*Journal of Experimental Psychology-Applied*	21	3	155.33	29.00	0.17	1.58
6	*Journal of Abnormal Psychology*	65	21	153.95	26.85	1.18	3.215
7	*Psychological Review*	20	7	161.86	25.50	0.39	6.75
8	*Group Dynamics-Theory Research and Practice*	21	3	153.33	25.33	0.17	0.17
9	*Journal of Autism and Developmental Disorders*	51	14	109.14	25.17	0.79	2.142
10	*Journal of Research in Personality*	39	2	127.50	25.00	0.11	0.905
11	*Journal of Cognitive Neuroscience*	102	48	136.92	23.40	2.71	6.096
12	*Development and Psychopathology*	35	14	109.43	22.40	0.79	4.121
13	*Neurobiology of Learning and Memory*	66	14	124.64	21.30	0.79	2.417
14	*British Journal of Developmental Psychology*	33	1	161.00	21.00	0.06	1.041
15	*Psychophysiology*	91	14	108.00	20.50	0.79	2.674
16	*Neuropsychologia*	234	45	98.38	20.05	2.54	3.184
17	*Infancy*	27	1	90.00	20.00	0.06	N/A
18	*Psychological Assessment*	43	9	138.78	19.80	0.51	2.041
19	*Psychology of Addictive Behaviors*	53	5	78.60	19.00	0.28	1.432
19	*Psychotherapy and Psychosomatics*	36	3	97.33	19.00	0.17	3.188
20	*Journal of Vocational Behavior*	45	7	165.86	18.60	0.39	1.99

Table 6. Journal ranking for psychology in 2002 by percentage of generative articles.

Rank	Journal title	No. of articles	No. of generative articles	Mean citations	Mean generativity	% Generative	Impact factor
1	*Journal of Personality and Social Psychology*	148	78	109.99	13.70	4.40	3.649
2	*Journal of Clinical Psychiatry*	206	73	98.38	12.29	4.11	4.333
3	*Journal of the American Academy of Child and Adolescent Psychiatry*	170	57	97.67	16.31	3.21	3.662
4	*Child Development*	118	53	99.92	12.58	2.99	3.272
5	*Journal of Cognitive Neuroscience*	102	48	136.92	23.40	2.71	6.096
6	*Neuropsychologia*	234	45	98.38	20.05	2.54	3.184
7	*Journal of Applied Psychology*	101	43	103.51	16.24	2.42	1.98
8	*Personality and Social Psychology Bulletin*	145	38	81.76	10.81	2.14	1.758
9	*Psychological science*	95	32	111.94	15.94	1.80	2.961
10	*Journal of Consulting and Clinical Psychology*	74	31	99.45	16.17	1.75	3.613
11	*Developmental Psychology*	73	30	93.97	10.08	1.69	2.496
12	*Physiology & Behavior*	265	28	98.32	14.54	1.58	1.652
13	*Psychological Medicine*	124	26	110.81	9.27	1.47	2.784
14	*Health Psychology*	73	23	86.61	9.33	1.30	3.5
15	*Behaviour Research and Therapy*	101	22	90.09	14.83	1.24	2.188
15	*Psychosomatic Medicine*	81	22	120.95	12.58	1.24	3.218
16	*Journal of Abnormal Psychology*	65	21	153.95	26.85	1.18	3.215
16	*Journal of Adolescent Health*	140	21	95.71	9.82	1.18	1.544
17	*Journal of Educational Psychology*	65	20	92.50	12.78	1.13	0.476
18	*Cognition*	67	19	95.63	17.33	1.07	3.099
18	*Journal of Child Psychology and Psychiatry and Allied Disciplines*	58	19	87.16	18.45	1.07	2.514
18	*Journal of Experimental Psychology-Human Perception and Performance*	90	19	95.58	10.90	1.07	2.335

Table 7. Journal ranking for psychology in 2002 by impact factor.

Rank	Journal title	No. of articles	No. of generative articles	Mean citations	Mean generativity	% generative	Impact factor
1	*Behavioral and Brain Sciences*	3	1	67.00	2.00	0.06	8.73
2	*Trends in Cognitive Science*	N/A	N/A	N/A	N/A	N/A	8.129
3	*Annual Review of Psychology*	N/A	N/A	N/A	N/A	N/A	7.898
4	*Psychological Bulletin*	13	6	140.17	13.00	0.34	7.011
5	*Psychological Review*	20	7	161.86	25.50	0.39	6.75
6	*Monographs of the Society for Research in Child Development*	N/A	N/A	N/A	N/A	N/A	6.625
7	*Psychological Inquiry*	N/A	N/A	N/A	N/A	N/A	6.25
8	*Journal of Cognitive Neuroscience*	102	48	136.92	23.40	2.71	6.096
9	*American Psychologist*	27	3	113.67	5.50	0.17	5.981
10	*Advances in Experimental Social Psychology*	7	5	201.60	29.00	0.28	4.7
11	*Journal of Clinical Psychiatry*	206	73	98.38	12.29	4.11	4.333
12	*Development and Psychopathology*	35	14	109.43	22.40	0.79	4.121
13	*Cognitive Psychology*	19	6	102.83	15.67	0.34	4.059
14	*Journal of the American Academy of Child & Adolescent Psychiatry*	170	57	97.67	16.31	3.21	3.662
15	*Journal of Personality and Social Psychology*	148	78	109.99	13.70	4.40	3.649
16	*Journal of Consulting and Clinical Psychology*	74	31	99.45	16.17	1.75	3.613
17	*Health Psychology*	73	23	86.61	9.33	1.30	3.5
18	*Cognitive Neuropsychology*	29	1	55.00	1.00	0.06	3.391
19	*Journal of Experimental Psychology: General*	32	13	88.15	10.60	0.73	3.348
20	*Child Development*	118	53	99.92	12.58	2.99	3.272

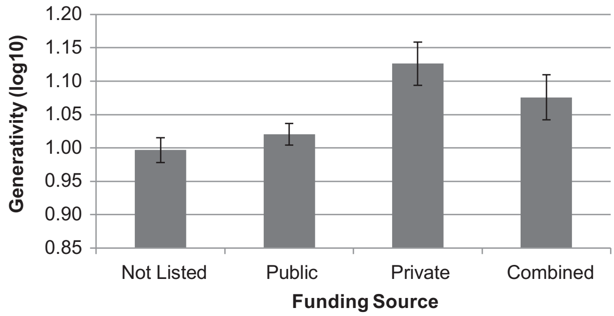

Figure 12. Generativity by funding source sector.

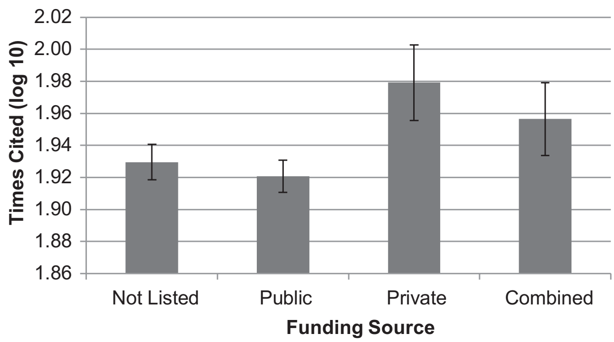

Figure 13. Times cited by funding source sector.

Table 8. Structural significance and funding sources.

Funding source	Mean generativity	SD	SEM	Mean TC	SD	SEM
Not listed	0.997	0.315	0.018	1.930	0.190	0.011
Public	1.020	0.334	0.016	1.921	0.207	0.010
Private	1.126	0.292	0.032	1.979	0.214	0.024
Combined – public/private	1.076	0.309	0.034	1.956	0.209	0.023
US	1.030	0.333	0.016	1.937	0.226	0.011
International	1.025	0.323	0.036	1.917	0.176	0.019
Combined – US/international	1.127	0.279	0.030	1.957	0.189	0.020
Sum	1.028	0.324	0.011	1.932	0.203	0.007

Exploratory inferential statistics

Sector and national origin

The addition of a National Origin categorization for funding sources leads to the suggestion that there might be a difference in the relationship between public and private funding between countries, because the nature of public funding institutions, both structurally and culturally, might vary between nations. Coding for National Origin allowed us to test for an interaction between the effect of Public vs. Private funding sources, and the effect of Domestic vs. International Funding sources, for Generativity (G log 10) and for Times Cited (TC log 10; See Table 8). This was followed by an ANOVA with G log 10 as the DV, and Funding Source category (Not Listed, U.S., International, Combined) as the IV. As when comparing public and private sources, this analysis was repeated, replacing the DV with TC log 10. The first set of tests did not indicate a significant interaction F (4,886) = 2.02, p = 0.088, partial η^2 = 0.009, a significant main effect of Public vs. Private Funding Source (F (2,886) = 0.87, p = 0.42), Partial η^2 = 0.002, or a significant main effect of Domestic vs. International Funding Source, F (2,886) = 0.388, p = 0.678, partial $\eta^2 = 0.001$. In short, in our exploratory analyses we did not see any interaction, or any effect of Funding Source on Generativity. Unsurprisingly, the second set of analyses, with TC log 10 as the DV also indicated no significant interaction F (4,886) = 0.30, p = 0.876, partial $\eta^2 = 0.001$, no significant main effect of Public vs. Private Funding Source, F (2,886) = 1.34, p = 0.263, partial $\eta^2 = 0.001$, and no significant main effect of Domestic vs. International Funding Source, F (2,886) = 0.29, p = 0.745, partial $\eta^2 = 0.001$.

Number of funding sources

During the course of analysis it was suggested that Generativity might vary based on the number of funding sources, because of the cautious reception we might expect for transformative ideas from funding institutions. We found no significant difference in Generativity (G log 10) between research supported by multiple funding sources (M = 1.06, SD = 0.309) and research supported by a single source (M = 1.02, I = 0.338), F (1,629) = 2.65 p = 0.104. The second analysis, with TC log 10 as the DV, also indicated no significant effect (Single Source: M = 1.92, SD = 0.22, Multiple Sources: M = 1.94, SD = 0.19, F (1,629) = 2.77, p = 0.097).

Discussion

As an attempt to define and assess transformational science, we developed a new measure of how structurally significant (i.e. its relation to new branches of knowledge) a scientific work is, namely "generativity." In short, generativity appears to be a valid new index of structurally significant science.

Validity of generativity as a new bibliometric indicator

Generativity offers a partial glimpse into the Tree of Knowledge, with unique advantages over other bibliometric measures. Situated at a level between the immediacy of pure citation count and the judgment of history, generativity balances the advantages of perspective with the demands of relevance. In the context of the debate about the nature of transformative research, a computational measure also has the advantage of reducing ambiguity, and hopefully may encourage more clarity in the definition of research that is (and is not) transformative.

The journal impact factor, one of the most widely-used bibliometric measures, was created to help librarians prioritize journals to include in their collections (Garfield 2006). Despite this original intent, the measure has since been used to influence decisions about hiring, promotions, tenure, awarding grants (Meho 2006; The PLoS Medicine Editors 2006), and in some cases even government funding (Adam 2002; Ferreira, Antoneli, and Briones 2013). Journal impact factor is unsuitable for these roles, both because of a lack of transparency in the measure (Thomson ISI, a private corporation, alone decides which papers are "citable"), and because it applies at the level of journals rather than individual contributions. Impact factor has also been criticized for the undue influence of a small number of highly cited articles (or a large number of uncited articles), the exaggerated impact of review articles, and the limited perspective of a two year "citation window" (Meho 2006). There is also reason to believe that reliance on impact factor underestimates the impact of social science (Hegarty and Walton 2012). In the past year, researchers at American Society for Cell Biology published a declaration decrying journal impact factor's flaws and abuses, and calling for better research metrics (San Francisco Declaration on Research Assessment 2013). As of this writing the declaration has more than 6000 signatures, but this is by no means the first time that journal impact factor has been subject to these criticisms (Campbell 2008; Kurmis 2003; Largent and Lane 2012; Opthof 1997; Seglen 1997).

Perhaps because of their accessibility, and certainly in part due to a perception of objectivity, quantitative measures can be misused. Julia Lane, the former programme director of SciSIP (Science of Science and Innovation Policy) at NSF wrote "Science should learn lessons from the experiences of other fields, such as business. The management literature is rich in sad examples of rewards tied to ill-conceived measures, resulting in perverse outcomes." (Lane 2010, 488) This is just as true for bibliometric measures as it is for IQ, Body Mass Index (BMI), standardized testing scores, or the Dow Jones Industrial Average. Often critics of these indicators argue that we should rely on more narrative evaluations, but this is an insufficient response. No bibliometric measure will ever be a substitute for expert judgment (and generativity is not an exception) but the cost of obtaining expert evaluation can quickly become prohibitive. It does not scale well, it is already strongly correlated with many bibliometric measures (Oppenheim 1996), and despite being a "gold standard", it is also worth considering that expert judgment itself might require some form of validation (Harnad 2008). One of the best answers to abuses of a quantitative measure is to provide a better quantitative measure.

Many bibliometric measures, like the journal impact factor or the *h*-index (Hirsch 2005), are derived from citation data, but are able to achieve a greater degree of nuance (Cronin and Meho 2006). Generativity is one such measure, but given its relative correspondence to raw citation count it may be possible, assuming enough time has passed for collecting generativity to be feasible, that it could be substituted as the basis for other citation derived metrics.

The form of generativity that we have explored here is in many ways an incomplete approximation, limited by time and resources. Although a version of the measure has been fully specified in this paper, it should not necessarily be understood as definitive. The core of the concept of generativity – of examining the contribution of individual articles by looking further down the branches of the tree – can be implemented in a variety of ways. This could be as simple as varying the threshold for citation counts, or as complex as basing the measure on the shape of the growth curve of citations. In either case, the central concept is the same. Present evidence suggests that, on the spectrum of bibliometric measures (Bollen, Van de Sompel, Hagberg and Chute, 2009), generativity or a measure derived from it will prove itself to occupy a novel and useful niche.

Remind readers of two main research questions and summarize findings on them here. First, we examined the question of whether the existence of funding was at all related to structurally significant science (i.e. generative science). First, funding is not evenly distributed among agencies. We saw that generativity was greater for those articles that reported their funding sources than those that did not. The NIH was, predictably, the largest single funding source for generative research in psychology. However, almost half of the articles in our sample were funded by sources that individually accounted for less than one half of one percent of the sample. When we look at funding sources by Sector and by National Origin, we also see a great deal more cooperation between the public and private sectors (and much of that within the U.S.) than we see between nations.

Second, we examined the question of whether public funding agencies are more conservative in their funding decisions, and therefore less likely to fund transformative research. The data support this assumption: generativity varied based on funding source, and it was greater for privately funded research. We also saw a difference in the same direction for citation count, but it was smaller and was not statistically significant. This pattern is consistent with the idea that both generativity and times cited are measures of structural significance but that generativity is the more sensitive measure. Given our assumptions, our results are consistent with (1) generativity containing information not provided by pure citation count, and (2) private funding sources (at least in the U.S.) recognizing and encouraging more generative research than public sources. This should not be construed as implying that privately funded science has more value than science that is publicly funded. It may be that private sources are free to pursue riskier ideas only in a context where more basic science (Kuhn's "puzzle solving" science) is publicly funded.

Limitations

We have only examined the top 10% of articles by citation count (published in English, in Psychology, and in 2002). This does not provide a picture of the overall funding situation. Although it may be that what we see at the top is a small-scale version of the whole distribution it is important to recognize that we are looking at the "winners", and that greater context could change the interpretation of our findings. Additionally, although we would argue that a more generative article is a more transformative one, we recognize that the measure requires further validation. We lack test-retest reliability as well as predictive validity of the generativity index. Finally, whereas our statistical techniques are robust against some degree of violation of normality (Howell 1997), it is possible that the skewed distribution of citation data renders some of our analysis suspect and in need of replication.

Future research

The present study suggests three kinds of future projects, (1) research that focuses on further validating generativity, (2) research that extends or improve on the quality of generativity, and (3) the development of tools to increase the ease of use of the measure.

We suggest two complimentary approaches to validating generativity. First, if expert ratings of transformativeness for a sample of articles (which have generativity scores) could be collected, and compared to citation count, we would expect that generativity scores would correlate more strongly with expert judgments than pure citation count. The second validation study would require extending generativity to the researcher level, so that it could be correlated against measures of lifetime achievement (awards, honors, etc.). Here generativity could be compared against citation count as well as a variety of other scientometric measures (*h*-index, creativity index, etc.).

Other future projects could include development of automated tools to ease in the collection of an even more complete generativity score, and research to fine tune the measure (varying aspects of the measure such as citation count thresholds) and to extend it to other levels (researcher, journal, society).

Ultimately the judgment of transformativeness belongs to the history of science, but such a judgment requires a perspective far removed from funding decisions that are being made today. It is our hope that generativity, or bibliometric measures derived from it, might provide decision makers with more complete information in an appropriately timely manner. We also hope that generativity might serve as a foundation for future empirical investigations into the structure and nature of transformative science.

Notes

1. The choice to focus on ideas should not be construed as denying the impact of the individual participants in shaping a particular course – 'Generic eventuality is not equivalent to specific inevitability' (Simonton 2004).
2. One possible drawback of the tree of life metaphor is that it implicitly downplays the impact of interdisciplinary work. These collaborations would be metaphorically equivalent to horizontal gene transfer, which in fact does occur in most branches (prokaryotes, bacteria, and archea) of the tree of life.
3. The ISI Web of Science uses two fields to categorize articles by subject, Subject Area and Web of Science Category. The Subject Areas correspond to thesauri managed by the indexers and editorial staff of Thomson Reuters. Notes that clarify and define the scope for the various subject areas, which are specific to each index, are available online (http://ip-science.thomsonreuters.com/mjl/scope/). Web of Science categories are assigned at the journal level. These categories are assigned in the Thomson Reuters Journal Citation Reports, and carry over to the Web of Science.
4. Ideally the percentage used in this ranking would be equal to the number of Generative articles divided by the number of published articles. In the present example (Table 6) the number of published articles only includes those collected in our sample.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Barrett Anderson is a graduate student in the *De*partment of Psychology,UC Santa Cruz, Santa Cruz, CA 95064. His research interests are learning and technology, educational games, and meta-science. His publications include “Achieving the Performance Benefits of Hands-on Experience When Using Digital Devices: A representational approach,”*Computers in Human Behavior 2*016 and “Self-representation in social anxiety disorder: Linguistic analysis of autobiographical narratives,” *Behaviour Research and Therapy* 2008.

Gregory J. Feist is a professor of Psychology in the Department of Psychology San Jose State University, San Jose, CA 95192. His research interests psychology of science, psychology of creativity, and development of scientific talent. *His publications include Handbook of the Psychology of Science* 2013; *The Psychology of Science and the Origins of the Scientific Mind* 2006; The Psychology of Scientific thought and Behaviour. *The Psychologist* 2013; Psychology of Science as a New Subdiscipline in Psychology. *Current Directions in Psychological Science* 2011; The Psychology of Science: Development and Personality in Scientific Thought, Interest and Achievement. *Review of General Psychology* 2006 and Psychology of science: Review and integration of a Nascent Discipline. *Review of General Psychology* 1998.

References

Adam, D. 2002. "Citation Analysis: The Counting House." *Nature* 415 (6873): 726–729.

Andrés, A. 2009. *Measuring Academic Research: How to Undertake a Bibliometric Study*. Oxford: Chandos Publishing.

Austin, F. C. 2008. "High-risk High-reward Research Demonstration Project." Presentation given to the NIH Council of Councils, Bethesda, MD.

Bollen, J., H. V. Van de Sompel, A. Hagberg, L. Bettencourt, R. Chute, A. A. Rodriguez, and L. Balakireva. 2009. "Clickstream Data Yields High-resolution Maps of Science." *PLoS One* 4 (3): e4803. doi:10.1371/journal.pone.0004803.

Bollen, J., H. V. Van de Sompel, A. Hagberg, and R. Chute. 2009. "A Principal Component Analysis of 39 Scientific Impact Measures." *PLoS One* 4 (6): e6022. doi:10.1371/journal.pone.0006022.

Bornmann, L., and H. D. Daniel. 2008. "What Do Citation Counts Measure? A Review of Studies on Citing Behavior." *Journal of Documentation* 64 (1): 45–80.

Bornmann, L., R. Mutz, C. Neuhaus, and H. D. Daniel. 2008. "Use of Citation Counts for Research Evaluation: Standards of Good Practice for Analyzing Bibliometric Data and Presenting and Interpreting Results." *Ethics in Science and Environmental Politics* 8: 93–102.

Campbell, P. 2008. "Escape from the Impact Factor." *Ethics in Science and Environmental Politics* 8 (1): 5–7.

Cronin, B., and L. Meho. 2006. "Using the H-index to Rank Influential Information Scientists." *Journal of the American Society for Information Science and Technology* 57 (9): 1275–1278.

De Bellis, N. 2009. *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*. Lanham, MD: Scarecrow Press.

Feist, G. J. 2006. *The Psychology of Science and the Origins of the Scientific Mind*. New Haven, CT: Yale University Press.

Ferreira, R. C., F. Antoneli, and M. R. S. Briones. 2013. "The Hidden Factors in Impact Factors: A Perspective from Brazilian Science." *Frontiers in Genetics* 4 (130): 1–2. doi:10.3389/fgene.2013.00130.

Frodeman, R., and J. B. Holbrook. 2012. "The Promise and Perils of Transformative Research." Report on the workshop: Transformative research: Ethical and societal implications, Arlington, VA.

Garfield, E. 2006. "The History and Meaning of the Journal Impact Factor." *JAMA-Journal of the American Medical Association* 295 (1): 90–93.

Gholson, B., W. R. Shadish, R. A. Neimeyer, and A. C. Houts, eds. 1989. *The Psychology of Science: Contributions to Metascience*. Cambridge: Cambridge University Press.

Harnad, S. 2008. "Validating Research Performance Metrics against Peer Rankings." *Ethics in Science and Environmental Politics* 8 (11). Accessed August 2, 2015. http://eprints.soton.ac.uk/265619/1/esep-harnad.html

Harzing, A. W. K., and R. van der Wal. 2008. "Google Scholar as a New Source for Citation Analysis." *Ethics in Science and Environmental Politics* 8 (1): 62–71.

Hegarty, P., and Z. Walton. 2012. "The Consequences of Predicting Scientific Impact in Psychology Using Journal Impact Factors." *Perspectives on Psychological Science* 7 (1): 72–78.

Hirsch, J. E. 2005. "An Index to Quantify an Individual's Scientific Research Output." *Proceedings of the National Academy of Sciences of the United States of America* 102 (46): 16569–16572.

Howell, D. C. 1997. *Statistical Methods for Psychology*. 4th ed. Belmont, CA: Duxbury Press.

H.R. 5116–111th Congress: America COMPETES Reauthorization Act of 2010. 2010.

Kuhn, T. 1962. *The Structure of Scientific Revolutions*. Chicago, IL: The University of Chicago Press.

Kurmis, A. P. 2003. "Understanding the Limitations of the Journal Impact Factor." *The Journal of Bone & Joint Surgery* 85 (12): 2449–2454.

Lane, J. 2010. "Let's Make Science Metrics More Scientific." *Nature* 464 (7288): 488–489.

Largent, M. A., and J. I. Lane. 2012. "STAR METRICS and the Science of Science Policy." *Review of Policy Research* 29 (3): 431–438.

Meho, L. I. 2006. The Rise and Rise of Citation Analysis. *Physics World* 20 (1): 32.

Meho, L. I., and K. Yang. 2007. "A New Era in Citation and Bibliometric Analyses: Web of Science, Scopus, and Google Scholar." *Journal of the American Society for Information Science and Technology* 58: 1–21.

Merton, R. K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago, IL: Chicago University Press.

Norris, M., and C. Oppenheim. 2007. "Comparing Alternatives to the Web of Science for Coverage of the Social Sciences' Literature." *Journal of Informetrics* 1 (2): 161–169.

Oppenheim, C. 1996. "Do Citations Count? Citation Indexing and the Research Assessment Exercise (RAE)." *Serials: The Journal for the Serials Community* 9 (2): 155–161.

Opthof, T. 1997. "Sense and Nonsense about the Impact Factor." *Cardiovascular Research* 33 (1): 1–7.
Osborne, J. 2002. "Notes on the Use of Data Transformations." *Practical Assessment, Research & Evaluation* 8 (6). Accessed July 25, 2015. http://pareonline.net/getvn.asp?v=8&n=6
The PLoS Medicine Editors. 2006. "The Impact Factor Game." *PLoS Medicine* 3 (6): e291. doi:10.1371/journal.pmed.0030291.
Popper, K. 1972. *Objective Knowledge: An Evolutionary Approach*. Oxford: Oxford University Press.
Radicchi, F., S. Fortunato, and C. Castellano. 2008. "Universality of Citation Distributions: Toward an Objective Measure of Scientific Impact." *Proceedings of the National Academy of Sciences* 105 (45): 17268–17272.
Rieppel, O. 2010. "The Series, the Network, and the Tree: Changing Metaphors of Order in Nature." *Biology & Philosophy* 25: 475–496.
San Francisco Declaration on Research Assessment. 2013. Accessed July 23, 2013. http://am.ascb.org/dora/
Seglen, P. O. 1997. "Why the Impact Factor of Journals Should Not Be Used for Evaluating Research." *BMJ: British Medical Journal* 314 (7079): 498–502.
Simonton, D. K. 2004. *Creativity in Science: Chance, Logic, Genius, and Zeitgeist*. Cambridge: Cambridge University Press.