Creativity Research Journal

Routledge
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hcrj20

The Importance of Separating Appropriateness into Impact and Feasibility for the Psychology of Creativity

Joel Chan & Christian D. Schunn

To cite this article: Joel Chan & Christian D. Schunn (2023): The Importance of Separating Appropriateness into Impact and Feasibility for the Psychology of Creativity, Creativity Research Journal, DOI: 10.1080/10400419.2023.2191919

To link to this article: https://doi.org/10.1080/10400419.2023.2191919

CREATIVITY RESEARCH JOURNAL
https://doi.org/10.1080/10400419.2023.2191919

Routledge
Taylor & Francis Group

ARTICLE

The Importance of Separating Appropriateness into Impact and Feasibility for the Psychology of Creativity

Joel Chan [a] and Christian D. Schunn [b]

[a]College of Information Studies, University of Maryland; [b]Learning Research and Development Center, University of Pittsburgh

ABSTRACT

Scientific progress on creativity research depends on having properly operationalized measures. In psychological research on creativity, it is common to operationalize creativity as the combination of novelty and appropriateness. However, the operationalization of appropriateness varies widely across researchers, studies, and domains (e.g. technical goodness, significance, elegance, usefulness, and feasibility). We argue that a core distinction between impact (how useful an idea is for solving the problem) and feasibility (how easy it is to realize the idea) underlies the variation. We further claim that this distinction is both possible to capture reliably in practice and psychologically significant. To test these claims, 318 ideas from 5 real-world social innovation problems (e.g. improving accessibility in elections) were rated for novelty, impact, and feasibility by a set of six experts selected for each of the 5 challenges. We find that all three constructs can be measured reliably and are statistically separable. Further, we show that distinguishing impact and feasibility reveals theoretically meaningful patterns of relationships with key psychological processes of creativity – analogy and conceptual combination – that would be difficult if impact and feasibility were conflated. These results demonstrate the theoretical importance of separating appropriateness into impact and feasibility for the psychology of creativity.

ARTICLE HISTORY
Received February 10, 2022

Introduction

Measuring the creativity of "products" is an important aspect of the scientific study of creativity. Early work on creativity sought to understand the characteristics of creative personalities, which required reliable ways of determining which people were creative and which ones were not (Guilford, 1950). In the current "sociocultural" approach to creativity research (Sawyer, 2012), other aspects of creativity are being studied: in Mel Rhodes' classic 4P distinction (Rhodes, 1961), creativity research across multiple disciplines and fields now studies the characteristics of people, processes, and pressures that constrain or facilitate the generation of creative products. Across these domains of creativity research, it is frequently critical to be able to conceptualize and measure creative products in a rigorous manner.

Many definitions of creative products have been proposed. One common thread is that both "novelty" and "appropriateness" are required for a product to be considered creative. As Sawyer (2012) described in his review of definitions of creativity, "Creativity is the generation of a product that is judged to be novel and also to be appropriate, useful, or valuable by a suitably knowledgeable social group." This is sometimes called the "standard definition of creativity" (Runco & Jaeger, 2012), and aligns with many other prominent theoretical definitions of creativity. For example, in Csikszentmihalyi's Systems Model (Csikszentmihalyi, 1997), a product is considered creative if it has contributed to a domain, based on judgments from an appropriate set of experts in that domain. Simonton's (2016) tripartite model distinguishes statistical infrequency of the idea and surprise (with respect to the utility of the idea) as aspects of novelty, in part to incorporate the core idea of "blindness" from his Darwinian model of creativity (Simonton, 1999), but also includes a notion of appropriateness. Amabile's (1983) componential conceptualization of creativity also emphasizes novelty and appropriateness as core aspects of creativity. A recent review of the literature on creativity (Puryear & Lamb, 2020) further documents the high regularity in definitions coalescing around the core dimensions of novelty and appropriateness. It is worth noting, however, that everyday judgments of creativity may weight the novelty criterion more highly than appropriateness (Pichot et al., 2022)

In this paper, we want to focus on the *appropriateness* dimension of the creativity of ideas. How should we

CONTACT Joel Chan joelchan@umd.edu College of Information Studies, University of Maryland, Hornbake Building, South Wing Room 2118E, 4130 Campus Drive, College Park, MD 20742

think about it and how should we measure it? And why might it matter? Although there does seem to be a general shared intuition that a creative idea should be "appropriate," there is substantial variability in how it is conceptualized and measured across domains. For example, novel products can be considered creative if they are judged by domain experts to also be *esthetically appealing* (e.g. in music, art; Amabile, 1982; Gorder, 1980), *relevant* to a problem (MacCrimmon & Wagner, 1994), *correct/rigorous* and/or *significant* (e.g. in scientific theories, mathematical proofs; Simonton, 2009), *elegant* (e.g. in dance, mathematical proofs), *technically good* (e.g. in art; Amabile, 1982), or *technically feasible* (e.g. in engineering designs; Chan et al., 2011; Shah, Vargas-Hernandez, & Smith, 2003).

At first glance, the variation in ways of operationalizing appropriateness appears to involve incommensurable domain-specificity. However, we argue here that a different conceptualization of appropriateness captures an important core aspect of these variations. In particular, we propose that there are two consistently important dimensions of variation in how researchers conceptualize appropriateness: 1) *impact* (what is the degree of potential *contribution* to the domain of interest) and 2) *feasibility* (to what extent is this contribution *possible*). To provide some pragmatic support for this conceptual distinction as well as suggestions of how it applies in different creative domains, we begin by reviewing a variety of cases that support the distinction.

Consider the NIH's primary review criterion of "overall **impact**/benefit" for funding proposals, relative to the role of pilot or "proof-of-concept" data (feasibility): the former criterion indexes the expected degree of potential contribution to the particular research or societal problem of interest, while the latter indexes the degree to which the proposed contribution is possible. Consider also how the NIH separates impact from feasibility in its "high risk, high reward" funding programs,[1] many of which do not require pilot data (lowering criteria for feasibility) in favor of extraordinary potential for broad impact. On the flip side, consider the many papers that are considered to be "incremental," not necessarily for lack of novelty (they are often quite different from prior work), or feasibility (they are often technically correct and/or rigorous!), but for lack of *impact* (advancement in some important aspects of understanding) (Sandberg & Alvesson, 2011; Sverdlov, 2018). In design and engineering, consider the difference between the many proposed mechanisms for achieving flight before the Wright brothers' breakthroughs, which were certainly high in potential impact but low in feasibility (they did not work): in other words, "the Wright brothers did not invent the concept of the airplane; rather, they invented the first practical airplane" (Anderson, 2004, p. vii). The recent landmark result on fusion ignition from the Lawrence Livermore National Laboratory is also significant because of how it is increasing the feasibility of the very high impact and novel idea of controlled fusion for clean renewable energy (Bishop, 2022) . Indeed, it is often easy to think of ideas that would be impactful in theory, but could never be realized (e.g. a perpetual motion machine).

While these examples come from scientific and engineering creativity where impact and feasibility have relatively straightforward definitions, the distinction between impact and feasibility can also be mapped to the variations in how people operationalize appropriateness across other domains of creative achievement. For example, creativity in art or music often strikes a delicate balance between maximizing impact (e.g. esthetic appeal and emotional resonance/depth) and feasibility (e.g. Do the notes actually work together? Can we even create this sculpture or successfully execute this special effect in the film?). Picasso's Cubism can be seen as an example of when this balance succeeds in bringing forth new ideas that are both highly impactful and feasible, and thereby radically transforming a domain (Sgourev, 2013). In sports, we often celebrate athletes as creative who not only defeat their opponents or break records, but do so while rewriting the rules for what is possible. Simone Biles is a recent example of this, who broke records in gymnastics by also inventing new gymnastics skills and expanding our sense of what is possible in gymnastics (Orbey, 2019). The intuition behind separating impact and feasibility is also consistent with prominent theoretical conceptualizations of creativity (though it is not clearly discussed as such): for example, Csikszentmihalyi's (1997) systems view of creativity proposes that creative products are those which a field of knowledgeable/qualified individuals (who draw on and work within a common domain) deem to be a useful contribution to their shared domain. From this viewpoint, it makes sense that the most creative products are both impactful (i.e. a meaningful potential contribution to the domain of interest) and feasible (i.e. there is a reasonable expectation that the potential contribution could be realized). Similarly, Boden's (2004) prominent notion of transformational creativity also rests on the difficulty of constructing novel products that are both impactful (recognized as a significant contribution to a domain) *and* feasible (rewrites the rules for what is possible).

In addition to arguing for the empirical separability of impact from feasibility, we also argue that failing to distinguish impact and feasibility can substantially limit the understanding of the psychological antecedents of

creativity. In particular, separating impact and feasibility will likely reveal distinct patterns of creativity antecedents with respect to the question of creative inspiration. People build on prior knowledge when generating new ideas, whether they like it or not (Jansson & Smith, 1991; Linsey et al., 2010; Ward, 1994), and whether they are aware of it or not (Marsh & Bower, 1993; Marsh & Landau, 1995; Marsh, Landau, & Hicks, 1997; Schunn & Dunbar, 1996). An active area of creativity research studies the conditions under which different kinds of knowledge – e.g. drawing from *conceptually distant* domains of knowledge, exploring unusual *conceptual combinations* of prior knowledge – are helpful or harmful to creative outcomes (Sio, Kotovsky, & Cagan, 2015) or vary in usage and perception as a function of expertise (Ball, Ormerod, & Morley, 2004; Bonnardel & Marmèche, 2005). Distinguishing when the variations in outcomes of certain knowledge sources signal acceptable risks (e.g. high potential impact but low feasibility) or dead-ends (low potential impact, regardless of feasibility) is both theoretically and practically important. Relatedly, a growing body of research on bias against highly novel research in peer review (Boudreau, Guinan, Lakhani, & Riedl, 2014; Wang, Veugelers, & Stephan, 2016)—of both publications and grants – could benefit from clearly investigating and distinguishing impact and feasibility: how much do reviewers' downwardly biased scores for novel research reflect their judgment of low (expected) impact vs. difficulty understanding how the proposed solutions could be feasibly created or used in the domain? As another example, distinguishing impact and feasibility might help explain why outsiders to a field seem to benefit from collaboration with insiders (Arts & Fleming, 2018): could the mechanisms be better identification of impactful problems or enhanced ability to develop the feasibility of their ideas?

In this paper, we investigate the feasibility and significance of empirically distinguishing impact and feasibility when measuring the creativity of ideas. We demonstrate that distinguishing impact and feasibility is both *possible* (i.e. they can each be measured reliably and separated statistically and are separately predictive of an aggregate measure of creativity) and *psychologically significant* (i.e. they can be tied to distinct psychological phenomena). For example, we show that 1) drawing from conceptually distant domains of knowledge can lead to less impactful, but not necessarily less feasible, ideas; and 2) exploring unusual conceptual combinations of prior knowledge can lead to less feasible, but not necessarily less impactful, ideas. We do this by examining creative outputs across a range of real tasks with people actually trying to solve them, rather than in a lab with a single fictitious task. While the use of a controlled task in a lab setting can produce critical insights, the gap between lab tasks and real-world creativity is often large, and establishing the validity of a construct is more meaningful when tested with a range of real-world tasks and contexts.

Study 1: Is it possible to separate impact and feasibility in practice?

Our first objective is to demonstrate that it is possible to measure impact and feasibility separately in practice. To do this, we investigate the following research questions: 1) can domain experts provide *reliable judgments* of impact and feasibility; 2) are impact and feasibility measures obtained from expert judges *statistically separable*; and 3) are separate impact and feasibility measures obtained from expert judges valid, in the sense that they are *predictive of aggregate creativity*?

Setting

The setting for this study is OpenIDEO, a Web-based innovation platform that crowdsources design ideas from people across the world for a variety of difficult social innovation problems, such as revitalizing struggling urban neighborhoods or increasing accessibility of voting. Concept submissions take the form of solution proposals. These proposals, although not fully completed designs, are typically well thought out, with substantial amounts of detail and often

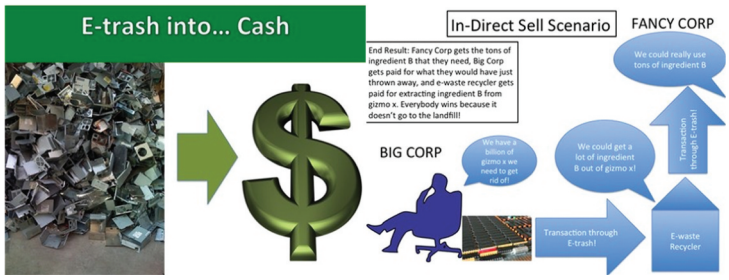
including media (e.g. videos, sketches, and diagrams) as supplements to illustrate the ideas involved (see Figure 1 for an example). The platform was run by expert designers from IDEO, a design firm with a strong track record of design innovation. At the time of data collection, there were 12 completed innovation challenges on the platform, with a total of 2,344 design concept submissions for the platform, submitted by 1190 unique contributors.

Each innovation challenge is sponsored by an organization with expertise in the domain of the problem. The OpenIDEO team assembles an expert panel for each challenge to select the most promising concepts to be *shortlisted* for further development. The shortlist typically consists of about 20 concepts. After additional development of the concepts (typically lasting 3–4 weeks), the expert panel then selects a small subset of the shortlisted concepts for immediate implementation. Thus, prior to the implementation phase, the concepts are not definitive proposals and still undergo refinement.

Through consultation with the OpenIDEO team, we identified shortlist status (whether each concept makes it

E-trash into real cash

Companies can end up with left-over electronics and components for electronics, imagine if there was a marketplace for them to sell their scrap, trash, and left-over chemicals to other companies that need it.



Example Use cases:

1: Big Corp makes 50,000 widgets that need ingredient A in the casing. Unfortunately, the widgets are discontinued and Big Corp is left with mountains of ingredient A that they don't foresee using the future. They are about to throw it all away since they need the space in their warehouse when Big Corp goes to E-trash.com and finds Fancy Corp who just decided to make 100,000 gizmos that really need ingredient A. E-trash facilitates the transaction and mountains of ingredient A don't go to the landfill!

2: Big Corp has thousands of version 1 doodads that they used for the last couple of years but now they need new version 2. They need to get rid of it quickly and so they go to E-trash.com and put it up to find out that Fancy Corp really needs doodads and version 1 works perfectly! Transaction made! Alternatively, version 1 just isn't applicable anymore but ingredient B in it could be very valuable so through E-trash.com they find a recycler who specializes in extracting ingredient B from old electronics and then selling it to other companies.

Description:

It would be an on-line marketplace for businesses to find business buyers for their large quantities of e-waste. Sellers could post, either publicly or to select partners, what "waste" they have available and then buyers could bid on the "waste" that they could actually use. Lots of e-products and electronic components can be re-used and re-purposed. This would provide a method for companies to make money off of their waste and to find necessary products and components at a discount.

Figure 1. Example concept (text and visuals) to illustrate typical amount of detail per concept.

into the shortlist) as their face valid measure of aggregate creativity; the final selection of winners is also driven by characteristics of the proposer, while the shortlist is a clearer measure of the creative potential of the ideas on their own. Finally, the shortlist panel selection by experts is consistent with the prominent use of expert panels in creativity research to determine creativity (Amabile, 1982).

We sampled five challenges involving a total of 318 concepts to collect expert ratings. Our sampling criteria included the number of contributed concepts, topical diversity, and the availability of appropriate experts to give additional ratings of impact and feasibility. The challenge titles were as follows:

(1) **Vibrant Cities** ($N = 119$ concepts): How might we restore vibrancy in cities and regions facing economic decline?
(2) **Bone Marrow** ($N = 41$ concepts): How might we increase the number of registered bone marrow donors to help save more lives?
(3) **Voting Accessibility** ($N = 47$ concepts): How might we design an accessible election experience for everyone?
(4) **Human Rights** ($N = 62$ concepts): How can technology help people working to uphold human rights in the face of unlawful detention?

(5) **Web Entrepreneurs** ($N = 49$ concepts): How might we support web entrepreneurs in launching and growing sustainable global businesses?

The details of the composition of the expert panel for each challenge were not always formally released, but shortlisted concepts were often announced in a blog post addressed to the OpenIDEO community that included some information about the expert panels. For the Vibrant Cities challenge, the panel included a group of Detroit and revitalization advisors, including experts from the Steelcase workplace furnishing company, and IDEO's Chicago office.[2] For the Bone Marrow challenge, the panel included experts from Stanford University's Haas Center for Public Service and students leading the grassroots 100KCheeks program for bone marrow registration drives.[3] For the Voting Accessibility challenge, the panel included experts from the Information Technology and Innovation Foundation, LA County, and IDEO.[4] For the Human Rights challenge, the panel consisted of four experts from Amnesty International.[5] And for the Web Entrepreneurs challenge, the panel included experts from the European Commission and local tech startup incubators and investing firms.[6]

Methods

For each of the challenges, we recruited at least six different expert raters who had at least 2–3 years of postgraduate work and/or industry experience that provided them relevant knowledge about the problem domain and prior solutions. Specifically, the qualifications and characteristics of the experts for each of the challenges were:

(1) **Vibrant Cities** ($N = 6$ experts): three Master's in City and Regional Planning students, with professional experience in community development work and historic preservation; one Master's in Architecture student, with professional experience in site development and civil engineering; one Facade Renovation Program manager at the Urban Redevelopment Authority (8+ years of experience); and one project development specialist at the Urban Redevelopment Authority, with a Master's in Public Management and 4+ years of experience.
(2) **Bone Marrow** ($N = 6$ experts): two PhD students in Behavioral Decision Sciences, with research focus on behavioral health interventions, nudges, and prosocial behavior; one donor search specialist at a large national bone marrow donor program, with 7+ years of experience; one Assistant Professor of Nursing, with extensive experience with bone marrow transplant patients; one graduate student in Public Health with 20 years of experience, including 5 years with a bone marrow transplant program; and one Outreach and Recruitment coordinator for a large international marrow matching organization.
(3) **Voting Accessibility** ($N = 6$ experts): two PhD students in Political Science; two PhD students in Assistive Technologies; one PhD student in Rehabilitation Science; one Master's in Public Administration student, who was personally visually impaired and an activist for accessibility.
(4) **Human Rights** ($N = 6$ experts): one master's student in International Human Rights; one Associate Director of Research at a Human Trafficking Center; one PhD student focused on International Human Rights Law, with an MA in Human Rights Studies; one Associate Director of Advocacy at a Human Trafficking Center, and an MA Candidate in International Administration; one Professor of Political Science; and one PhD student in Criminal Justice.
(5) **Web Entrepreneurs** ($N = 7$ experts): one adjunct faculty focused on Entrepreneurship; one master's student in Entrepreneurship; one MBA student who was a CEO and co-founder of a startup; one innovation consultant with a Master's in Entrepreneurship; one co-founder of startup who was a PhD student in Entrepreneurship and had an MA in Entrepreneurship; one Professor of Entrepreneurship, who was a former CIO and CTO at a large multinational tech incubator, with a PhD in Information Systems and an MBA; and one innovation specialist at a tech startup incubator.

The expert raters were each provided with the materials for just their focal domain problem: the challenge brief (with associated materials in the brief) and all concepts submitted for the challenge. They were instructed to rate each concept on the three dimensions of *NOVELTY*, *IMPACT*, and *FEASIBILITY*, using a Likert-like scale ranging from 1 (worst) to 6 (best). Raters spent an average of 4 min rating each concept. Table 1 shows how the dimensions were described and anchored for the raters.

Analyses used a mean rating across expert raters for each concept for each dimension. Table 2 shows examples of concepts whose mean expert ratings exemplified

Table 1. Rating Scale for Expert Rating of Ideas, with Anchors for the Scale Endpoints.

Dimension	Description	Scale Endpoints	Scale Endpoint Anchors
Novelty	How surprising was this concept to you? How different do you think it is from the "status quo"?	6	Took me completely by surprise, a really new and interesting concept
		1	Very obvious/standard solution and/or duplicate of something I already know about
Impact	Assuming it works (i.e. leaving aside implementation feasibility), how much potential for impact do you think this concept has? How well do you think this concept, if implemented, would solve the major dimensions of the problem?	6	Will make highly significant progress on multiple major dimensions of the problem
		1	Will not help with any of the major dimensions of the problem
Feasibility	How easy will it be to implement this concept, given your knowledge of current available resources? (intuition: how many barriers are there to implementation, and how hard would it be to overcome those barriers? fewer and easier barriers means higher feasibility)	6	Feasible to implement now in most settings where this problem exists
		1	Effectively impossible to implement now in any setting where this problem exists

Table 2. Example Concepts with Low and High Impact/Feasibility Scores Based upon Mean Expert Ratings from the "Voting Accessibility" Challenge. Concepts paraphrased/condensed from original descriptions.

		Feasibility	
		Low	High
Impact	Low	Represent each vote with a ball and fill a clear visible box with the balls (*Impact*=1.5; *Feasibility*=2.0)	Have children accompany their parents to the polling stations (*Impact*=2.3, *Feasibility*=4.0)
	High	Create a long-term project to research and prototype an ideal universal polling place *(Impact*=4.3, *Feasibility*=2.2)	Create and use secure online/mobile voting systems for people with accessibility issues *(Impact*=4.8, *Feasibility*=4.7)

combinations of low vs. high for impact and feasibility from the "Voting Accessibility" challenge. To illustrate concepts varying in novelty, a low-rated concept for the Voting challenge was *"Get a reminder in the mail to vote"* and a high-rated concept for the same challenge was *"Use near-field*

communications devices to automate configuring common settings for different accessibility needs."

Table 3 shows the inter-rater reliability of each measure for each challenge and summarized overall. In all cases, the raters were able to produce ratings with at least moderate reliability, and, as is typical in measures of complex real-world creativity, in no cases was the task so easy as to produce very high reliabilities. Indeed, it is not uncommon for creativity studies to use 10 or more raters to achieve high aggregate reliability, even with expert judges (Amabile, 1982; Kaufman, Baer, Cole, & Sexton, 2008). Features of the rating task that make it challenging include aggregating judgments over different aspects of the overall concept (e.g. some parts may be more novel or more feasible than others) and evaluating concepts that are not completely fleshed out (e.g. impact and feasibility judgments may require additional information).

Results & discussion

Table 4 shows that all three measures had meaningful variability. The intercorrelations were relatively small, which is consistent with the claim of separability of the constructs (including *IMPACT* and *FEASIBILITY*). We stress here that our claim is not that impact and feasibility

are completely independent, but rather that they are separable. Further, each of our three measures captured some important aspects of the *SHORTLIST* measure. The bivariate correlations reveal that shortlisted concepts are, on average, more novel, impactful, and feasible. However, there was substantial variability in the magnitude of the correlations: *IMPACT* was most predictive of *SHORTLIST*, $r = .26$, followed by *FEASIBILITY*, $r = .20$, and then *NOVELTY*, $r = .13$.

To explore how these constructs jointly relate to *SHORTLIST*, we fitted a generalized linear mixed model (GLMM) predicting concepts' *SHORTLIST* as a function

Table 3. Inter-Rater Reliability[a] for Each of the Measures, Overall and for Each of the Challenges

	Overall	Vibrant Cities	Bone Marrow	Voting Access	Human Rights	Web Entrepr.
NOVELTY	0.57	0.51	0.50	0.62	0.69	0.54
IMPACT	0.65	0.64	0.64	0.64	0.64	0.68
FEASIBILITY	0.70	0.67	0.71	0.75	0.67	0.70
N concepts	318	119	41	47	62	49
N raters		6	6	6	6	7

[a]Intra-class correlation (ICC), Type II, consistency.

Table 4. Mean, Standard Deviation, Range, and Bivariate Correlations for Creativity Measures

			Correlations		
Variable	M *(SD)*	Min-Max	*NOVELTY*	*IMPACT*	*FEASIBILITY*
SHORTLIST	.12 *(.33)*	0–1	.13*	.26***	.20**
NOVELTY	3.2 *(0.8)*	1–6		.25***	.00
IMPACT	3.2 *(0.8)*	1–6			.29***
FEASIBILITY	3.5 *(0.9)*	1–6			–

Note: * $p < .05$; ** $p < .01$; *** $p < .001$.

of their *IMPACT*, *FEASIBILITY*, and *NOVELTY*, aggregated across experts, with random effects of challenge (i.e. a categorical variable to indicate which OpenIDEO challenge a given concept was a part of) and author (i.e. a categorical variable to indicate which participant authored the concept since participants could author multiple concepts), to account for potential within-challenge and within-author statistical dependencies. Consistent with the OpenIDEO team's (2013) claim that the potential for impact was the most significant (but not the only) consideration for being shortlisted, a concept's estimated *IMPACT* was the most predictive of being shortlisted, $\gamma_{10} = 0.88$, 95% CI = [0.37, 1.42], $p < .001$ followed by *FEASIBILITY*, $\gamma_{20} = 0.56$ [0.13, 1.02], $p = .01$. Interestingly, *NOVELTY* was *not* a significant independent predictor of being shortlisted, $\gamma_{30} = 0.39$ [−0.12, 0.90], $p = .14$, conditional on impact and feasibility. Perhaps, novelty is helpful to produce ideas that can be impactful and feasible, but itself is not worth valuing in selecting ideas for implementation – we turn to the issue of process in Study 2. This overall model is a good fit for the data, improving fit over a baseline null model (with just random effects of challenge and author), $\chi^2(3) = 30.17$, $p < .001$.

Study 2: Is separating impact and feasibility psychologically significant?

Our second objective is to demonstrate that separating impact and feasibility is psychologically significant. To do this, we explore the extent to which these constructs have different relations to two psychological phenomena that are often related to creativity: analogy and conceptual combination. We discuss these constructs in the following paragraphs, along with how they might interact differently with the constructs of impact and feasibility.

Analogies are previous/existing ideas that share relational similarities with a target problem (Gentner, 1983; Gick & Holyoak, 1980). They have been shown to be an important source of creative ideas, both in analyses of real-world creativity (Chan & Schunn, 2015; Dunbar, 1997; Gentner et al., 1997; Holyoak & Thagard, 1996) and in experimental observations of creative ideation (Chan et al., 2011; Dahl & Moreau, 2002). Early theoretical work (Gentner & Markman, 1997; Holyoak & Thagard, 1996; Poze, 1983; Ward, 1998) argued that conceptually *far* analogies are especially helpful for creativity, and many studies have shown positive effects of far analogies, at least for novelty (Chan et al., 2011; Chiu & Shu, 2012; Dahl & Moreau, 2002; Hender, Dean, Rodgers, & Jay, 2002). By far, we mean analogies that share relational similarities but differ on a number of other "surface" attributes that are not relevant to the core relational mapping (e.g. atom and solar system, which share relational similarities w.r.t. the schema of a core body being orbited by satellite entities but differ in the attribute of scale). However, a growing number of studies (Chan, Dow, & Schunn, 2015; Dunbar, 1997; Fu et al., 2013; Gonçalves, Cardoso, & Badke-Schaub, 2013; He & Luo, 2017; Madjar, Shalley, & Herndon, n.d.) suggest that far analogies actually lead to *less* creative ideas. However, these studies sometimes did not measure appropriateness explicitly, or conflated impact and feasibility subdimensions in their measurement of appropriateness. Here, we explore whether separating impact and feasibility gives us further insight into the relationship between conceptual distance of analogies and creative output. It is possible that far analogies produce *less feasible* (by virtue of more technical gaps in supporting components that are not yet resolved), but potentially *more impactful*, ideas (in which case the risk of pursuing these ideas is worth accepting). Alternatively, far analogies could in fact lead to *less impactful* ideas as well by relying on approaches that are generally better for other purposes, even if they are feasible (in which case greater caution in pursuing these ideas is warranted). However, it could also be that far analogies just tend to lead to less feasible *and* less impactful ideas. If separating impact and feasibility is psychologically significant, we should be able to detect

these distinct patterns of relationship of conceptual distance of analogical sources with impact vs. with feasibility.

Conceptual combination is a psychological process in which people combine or integrate two or more ideas into a new idea with properties that cannot be reduced to its constituent ideas (e.g. "fusion" cooking or "mash-ups" and hip-hop sampling in music). This process has also been shown to be an important source of creative ideas (Ward, 2001). Here, too, conceptual distance is a potentially important variable: prominent theories of creativity propose that conceptually far combinations are more likely to lead to creative ideas (Blasko & Mokwa, 1986; Koestler, 1964; Mednick, 1962; Rothenberg, 1979). Empirical work mostly confirms that conceptually far combinations do tend to improve novelty of ideas (Doboli, Umbarkar, Subramanian, & Doboli, 2014; Gielnik, Frese, Graf, & Kampschulte, 2011; Mobley, Doares, & Mumford, 1992; Nagai, Taura, & Mukai, 2009; Wilkenfeld, 1995; Wilkenfeld & Ward, 2001; Wisniewski, 1997), but effects on appropriateness are less clear, with some showing negative effects (Baughman & Mumford, 1995; Mobley, Doares, & Mumford, 1992), and others showing no effect on appropriateness (Doboli, Umbarkar, Subramanian, & Doboli, 2014) or aggregate creativity (Chan & Schunn, 2015). Here, we explore whether separating impact and feasibility can give us further insight into the relationship between far conceptual combinations and creative output. For example, do far combinations lead to increased risk of worse ideas, ideas that are difficult to implement (but still potentially impactful), or some other configuration of effects?

In sum, our research questions in this study are as follows:

(1) Does conceptual distance of inspiration sources from the problem domain have statistically different patterns of relationship to the impact vs. feasibility of resulting ideas?
(2) Does conceptual combination distance of inspiration sources have statistically different patterns of relationship to the impact vs. feasibility of resulting ideas?

We emphasize that our goal here is *not* to test *a priori* hypotheses about the effects of far analogies and conceptual combinations. Alternative predictions about the direction or magnitude of statistical relationships could be made, given the general nature of current theories and lack of prior empirical work on this divide. Instead, the purpose is simply to conduct exploratory statistical tests of whether separating impact and feasibility reveals meaningfully distinct patterns of relationships to conceptual distance of analogical sources and to conceptual distance of conceptual combinations. If such differential relationships are found, then this provides important evidence that separating impact and feasibility is psychologically significant.

Methods

Inspiration sources for concepts

The OpenIDEO dataset provides a convenient way for us to explore how properties of the inspiration sources for concepts are predictive of the concepts' creativity (or novelty, impact, and feasibility): all challenges on OpenIDEO included an inspiration phase (now called the research phase), where contributors post *inspirations* meant to provide raw materials (e.g. case studies of stakeholders and descriptions of solutions to analogous problems) for generating later concepts. During the formal phase for collecting concepts, OpenIDEO contributors are strongly encouraged to cite inspirations they built on to generate their concepts. This encouragement came in the form of norms (e.g. explicit instructions for the concepting phase and comments on posts from the OpenIDEO facilitators encouraging participants to cite inspirations) and the structure of the challenges, as well as explicit instructions and affordances in the user interface for uploading and editing ideas. Our sample includes 555 unique inspirations that were cited across the 5 challenges.

Measuring conceptual distance

Because of the large number of needed similarity judgments (e.g. 153,735 judgments for all possible pairwise comparisons between inspirations), we used a computational approach to measure conceptual distance. Specifically, we built a Latent Dirichlet Allocation (LDA) probabilistic topic model (Blei, Ng, Jordan, & Lafferty, 2003)—which is an unsupervised machine learning technique for learning topical structures from unstructured text – to capture how problem statements and inspirations differ in their topical compositions.

The LDA topic model jointly learns the latent topical structure (i.e. the presence of topics defined as statistical distributions over words) within a corpus of documents, along with the estimated topical composition (defined as a statistical distribution over the set of learned topics) of each document in the corpus. The approach

represents the topical composition of each document as a vector of weights for each of the learned topics. These dimensions can then be used to compare documents or words based on how similar or different they are with respect to these dimensions, by computing the Euclidean dot product of their respective topic weight vectors. This dot product – also called "cosine similarity"[7] — ranges from −1, for an identical topical composition, to 1, for a very different topical composition.

The LDA computational approach is analogous to the method of Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997), which is more commonly used in psychological research, in that both approaches rely on distributional semantics or "words of a feather flock together." That is, both methods essentially estimate a latent set of semantic dimensions in a corpus from co-occurrence patterns of words within and across documents. Here, we selected LDA for its advantages in terms of interpretability and coherence of topical dimensions (for a recent demonstration of this on scientific documents, see Bellaouar, Bellaouar, & Ghada, 2021), which we used to guide our development of the model. LDA has also previously been validated for capturing semantics in other settings (Griffiths & Steyvers, 2004; Griffiths, Steyvers, & Tenenbaum, 2007; Schwartz et al., 2013).

The dataset we used to construct this LDA model consisted of all documents in the full OpenIDEO dataset, i.e. 2,341 concepts, 4,557 inspirations, and 12 challenge briefs (6,910 total documents). The documents were preprocessed by splitting the documents into tokens using the TreeBank Tokenizer from the open-source Natural-Language Toolkit Python library and removing stopwords (e.g. "the" and "this") using the open-source MAchine Learning for LanguagE Toolkit (MALLET) (McCallum, 2002) stopword list. We then used MALLET to train an LDA model on this corpus with 400 topics, with default parameters (asymmetric priors for topic-document and topic-word distributions, optimized using MALLET's in-package optimization option). With the resulting learned topical dimensions, we then compute cosine similarities between all pairs of documents.

We applied two transformations to the cosine similarity scores before running our analysis. First, to enable comparability of cosine scores across challenges, we normalized the cosines within each challenge as z-scores. We did this because mean conceptual distance of inspirations from the problem domain (but not between inspirations within a domain) varied significantly across challenges Second, because we are interested in conceptual *distance* as a construct, rather than similarity, we subtracted the cosine similarity z-scores from 0, so that larger numbers indicate *more distance* between two documents. This enables us to interpret results in terms of the effects of distance (higher is more distant) rather than similarity. We call this transformed measure the cosine distance between two documents.

With these transformations, we then operationalize conceptual distance as follows:

- *ANALOGY-DIST* (conceptual distance of an analogical source to the challenge domain) is the mean of the cosine *distances* between the challenge brief document and each inspiration cited by the concept.
- *COMBIN-DIST* (conceptual distance between conceptual combinations of concepts) is the mean of the cosine *distances* between each pair of inspirations cited by the concept.

While it may seem unusual and inappropriate to measure analogical distance with an overall computational semantic distance measure, we believe this approach is reasonable in this case for a few reasons. First, the analogical distance is multifaceted: what aspects of a problem or idea count as "structural" for computation of structural similarity (Gentner & Markman, 1997) is not an inherent property of the items but is jointly determined by the context and the nature of the pair of items being mapped (Holyoak & Thagard, 1989). In this specific setting, too, contributors to the platform have already placed a floor on the structural distance because they have already made a mapping to the problem; on the margin, then, overall semantic distance is a reasonable proxy for the dimension of "surface dissimilarity" that is central to the concept of analogical distance.

As further validation, previous work using this LDA approach applied to OpenIDEA data (omitted for review) observed a substantial ($r = .53$) correlation between the cosine distances and human Likert-like judgments of conceptual distance between a subset ($N = 544$) of document pairs. Notably, this correlation was larger than the highest correlation between individual human judges' ratings of conceptual distance ($r = .48$).

To give intuition for the distance measure, consider these examples of inspirations from the Electronic Waste challenge that were near/far from their challenge brief document (descriptions condensed/paraphrased):

- Near *ANALOGY-DIST* (cosine distance =.37): *Bicycle turns e-waste into copper (describes a self-contained portable cable recycling unit in Africa that recycles copper from cables)*
- Far *ANALOGY-DIST* (cosine distance =.99): *Emotional Accounting (describes how people value possessions*

not just in terms of their "fair value" but also in terms of their "affective tag," i.e., how it makes them feel, and both of these valuations influence whether people dispose of their possessions)

Analyses

Because concepts are clustered within authors (who can post multiple concepts) and within challenges (each challenge accepts many concepts), we use a linear mixed model to ensure that estimates of the effects of ANALOGY-DIST and COMBIN-DIST on impact and feasibility are not artifacts of random effects of challenge (are some challenges more difficult?) and author (are some authors more creative?). We fit models using the lme4 package's (Bates, Maechler, Bolker, & Walker, 2013) lmer function in R with default settings, including restricted maximum likelihood (REML) estimation of the model parameters.

All concepts had the opportunity to receive feedback (in the form of comments) and update their concepts in response to feedback. Feedback and iteration are often positive contributors to the creativity of concepts; for example, suggestions and ideas can be incorporated into ideas to improve their creativity (Chan, Dang, & Dow, 2016), feasibility, or diversity (Dow et al., 2010); comments and feedback can also motivate more effort, which can improve creative performance (Roy, Gauvin, & Limayem, 1996). We therefore include FEEDBACK—the number of comments received from others (i.e. excluding author responses to comments)—as a control variable in all statistical models.

Results

Far inspirations tend to yield useless ideas

Table 5 shows analogy-distance model results for each of the dependent measures. ANALOGY-DIST has no statistically significant effect on NOVELTY (although the coefficient leans in the predicted, positive, direction), and no statistically significant effect on FEASIBILITY. The primary finding is that ANALOGY-DIST has a statistically significant *negative* relationship with IMPACT, such that an increase of 1 standard deviation in inspiration distance predicts a decrease of approximately 0.14 points on expert-rated IMPACT. FEEDBACK has a significant positive relationship with all measures.

Far combinations tend to yield impossible ideas

In this set of analyses, our effective N drops to 200 concepts because not all concepts cited at least 2 inspirations. Table 6 shows the results of our models for each of the dependent measures. COMBIN-DIST has no statistically significant relationship with NOVELTY (although the coefficient leans in the predicted, positive, direction), and no statistically significant relationship with IMPACT (although the coefficient leans in the negative direction). The primary finding is that COMBIN-DIST has a statistically significant

Table 5. Model Estimates and Fit Statistics for Cross-Classified Multilevel Linear Regressions of *NOVELTY*, *IMPACT*, and *FEASIBILITY* on *ANALOGY-DIST*

	NOVELTY	IMPACT	FEASIBILITY
Fixed effects			
ANALOGY-DIST	0.04 [−0.07, 0.15]	−0.14* [−0.25, −0.03]	−0.01 [−0.14, 0.12]
FEEDBACK	0.01* [0.00, 0.02]	0.02*** [0.01, 0.03]	0.02*** [0.01, 0.03]
Intercept	3.15[2.96, 3.33]	3.06[2.95, 3.18]	3.33[3.21, 3.47]
Model fit statistics			
Deviance	708.10	717.70	809.90
Reduction from null	6.19*	26.47***	12.14***

Note:: ^{m} $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$; 95% CI (Wald) = [lower, upper].

Table 6. Model Estimates and Fit Statistics for Cross-Classified Multilevel Linear Regressions of *NOVELTY*, *IMPACT*, and *FEASIBILITY* on *COMBIN-DIST*

	NOVELTY	IMPACT	FEASIBILITY
Fixed effects			
COMBIN-DIST	0.01 [−0.09, 0.09]	−0.07 [−0.17, 0.02]	−0.15** [−0.25, −0.05]
ANALOGY-DIST	0.08 [−0.10, 0.25]	−0.22* [−0.41, −0.04]	−0.15 [−0.37, 0.06]
FEEDBACK	0.01 [−0.00, 0.02]	0.02** [0.01, 0.03]	0.01* [0.00, 0.02]
Intercept	3.17[2.90, 3.44]	3.02[2.79, 3.25]	3.07[2.79, 3.32]
Model fit statistics			
Deviance	427.56	444.36	482.69
Reduction from baseline	0.00	2.56	8.03**
Reduction from null	2.61	17.89***	16.94***

Note:^{m} $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$; 95% CI (Wald) = [lower, upper].

negative relationship with FEASIBILITY, such that an increase of 0.10 units in COMBIN-DIST (on a scale of 0 to 1) predicts a decrease of 0.15 points on expert-rated FEASIBILITY.

General discussion

Summary and interpretation of findings

In this paper, we have argued that it is theoretically important to separate impact and feasibility when examining changes in overall appropriateness during creative cognition. Our data provide two points of support for this argument. First, we show that impact and feasibility are predictors of a face-valid aggregate creativity measure (shortlisting of ideas by an expert panel on OpenIDEO) but are also statistically separable, with only a modest correlation between the two measures ($r = .29$). Second, we show that separating these sub-dimensions of appropriateness

reveals psychologically distinct patterns of creative phenomena (see Figure 2): far inspirations yield *useless* ideas (i.e. negative effects on impact), while far combinations yield *impossible* ideas (i.e. negative effects on feasibility). This suggests that the distinction between impact and feasibility is psychologically significant when it comes to distinguishing potential impacts of analogical distance and combination distance of inspiration sources for creative outcomes.

We hypothesize that other theoretically important questions will gain empirical clarity by distinguishing impact and feasibility. For example, separating impact and feasibility may allow us to clarify the intuitive notion of "useful bad ideas" in brainstorming or other creative processes (Dix et al., 2006): we argue that these ideas are high in potential impact but low in feasibility (hence "bad") but can nevertheless spur forward creative progress. To illustrate, consider Gruber and Barrett (1974)'s close analysis of Darwin's creative process from his detailed notebooks, which identified several prior theories that were potentially impactful but technically incorrect. For example, the theory of coral formation aimed to explain a core puzzle that Darwin's theory of evolution by natural selection ultimately explained, but was incorrect in its entirety. Gruber and Barrett argued that this theory provided a crucial intuition of struggle and branching. Another example was the monad theory of evolution, which provided a crucial intuition of an irregular branching tree of nature (Gruber & Barrett, 1974). Gruber and Barrett called these useful bad ideas "by-productive thinking" because they seemed to be "important steps toward the theory of evolution through natural selection being taken as by-products of efforts that seemed to move in other directions" (Gruber & Barrett, 1974, p. 111). Similar patterns of by-productive thinking (and their antecedents) may be possible to study in other creative settings by carefully distinguishing impact and feasibility. There may also be useful connections here between this conception of "useful bad ideas" and Corazza's dynamic creativity framework (Corazza, 2016; Corazza, Agnoli, & Mastria, 2022), which highlights the importance of "inconclusive outcomes" to the creative process, such as

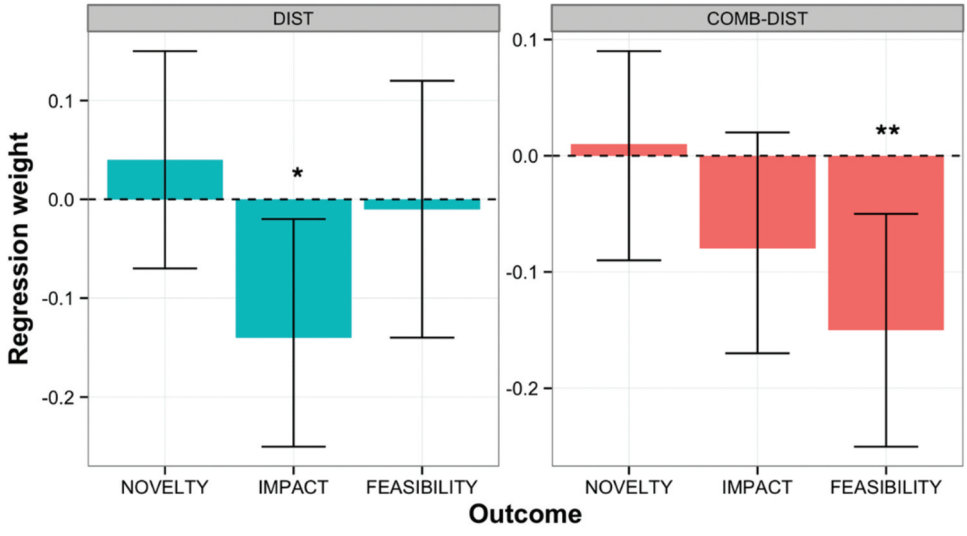


Figure 2. Summary of results. Analogical distance has a significant negative relationship with impact (left panel), while combination distance has a significant negative relationship with feasibility (right panel). * $p < .05$; ** $p < .01$.

Edison's "failed" prototypes or Van Gogh's "scrapped" sketches.

As another example, Weisberg (2015) argues that because appropriateness can be quite unstable over time (e.g. people becoming less creative over time or becoming more creative during one's lifetime after death), creativity research might be better off narrowing the research focus to understanding the people and processes that reliably produce *intentional* novelty. We believe this restriction gives away too much and argue instead that separating impact and feasibility can address at least some of these criticisms of the construct of appropriateness and its place in the definition of creative people/products. For example, feasibility (because of its conceptualization) might be easier to gain stable consensus on (in contrast to impact, which might involve uncertain predictions of the success of an idea) ideas that are both novel and feasible may be more likely to be creative (i.e. also impactful), as Uzzi, Mukherjee, Stringer, and Jones (2013) found.

Based on these empirical results and theoretical implications, we recommend that future research on the psychology of creativity makes clear distinctions between impact and feasibility.

Limitations and future work

Our evidence for the separability of impact and feasibility was obtained from the analysis of a wide range of OpenIDEO's concepts across diverse, real-world problems, but these are largely in the domain of design. Additional research is needed to formally test broader uses. We expect that a similar breakdown (e.g. significant vs. technically correct) is both possible and desirable for other domains, such as mathematics, scientific discovery, and the arts. Furthermore, the OpenIDEO setting included a wide variety of problem domains, lending some empirical warrants to hypothesizing generalizability of our measurement approach.

Relatedly, the most natural scope of generalization for our claims about creative products is to creative ideas or concepts since that is what we directly test empirically. However, we expect that distinguishing impact and feasibility will still be theoretically useful and psychologically significant for other forms of creative products, but may be more difficult to measure reliably or require adaptation of measurement procedures. For example, impact and feasibility may be less reliable for earlier stage creative products, such as "idea fragments" (Sosa, 2019), or preinventive structures or sketches (Finke, Ward, & Smith, 1996), given that they may be by necessity more vague and unformed (Suwa & Tversky, 1997).

Methodologically, there is a potential bias in our separability results due to our rating procedure since we explicitly asked experts to assess impact without considering feasibility (and vice versa). While our experts did raise concerns about some aspects of the rating procedure (e.g. requesting a change in anchoring from 1-to-6 to 6-to-1), they reported finding the distinction between impact and feasibility to be quite natural. Nevertheless, we acknowledge that a more stringent test of the separability claims would be to ask for ratings of impact and feasibility without explicitly asking them to ignore the other dimensions.

We acknowledge that our inter-rater reliabilities and effect sizes, while meaningful, were relatively low. Both of these issues are not uncommon in field research, representing a tradeoff when moving from tightly controlled lab settings to messier, naturalistic settings; however, we do think it is important to complement our externally valid but somewhat noisy evidence with more precisely measured evidence. For instance, it may be possible to achieve higher levels of reliability in measuring novelty/impact/feasibility by increasing the number of expert raters to 8–10 or more; for example, in Amabile's original studies with the Consensual Assessment Technique, 10 or more judges were typically required to achieve high levels of reliability (Amabile, 1983), though this may require simpler problems for which it is easier to find experts (vs. the more complex, expertise-intensive real-world problems we studied). It may also be useful to more precisely measure conceptual or combination distance (e.g. through controlled stimuli measured exclusively using human judgments) rather than measuring them at scale using computational text analysis methods.

Additionally, our evidence for the psychological significance of separating impact and feasibility comes primarily from a correlational study that is able to identify relationships, not causal effects. However, discovering causal effects was not our primary goal: we aimed to test whether measuring impact and feasibility could reveal psychological relationships that are meaningful for creativity but impossible/difficult without distinguishing impact and feasibility. We look forward to future work that uses separate measures of impact and feasibility to bring new empirical clarity to the psychology of creativity.

Notes

1. https://commonfund.nih.gov/highrisk.

2. https://web.archive.org/web/20120517125857mp_/http://www.openideo.com/fieldnotes/openideo-team-notes/vibrantcitiesrefinement/.
3. https://web.archive.org/web/20120522231743mp_/http://www.openideo.com/fieldnotes/openideo-team-notes/evaluation-and-the-bone-marrow-donation-challenge.
4. https://web.archive.org/web/20120523003400mp_/http://www.openideo.com/fieldnotes/openideo-team-notes/announcing-our-voting-challenge-shortlist/.
5. https://web.archive.org/web/20120519033623mp_/http://www.openideo.com/fieldnotes/openideo-team-notes/welcome-to-the-amnesty-refinement-phase/.
6. https://web.archive.org/web/20120515061010mp_/http://www.openideo.com/fieldnotes/openideo-team-notes/web-start-up-shortlist/.
7. Cosine similarity is related to, but distinct from, the cosine trigonometric function: smaller cosine angles between two vectors correlate with values closer to 1 (more similar, pointing in the same direction).

Acknowledgments

We thank anonymous reviewers for thoughtful and constructive feedback and our expert raters for contributing their expertise.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by an NSF Doctoral Dissertation Research in Science of Science and Innovation Policy Grant #1360013 to the first author.

ORCID

Joel Chan http://orcid.org/0000-0003-3000-4160

References

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, *43*(5), 997–1013. doi:10.1037/0022-3514.43.5.997

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, *45*(2), 357. doi:10.1037/0022-3514.45.2.357

Anderson, J. D. (2004). *Inventing flight: the wright brothers & their predecessors*. Baltimore, MD: JHU Press.

Arts, S., & Fleming, L. (2018). Paradise of novelty—or loss of human capital? ExplorIng new fields and inventive output. *Organization Science*, *29*(6), 1074–1092. doi:10.1287/orsc.2018.1216

Ball, L. J., Ormerod, T. C., & Morley, N. J. (2004). Spontaneous analogising in engineering design: A comparative analysis of experts and novices. *Design Studies*, *25*(5), 495–508. doi:10.1016/j.destud.2004.05.004

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *Lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-5*. http://CRAN.R-project.org/package=lme4

Baughman, W. A., & Mumford, M. D. (1995). Process-analytic models of creative capacities: Operations influencing the combination-and-reorganization process. *Creativity Research Journal*, *8*(1), 37–62. doi:10.1207/s15326934crj0801_4

Bellaouar, S., Bellaouar, M. M., & Ghada, I. E. (2021). Topic modeling: CompaRison of lsa and lda on scientific publications. *2021 4th International Conference on Data Storage and Data Engineering*, 59–64. 10.1145/3456146.3456156

Bishop, B. (2022). *National Ignition Facility achieves fusion ignition*. https://www.llnl.gov/news/national-ignition-facility-achieves-fusion-ignition

Blasko, V. J., & Mokwa, M. P. (1986). Creativity in advertising: A Janusian perspective. *Journal of Advertising*, *15*(4), 43–72. doi:10.1080/00913367.1986.10673037

Blei, D. M., Ng, A. Y., Jordan, M. I., & Lafferty, J. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms*. New York, NY: Routledge.

Bonnardel, N., & Marmèche, E. (2005). Towards supporting evocation processes in creative design: A cognitive approach. *International Journal of Human-Computer Studies*, *63*(4–5), 422–435. doi:10.1016/j.ijhcs.2005.04.006

Boudreau, K., Guinan, E., Lakhani, K. R., & Riedl, C. (2014). *LookIng across and looking beyond the knowledge frontier: Intellectual distance and resource allocation in science* (SSRN Scholarly Paper ID2478627). Social Science Research Network. 10.2139/ssrn.2478627

Chan, J., Dang, S., & Dow, S. P. (2016). ImproVing crowd innovation with expert facilitation. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 1223–1235. 10.1145/2818048.2820023

Chan, J., Dow, S. P., & Schunn, C. D. (2015). Do the best design ideas (really) come from conceptually distant sources of inspiration? *Design Studies*, *36*, 31–58. doi:10.1016/j.destud.2014.08.001

Chan, J., Fu, K., Schunn, C., Cagan, J., Wood, K. L., & Kotovsky, K. (2011). On the benefits and pitfalls of analogies for innovative design: Ideation performance based on analogical distance, commonness, and modality of examples. *Journal of Mechanical Design*, *133*(8), 081004. doi:https://doi.org/10.1115/1.4004396

Chan, J., & Schunn, C. (2015). The impact of analogies on creative concept generation: Lessons from an in vivo study in engineering design. *Cognitive Science*, *39*(1), 126–155. doi:10.1111/cogs.12127

Chan, J., & Schunn, C. D. (2015). The importance of iteration in creative conceptual combination. *Cognition*, *145*, 104–115. doi:10.1016/j.cognition.2015.08.008

Chiu, I., & Shu, L. H. H. (2012). Investigating effects of oppositely related semantic stimuli on design concept

creativity. *Journal of Engineering Design*, *23*(4), 271–296. doi:10.1080/09544828.2011.603298

Corazza, G. E. (2016). PotEntial originality and effectiveness: The dynamic definition of creativity. *Creativity Research Journal*, *28*(3), 258–267. doi:10.1080/10400419.2016.1195627

Corazza, G. E., Agnoli, S., & Mastria, S. (2022). THe dynamic creativity framework. *European Psychologist*, *27*(3), 191–206. doi:10.1027/1016-9040/a000473

Csikszentmihalyi, M. (1997). *Flow and the psychology of discovery and invention*. New York, NY: HarperPerennial.

Dahl, D. W., & Moreau, P. (2002). The Influence and value of analogical thinking during new product ideation. *Journal of Marketing Research*, *39*(1), 47–60. doi:10.1509/jmkr.39.1.47.18930

Dix, A., Sas, C., Gomes da Silva, P., McKnight, L., Ormerod, T., & Twidale, M. (2006). Why bad ideas are a good idea. HCIEd'06.

Doboli, A., Umbarkar, A., Subramanian, V., & Doboli, S. (2014). Two experimental studies on creative concept combinations in modular design of electronic embedded systems. *Design Studies*, *35*(1), 80–109. doi:10.1016/j.destud.2013.10.002

Dow, S. P., Glassco, A., Kass, J., Schwarz, M., Schwartz, D. L., & Klemmer, S. R. (2010). PaRallel prototyping leads to better design results, more divergence, and increased self-efficacy. *ACM Transactions on Computer-Human Interaction*, *17*(4), 1–24. doi:https://doi.org/10.1145/1879831.1879836

Dunbar, K. N. (1997). How scientists think: On-line creativity and conceptual change in science. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 461–493). Washington, DC: American Psychological Association.

Finke, R. A., Ward, T. B., & Smith, S. M. (1996). *Creative cognition: Theory, research, and applications*. Cambridge, MA: MIT Press.

Fu, K., Chan, J., Cagan, J., Kotovsky, K., Schunn, C., & Wood, K. (2013). The meaning of near and far: the impact of structuring design databases and the effect of distance of analogy on design output. *Journal of Mechanical Design*, *135*(2), 021007. doi:10.1115/1.4023158

Gentner, D. (1983). StruCture mapping: a theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170. doi:10.1207/s15516709cog0702_3

Gentner, D., Brem, S., Ferguson, R. W., Wolff, P., Markman, A. B., & Forbus, K. D. (1997). Analogy and creativity in the works of johannes kepler. In T. B. Ward, J. Vaid, & S. M. Smith (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 403–459). Washington, DC: American Psychological Association.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *The American Psychologist*, *52*(1), 45–56. doi:10.1037/0003-066X.52.1.45

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*(3), 306–355. doi:10.1016/0010-0285(80)90013-4

Gielnik, M. M., Frese, M., Graf, J. M., & Kampschulte, A. (2011). Creativity in the opportunity identification process and the moderating effect of diversity of information. *Journal of Business Venturing*, *27*(5), 559–576. doi:10.1016/j.jbusvent.2011.10.003

Gonçalves, M., Cardoso, C., & Badke-Schaub, P. (2013). Inspiration peak: Exploring the semantic distance between design problem and textual inspirational stimuli. *International Journal of Design Creativity and Innovation*, *1*(4), 215–232. doi:10.1080/21650349.2013.799309

Gorder, W. D. (1980). DivergEnt production abilities as constructs of musical creativity. *Journal of Research in Music Education*, *28*(1), 34–42. doi:10.2307/3345051

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, *Suppl 101*(suppl_1), 5228–5235. doi:10.1073/pnas.0307752101

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211–244. doi:https://doi.org/10.1037/0033-295X.114.2.211

Gruber, H. E., & Barrett, P. H. (1974). In E. P. Dutton (Ed.), *Darwin on man: A psychological study of scientific creativity*. Chicago, IL: University of Chicago Press.

Guilford, J. P. (1950). Creativity. *The American Psychologist*, *5*(9), 444–454. doi:10.1037/h0063487

He, Y., & Luo, J. (2017). The novelty 'sweet spot' of invention. *Design Science*, *3*. doi:10.1017/dsj.2017.23

Hender, J. M., Dean, D. L., Rodgers, T. L., & Jay, F. F. (2002). An examination of the impact of stimuli type and gss structure on creativity: brainstorming versus non-brainstorming techniques in a gss environment. *Journal of Management Information Systems*, *18*(4), 59–85. doi:10.1080/07421222.2002.11045705

Holyoak, K. J., & Thagard, P. (1989). AnaloGical mapping by constraint satisfaction. *Cognitive Science*, *13*(3), 295–355. doi:10.1207/s15516709cog1303_1

Holyoak, K. J., & Thagard, P. (1996). *Mental Leaps: Analogy in Creative Thought*. Cambridge, MA: MIT Press.

Jansson, D. G., & Smith, S. M. (1991). Design fixation. *Design Studies*, *12*(1), 3–11. doi:10.1016/0142-694X(91)90003-F

Kaufman, J. C., Baer, J., Cole, J. C., & Sexton, J. D. (2008). A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal*, *20*(2), 171–178. doi:10.1080/10400410802059929

Koestler, A. (1964). *The act of creation*. New York, NY: Macmillan.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. doi:10.1037/0033-295X.104.2.211

Linsey, J. S., Tseng, I., Fu, K., Cagan, J., Wood, K. L., & Schunn, C. D. (2010). A Study of design fixation, its mitigation and perception in engineering design faculty. *Journal of Mechanical Design*, *132*(4), 041003. doi:10.1115/1.4001110

MacCrimmon, K. R., & Wagner, C. (1994). Stimulating ideas through creative software. *Management Science*, *40*(11), 1514–1532. doi:10.1287/mnsc.40.11.1514

Madjar, N., Shalley, C. E., & Herndon, B. (n.d.). Taking time to incubate: The moderating role of 'what you do' and 'when you do it' on creative performance. *The Journal of Creative Behavior*, *0*(0). doi:10.1002/jocb.362

Marsh, R. L., & Bower, G. H. (1993). Eliciting cryptomnesia: Unconscious plagiarism in a puzzle task. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *19*(3), 673–688. doi:10.1037/0278-7393.19.3.673

Marsh, R. L., & Landau, J. D. (1995). Item availability in cryptomnesia: Assessing its role in two paradigms of unconscious plagiarism. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *21*(6), 1568–1582. doi:10.1037/0278-7393.21.6.1568

Marsh, R. L., Landau, J. D., & Hicks, J. L. (1997). Contributions of inadequate source monitoring to unconscious plagiarism during idea generation. *Journal of Experimental Psychology Learning, Memory, and Cognition*, *23*(4), 886–897. doi:10.1037/0278-7393.23.4.886

McCallum, A. K. (2002).

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, *69*(3), 220–232. doi:10.1037/h0048850

Mobley, M. I., Doares, L. M., & Mumford, M. D. (1992). Process analytic models of creative capacities: Evidence for the combination and reorganization process. *Creativity Research Journal*, *5*(2), 125–155. doi:10.1080/10400419209534428

Nagai, Y., Taura, T., & Mukai, F. (2009). Concept blending and dissimilarity: Factors for creative concept generation process. *Design Studies*, *30*(6), 648–675. doi:10.1016/j.destud.2009.05.004

Orbey, E. (2019). ThE international federation of gymnastics needs to keep up with Simone Biles | the New Yorker. *The New Yorker*. https://www.newyorker.com/sports/sporting-scene/the-international-federation-of-gymnastics-needs-to-keep-up-with-simone-biles

Pichot, N., Bonetto, E., Pavani, J. -B., Arciszewski, T., Bonnardel, N., & Weisberg, R. W. (2022). The construct validity of creativity: EmpirIcal arguments in favor of novelty as the basis for creativity. *Creativity Research Journal*, *34*(1), 2–13. doi:10.1080/10400419.2021.1997176

Poze, T. (1983). Analogical connections: The essence of creativity. *The Journal of Creative Behavior*, *17*(4), 240–258. doi:10.1002/j.2162-6057.1983.tb00359.x

Puryear, J. S., & Lamb, K. N. (2020). Defining creativity: How far have we come since plucker, beghetto, and Dow? *Creativity Research Journal*, *32*(3), 206–214. doi:10.1080/10400419.2020.1821552

Rhodes, M. (1961). An Analysis of Creativity. *Phi Delta Kappan*, *42*(7), 305–310.

Rothenberg, A. (1979). *The emerging goddess: The creative process in art, science, and other fields*. Chicago, IL: University of Chicago Press.

Roy, M. C., Gauvin, S., & Limayem, M. (1996). ElectrOnic group brainstorming: The role of feedback on productivity. *Small Group Research*, *27*(2), 215–247. doi:10.1177/1046496496272002

Runco, M. A., & Jaeger, G. J. (2012). ThE standard definition of creativity. *Creativity Research Journal*, *24*(1), 92–96. doi:10.1080/10400419.2012.650092

Sandberg, J., & Alvesson, M. (2011). Ways of constructing research questions: Gap-spotting or problematization? *Organization*, *18*(1), 23–44. doi:10.1177/1350508410372151

Sawyer, R. K. (2012). *Explaining creativity: The science of human innovation* (2nd ed.). Oxford, United Kingdom: Oxford University Press.

Schunn, C. D., & Dunbar, K. N. (1996). Priming, analogy, and awareness in complex reasoning. *Memory & Cognition*, *24*(3), 271–284. doi:10.3758/BF03213292

Schwartz, A. H., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M. … Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The Open-Vocabulary approach. *PLoS One*, *8*(9), e73791. doi:10.1371/journal.pone.0073791

Sgourev, S. V. (2013). How Paris gave rise to cubism (and Picasso): Ambiguity aNd fragmentation in radical innovation. *Organization Science*, *24*(6), 1601–1617. doi:10.1287/orsc.1120.0819

Shah, J. J., Vargas-Hernandez, N., & Smith, S. M. (2003). Metrics for measuring ideation effectiveness. *Design Studies*, *24*(2), 111–134. doi:10.1016/S0142-694X(02)00034-0

Simonton, D. K. (1999). Creativity as blind variation and selective retention: is the creative process Darwinian? *Psychological Inquiry*, *10*(4), 309–328.

Simonton, D. K. (2009). Varieties of (scientific) creativity: A hierarchical model of Domain-SpeciFic disposition, development, and achievement. *Perspectives on Psychological Science*, *4*(5), 441–452. doi:10.1111/j.1745-6924.2009.01152.x

Simonton, D. K. (2016). Defining Creativity: Don't we also need to define what is not creative? *The Journal of Creative Behavior*, *52*(1), n/a–n/a. doi:10.1002/jocb.137

Sio, U. N., Kotovsky, K., & Cagan, J. (2015). Fixation or inspiration? A meta-analytic review of the role of examples on design processes. *Design Studies*, *39*, 70–99. doi:10.1016/j.destud.2015.04.004

Sosa, R. (2019). Accretion theory of ideation: Evaluation regimes for ideation stages. *Design Science*, *5*, e23. doi:10.1017/dsj.2019.22

Suwa, M., & Tversky, B. (1997). What do architects and students perceive in their design sketches? A protocol analysis. *Design Studies*, *18*(4), 385–403. doi:10.1016/S0142-694X(97)00008-2

Sverdlov, E. D. (2018). Incremental Science: PapErs and grants, yes; discoveries, no. *Molecular Genetics, Microbiology and Virology*, *33*(4), 207–216. doi:10.3103/S0891416818040079

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468–472. doi:10.1126/science.1240474

Wang, J., Veugelers, R., & Stephan, P. (2016). *Bias agaiNst novelty in science: A cautionary tale for users of bibliometric indicators* (Working Paper No. 22180). National Bureau of Economic Research. http://www.nber.org/papers/w22180

Ward, T. B. (1994). Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology*, *27*(1), 1–40. doi:10.1006/cogp.1994.1010

Ward, T. B. (1998). Analogical distance and purpose in creative thought: Mental leaps versus mental hops. In K. J. Holyoak, D. Gentner, & B. Kokinov (Eds.), *Advances in analogy research: integration of theory and data from the cognitive, computational, and neural sciences* (pp. 221–230). Sofia, Bulgaria: New Bulgarian University.

Ward, T. B. (2001). Creative cognition, conceptual combination, and the creative writing of Stephen R. Donaldson. *The American Psychologist*, *56*(4), 350–354. doi:10.1037/0003-066X.56.4.350

Weisberg, R. W. (2015). On the usefulness of "value" in the definition of creativity. *Creativity Research Journal*, *27*(2), 111–124. doi:10.1080/10400419.2015.1030320

Wilkenfeld, M. J. (1995). *Conceptual combinations: Does similarity predict emergence?* (Master's thesis, Texas A&M University).

Wilkenfeld, M. J., & Ward, T. B. (2001). SimilaRity and emergence in conceptual combination. *Journal of Memory and Language*, *45*(1), 21–38. doi:10.1006/jmla.2000.2772

Wisniewski, E. J. (1997). Conceptual combination: Possibilities and esthetics. In T. B. Ward, S. M. Smith, & J. Viad (Eds.), *Creative thought: An investigation of conceptual structures and processes* (pp. 51–81). Washington, DC: American Psychological Association.