



ARTICLE

Transformative research focus considered harmful

Michael Cooper¹ | John Licato²

¹Philosophy, University of Tampa, Tampa, Florida, USA

²Computer Science, University of South Florida, Tampa, Florida, USA

Correspondence

Michael Cooper, Philosophy, University of Tampa, 401 W. Kennedy Blvd. Tampa, FL 33606.

Email: mjcooper@ut.edu

Abstract

Researchers are often encouraged to pursue nothing short of revolutionary advances, and those who work in artificial intelligence are no exception. However, an *exclusive* focus on revolutionary breakthroughs is often counterproductive in science. As explained by Kuhn almost 50 years ago, dramatic breakthroughs usually rely on a foundation of less dramatic advances, which uncover anomalies and make marginal improvements to current efforts. Progress relies on an *essential tension* between convergent and divergent thinking, each being complementary aspects of the same process. We argue that an overemphasis on, and exclusive rewarding of, divergent thinking in contemporary AI—whether in the form of rejecting funding for nontransformative research, or peer-review criteria rejecting papers for lack of novelty—is counterproductive to artificial intelligence and machine learning research, and may even be fundamentally harmful to progress in the field. To reckon with this problem, we recommend increasing funding for iterative improvement of theories, better guidance for reviewers, and more transparency in public funding.

INTRODUCTION

Doing AI research can feel like building on shifting sands. Any time it seems like the field is starting to establish foundations, there is always a persistent threat that yet another paper will come out with a catchy name and a fancy new approach that will revolutionize the way we think—at least, so claim their abstracts. Furthermore, the impetus to continually revolutionize rather than stabilize is actively encouraged, from advisors who tell their students to “only write the first or the last paper on a topic,” to funding agencies that allow “not transformative” to constitute a valid reason for proposal rejection.

Transformative research is crucial to the advancement of science. However, an exclusive and myopic focus on transformation may, ironically, prevent it from occurring. Transformative research tries to change the dominant

paradigm of a field, shaking up the way things are done and the things that are believed. The NSF, to cite a prominent example, has fully embraced transformativity, as all proposals submitted to the agency must be evaluated in part on how well they “suggest and explore creative, original, or *potentially transformative* concepts.” As a result, many AI researchers can point to at least one funding proposal that was rejected due to reviewers who cited insufficient transformative potential. There is danger in this strategy, however. Increasing research shows that the ability for reviewers to properly and accurately assess transformative potential is suspect at best, and may even be actively harmful to truly transformative ideas (Smith 2010; Gravem et al. 2017). Furthermore, revolutionary breakthroughs have often relied on a foundation of less exciting work—in fact, scientific process is made possible by what Thomas Kuhn called an *essential*

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *AI Magazine* published by Wiley Periodicals LLC on behalf of the Association for the Advancement of Artificial Intelligence.



tension between convergent and divergent research (Kuhn 1977).

We will now briefly re-examine Kuhn's arguments in the context of contemporary AI research. We will then summarize research on the limitations and harms of assessing transformative potential in peer review, and make recommendations for changes. In this article, we argue against the harms of *transformative potential as a mandatory review criterion*, particularly focusing on its use in selecting funding for AI research.

WHAT IS TRANSFORMATIVE RESEARCH?

When Thomas Kuhn first taught the history of science at Harvard for nonscience majors, he was working under the assumption that many of these elite students would end up in funding or governmental roles, which related to science. Under the advice of James Conant, then-President of Harvard, he tried to describe the general nature of science so that these students could respond well to any new scientific advance that arose. This set his students apart from those expected to become working research scientists, who needed, above all else, to understand the current theories of their disciplines well (Jacobs 2010; Conant 1947). It was in this context that he wrote his famous work, *The Structure of Scientific Revolutions*, in which he made the distinction between “normal” and “revolutionary” science, and coined the now-popular usage of the word “paradigm” (Kuhn 1962).¹ He used the term “normal science” to refer to the periods of time in which a scientific discipline advances slowly, without revolutionary breaks in theory. As he later explained, scientific revolutions depend on long periods of convergent thinking applied to normal science; “[R]evolutionary shifts of a scientific tradition are relatively rare, and extended periods of convergent research are the necessary preliminary to them” (Kuhn 1977, 227). Kuhn argues that in order to produce the best new research that nurtures broad changes in theory, we must embrace the “essential tension” between convergent and divergent modes of research.

Kuhn's arguments suggest that the best way to stimulate effective, novel research is to build a foundation that details and elaborates on current theories, noting the anomalies that appear along the way. He argues that scientific research is often better served by working to improve upon current theories rather than constantly stopping to re-examine the theoretical foundations of the work. Revolutionary changes in scientific theory usually depend on understanding the anomalies in the current theory. These anomalies are defined by the comparison between a detailed theory that makes precise

predictions and real-world conditions that diverge from those predictions.

One argument opposing Kuhn holds that some revolutionary advances in science come about through chance, rather than as a result of the processes of normal science, as Kuhn describes them. One possible case of this happening was in the revolutionary discovery of the double-helix structure of DNA by Watson, Crick, and Wilkins through a “chance” encounter with Rosalind Franklin's “photograph 51,” which was produced using x-ray crystallography. It is worth noting, though, that Franklin produced this photograph by making “evolutionary” advances to the field of x-ray crystallography. Further, the photograph is not interpretable without a thorough grounding in the theories of the physical structure of DNA and observed heritable traits. If someone without the requisite training in the relevant theories views photograph 51, they might very well mistake it for a picture of a vinyl record, with its label blurred by the motion. It was not the picture itself, but how the picture related to current theories that caused the revolutionary discovery of DNA. It was not the photograph alone, but the ways in which the photograph corresponded with or diverged from current theories that created the essential tension that marks this revolutionary advance (Bird 2000; Politi 2018).

Can (or should) all research be transformative?

In 1959, Kuhn was invited to speak at a conference on the subject of nurturing divergent approaches in scientific research. In his keynote, he surprised his audience by declaring the dangers of an exclusive focus on transformative research marked by divergent thinking, arguing that this focus may underestimate the importance of “incremental” research marked by convergent thinking. Kuhn argues that convergence on similar theories within a discipline is a precondition for calling a discipline a “science” at all. “[H]istory strongly suggests that, though one can practice science—as one does philosophy or art or political science—without a firm consensus, this more flexible practice will not produce the pattern of rapid consequential scientific advance to which recent centuries have accustomed us” (Kuhn 1977, 232). This is not to say that disciplines with conflicting theories cannot advance, or that their contributions are not worthwhile. Rather, Kuhn argues that a scientific discipline's convergence on a few accepted theories contributes to scientific development by allowing researchers to focus on solving small puzzles of the current theory, instead of having to review and re-argue their theory for each new experiment. He says that great advances in theory often come from engaging in research

that converges on and improves upon current theory. This, he argues, is a necessary part of creating the conditions needed for scientific revolutions to occur.

According to Kuhn, most research scientists, including great luminaries, spend the vast majority of their time *expanding* current theory, rather than exclusively attempting to overturn it. They do this by working out the theory's implications and trying to solve small puzzles whose nature and standards for acceptable solutions come from the established theory, which is broadly accepted within the discipline. He argues that by focusing their attention on small, solvable puzzles, these scientists not only incrementally advance science, but set the stage for greater breakthroughs in understanding to emerge (Kuhn 1962, 10). Though ordinary scientific research involves small divergences from accepted theory, there are also examples of revolutionary thinkers who made larger divergences, such as Copernicus, Darwin, and Einstein. These revolutionary thinkers promoted theories, which were broadly incompatible with the theories that came before, and all had their insights vindicated by future generations of researchers. The previous theories these thinkers worked with defined the problems that their revolutions were to address.

We argue that Kuhn's observations relating to how normal and transformative science occur and can flourish are still applicable today, including to research in artificial intelligence and its various subfields. But current mechanisms for selecting which AI research to publish or fund draw on a mistaken conception of transformativity in science. Consider, for example, that the National Science Foundation (NSF) requires reviewers for *every* submitted proposal to evaluate the extent to which submitted proposals "suggest and explore [...] *potentially transformative* concepts," and "program officers are instructing reviewers to pay special attention to those proposals that may include potentially transformative research."²

In defense of their focus on transformative research, it was suggested that a foundation-wide change would "provide a clear indication to the entire scientific and engineering community that NSF welcomes, encourages, and supports research ideas that push the frontiers or challenge current orthodoxies [and would] be only the first step toward achieving a broader and longer-term capacity for supporting revolutionary ideas within NSF and, more importantly, toward providing the freedom that encourages greater boldness of ideas and aspirations within the research community" (Task Force on Transformative Research 2007, 9). Insofar, as their approach seeks to avoid an environment that stifles exotic ideas, and to encourage the exploration of such ideas in order to more rapidly advance scientific progress, their approach is clearly well-intentioned. Does it, however, achieve

these goals? Although such emphases on transformativity have admirable goals, their current implementation may be counterproductive and should be re-assessed. Kuhn's ideas, we argue here, give us a way to do just that.

The NSF is far from alone in its overzealous focus on transformativity. To cite just a few examples from US federal sources of science funding: The National Institute of Health (NIH) is also looking to make big breakthroughs by pursuing "high-risk, high-reward advances in computer and information science, engineering and technology, behavior, cognition, robotics and imaging [...] Realizing the promise of disruptive transformation in health."³ A recent broad agency announcement from the Defense Advanced Research Projects Agency's (DARPA) Information Innovation Office, which funds AI research, demonstrates their explicit and exclusive focus on making breakthroughs: "Proposed research should investigate innovative approaches that enable revolutionary advances in software science, technology, or systems. Specifically excluded is research that primarily results in evolutionary improvements to the existing state of the art" (Defense Advanced Research Projects Agency, 6). The Air Force Office of Scientific Research's 2014 Strategic Plan proposes to "seek out and support the best, most transformational basic research with the greatest potential for impact on the future Air Force" (Air Force Office of Scientific Research 2014, 4), by focusing on "early recognition of unexpected advances in science and technology, emerging scientific breakthroughs, and disruptive technologies" (Air Force Office of Scientific Research 2014, 7). The Office of Naval Research's guide to peer review suggests asking whether a proposal for funding has "the higher risk and high payoff characteristics normally associated with basic research."⁴ The term "high-risk, high-reward" is often used as a synonym for transformative research (Committee on Science and Technology 2009, 7). A note from the director of the Army Research Office states that their "commitment toward active program management and understanding in the nonlinearity of science that allows us to identify high-risk, high-payoff investments in science and engineering research and education" (Army Research Office 2012, 4).

These governmental funding sources are particularly important, as they represent the dominant source of funding for research without any prospects for generating a short-term profit. As the American Association for the Advancement of Science has argued, incentives in private research may not spur enough research in areas that do not quickly lead to a return on investment. "[M]any worthy research projects are risky, with uncertain prospects for success, and may require long-term commitments of resources and infrastructure. These qualities of the research enterprise can lead to underinvestment by private industry, which in general is more focused on



lower-risk research and product development for shorter-term results. This is why industry spends about 80 cents of every R&D dollar on development, and only about 20 cents on basic and applied research. For federal non-defense agencies, the ratio is reversed” (Hourihan and Parkes 2016). Much of the work of uncovering anomalies in theories and trying to address those puzzles, the activity of normal science, often does not lead to immediate, profitable applications, so this presents a problem.

These examples highlight the pervasiveness of the desire for transformative research in the public funding of science, and the active discouragement of anything else. Our concern is that by making transformation a dominant criterion for receiving funding, we may inadvertently starve more evolutionary advances of funding. Absent these evolutionary advances, we may fail to create the conditions necessary for revolutionary advances to emerge. When a focus on transformation becomes expected practice, it becomes much more difficult to perform the extensive data gathering in Kuhnian normal science. Research carried out by current and aspiring faculty who hope to get funding in the future will trend towards what is fundable, and that may not serve the best interests of advancing the discipline.

Types of normal science

Precisely what kind of science is in danger of being marginalized when there is too much emphasis on transformative research? Kuhn points out three types of *normal* (as opposed to *revolutionary*) research projects that have often yielded results in the history of science. Kuhn argues that even the greatest research scientists spend the majority of their time pursuing such normal projects which, on their own, are unlikely to produce fundamental or revolutionary changes in theory. Instead, these projects collectively generate the conditions for transformative research to occur, advancing the state of knowledge by extending the current theory without seeking to overturn it. Kuhn offers examples of each of these common ways that normal research can provoke transformative change (Kuhn 1977, 233). We present them here to show how an exclusive focus on transformation can prevent this research from occurring:

(1) Explaining anomalies from previous experiments

Anomalies represent puzzles to be solved that extend beyond the standard problem sets within a scientific discipline. They provide opportunities to extend current theories by discovering new abstractions that pertain to the anomaly’s case. This puzzle-solving activity in turn makes theory conform better to experimental data, by repeating

experiments relating to established problems. Although this kind of research often changes theory, since it tends not to provide fundamentally new beliefs or ideas (in the sense that it extends, but does not overthrow existing theories), it is often dismissed as “incremental” by AI conferences and journals, a criticism that is often accepted as a credible reason to reject.⁵ Such an attitude is counter-productive, as Kuhn argued: transformative theories that provoke dramatic upheavals of theory depend on a careful and detailed exploration of anomalies and failures that arise within the standard theory. It is deep and concrete understanding of these failures, and the reasons for them, that creates the pressure that pushes researchers towards better theories.

Thomas Kuhn chiefly modeled his concept of scientific revolutions after Copernicus’ 1543 *De Revolutionibus*, naming his own first book *The Copernican Revolution* (Kuhn 1957). According to Kuhn, this revolution took about 250 years to be completed, which was not necessarily an indictment of contemporary astronomers’ resistance to change. Instead, he argued, there were good reasons to hesitate in accepting Copernicus’ model of the solar system and rejecting the dominant Ptolemaic model. Kuhn was particularly blunt in his assessment of the quantitative aspects of Copernicus’ revolution: “The seven-circle system presented in the First Book of the *De Revolutionibus*, and in many modern elementary accounts of the Copernican system, is a wonderfully economical system, but it does not work” (Kuhn 1957, 169). He argued that although Copernicus provided a more elegant top-down model for astronomy, it was not immediately a more accurate predictive model than the Ptolemaic model that came before, which had centuries of elaborations to deal with anomalous findings. This extreme lag time in accepting a successful scientific revolution can make it extremely difficult to determine when a new theory will end up being transformative. Indeed, scientific revolutions are rarely recognized as transformative at the time of proposal, sometimes even at the time of experimentation. Research supports this idea: among 72 highly cited ecologists, only four realized that their work was transformative during the proposal stage (Gravem et al. 2017).

Another interesting takeaway from the Copernican example is that revolutionary new models are not necessarily immediately better than previous models at explaining all anomalous data points. Instead, these new models may require persistent and stubborn support by dedicated proponents, even (and especially) when the models do not produce immediately quantifiable improvements. As with Copernican theory, it may take improving an inferior model over the course of many years to bring it in line with prominent anomalous findings to finally achieve a technological revolution. This may be

surprising to current AI researchers, who are used to accepting new models only when they demonstrate superior performance on established benchmark tasks, or funding only those proposals that present convincing preliminary data. One wonders how many transformational but unpopular ideas today are being rejected for support by funding agencies because they are unable to immediately explain anomalies. And are peer reviewers for such proposals instructed that truly transformative ideas may contain these sorts of anomalies? In our experience, this is not the case.

Consider, for example, how the state of neural network research in the so-called “AI Winter” that followed Minsky and Papert (1969) paralleled Copernicus’ work, in that it did not obviously outperform existing approaches, nor explain anomalies in the dominant approach in a thoroughly consistent way. Backpropagation, in the 1990s, re-emerged as an interesting but underperforming way to train multilayer neural networks, leading to a “second AI winter,” according to Hinton in his 2018 ACM Turing Award Lecture (in which he also notes that the 2009 NIPS conference rejected one of his deep learning papers because “they already accepted a paper on deep learning and two papers on the same topic would be excessive”).⁶ Deep learning, as the story goes, was nevertheless kept alive by researchers who persisted in studying them, despite reduced public interest and available funding.

To be fair, the current AI boom is too large for publication venues and funding agencies to adequately support all ideas, and the identification of unexplained anomalies may be a tie-breaking criterion adopted out of practical necessity. But this makes it unclear where the first type of “normal” research can flourish, or even to secure necessary funding to be carried out at all.

(2) Applying standard problem solutions to new problems

The second kind of advance that comes from convergent thinking is the strategy of applying a standard problem solution to a problem on which it has never been used. A classic example of this type from the history of science is Darwin’s application of well-established statistical methods of geology to problems in biology. By applying this interdisciplinary knowledge, Darwin generated a revolution using accepted methods of dating and classifying rock strata, particularly making use of terms popularized by Alexander von Humboldt (Darwin et al. 1844; Darwin 1859). It was only by utilizing established theories such as those concerning lava flows, sedimentation, and the development of coral was he able to extract useful information about the age of fossils, which was key to his biological insights (Bressan 2012). Though these methods were well-

known in geology, and thus considered normal science, their interdisciplinary application to biology was new and led to transformative advances.

In contemporary research, there appears to be no shortage of attempts to apply advances in AI to other fields. But although there is much potential for transformative research in AI to come from inter- and cross-disciplinary work, such work also faces higher barriers to entry for funding (Bromham, Dinnage, and Hua 2016). For example, funding proposals must convince reviewers that the idea is novel by summarizing the current state of the relevant field (else be rejected for not demonstrating awareness of prior work), and ensure that standard discipline-specific concerns have been considered (else be rejected for not having appropriate expertise). These are difficult enough for any author to do in the amount of pages allowed. But for cross-disciplinary proposals, the authors must somehow do the same things for multiple fields in one proposal—*typically within the same page limit*—and reviewers are allowed to use the same kinds of criticisms that would apply to single-field proposals. Much is written elsewhere about additional difficulties of assessing interdisciplinary research, and of overcoming the “silo effect” barriers that preserve and enforce disciplinary boundaries (Institute of Medicine 2005; Hein et al. 2018). Occasionally, opportunities for funding arise that purport to be explicitly cross-disciplinary. But even these will often eschew this second type of Kuhnian normal science—as an anecdotal example, a recent NSF solicitation, which “requires interdisciplinary teams, with expertise across disciplines” explicitly refuses to fund “[i]ncremental advances in existing technologies or deployment/implementation of existing technologies in novel learning contexts.”⁷

(3) Gathering concrete data to extend a theory

Extensive long-term data gathering is necessary for spotting unexpected anomalies, as described earlier. Kuhn argued that Tycho Brahe made this kind of advance when he recorded extensive astronomical observations with a higher degree of accuracy than ever before (Kuhn 1957). However, as Gravem et al. (2017) argues, some research requires “lengthy periods of data collection, which can lead to their categorization as ‘incremental.’” AI subfields exist that are all too familiar with this problem. For example, work in computational cognitive architectures, which seek to build cognitively plausible computational models of the *whole* human mind, must continually show how various cognitive phenomena can or can not be explained by their architectures (Sun 2004). This is not merely an issue of adding new convenient functionality or making a tweak to the architecture to solve a task, as may be the case with language models in deep learning. Rather, cognitive architectures are bound by *ontological commitments*:



Every change made to a cognitive architecture must be warranted by (and make predictions about) the architecture's underlying theory of the mind, else the entire project loses its ability to be explanatory of human cognition (Bricker 2016). Given the constantly expanding number of cognitive phenomena reported every year, cognitive architectures require decades of what might be seen as incremental work. Obtaining continued funding for such work seems impossible under current transformative-only criteria, and perhaps for this reason, it is virtually standard practice to give a new catchy name to every slight architectural variation (in the hopes of convincing reviewers that the proposed work really is fundamentally new), resulting in an unwieldy number of cognitive architectures (Kotseruba and Tsotsos 2018). Determining how this plethora of architecture variations relate to each other, much less jointly contribute to a common underlying theory of computational cognitive architectures, is a virtually impossible task. It is not unreasonable to suspect that if the taboo on incremental work was somehow lifted, researchers would be less reluctant to make connections to previous work more explicit.

WHAT CAN WE DO DIFFERENTLY?

The three kinds of “normal” science, which Kuhn uses as examples (Kuhn 1977, 233) depend on broad acceptance of the current theory in order to function. They also serve to highlight and explore the problems and inaccuracies of the current theory. However, this description is not antagonistic to the goal of developing transformative or revolutionary changes in theory. As Kuhn argues, great and novel discoveries in the science only stand out against a backdrop of detailed, established theories: “In the mature sciences the prelude to much discovery and to all novel theory is not ignorance, but the recognition that something has gone wrong with existing knowledge and beliefs” (Kuhn 1977, 235). By uncovering the areas of trouble where theoretical predictions fail to match experimental evidence, normal science concentrates attention on exactly those areas most useful in creating fundamental change to scientific theory. So, by advancing these kinds of normal science projects, a discipline creates a kind of tension that sets the stage for a possible transformation. This foundation of normal science can be advanced by carefully choosing strategies for funding and peer review to better facilitate the creation of this tension. In order to pursue that goal, we suggest three high-level recommendations to move AI research forward, which we intend to be applicable to all funding agencies and peer-reviewed journals and conferences.

Recommendation 1: Eliminate transformative potential as a mandatory and universal review criterion

We have spent much of this paper arguing that overly prioritizing transformative potential, to the detriment of supporting Kuhnian normal science, is harmful to the field overall, and paradoxically may slow transformative advances. But can transformative potential then be useful as a helpful tie-breaking criterion, to be invoked only when other criteria are inconclusive at deciding which proposal should be selected for funding? We argue that even in this limited role, it is counter-productive and will stifle support for normal science. The use of “potentially transformative” as a criterion to evaluate proposals, assessed by reviewers who are not properly trained, may invite problematic reasoning that can infect the rest of the review (and other reviewers). It is well-established that scientific peer review, when left completely unrestricted, can unfairly disadvantage underrepresented genders, ethnic minorities, and nonelite institutions (Ginther et al. 2011; 2018; Smith 2010). In general, even field-appropriate scientific experts are unreliable at picking out potentially transformative research (Gravem et al. 2017), and AI researchers are no exception: An analysis of ACL 2018 reviews found that author rebuttals were largely ineffective at changing the minds of reviewers (Gao et al. 2019). Reviewer ratings from AI conferences have been found to be poor predictors of a paper's ultimate impact (Cortes and Lawrence 2021; Tran et al. 2020), although papers, which ultimately go on to receive high citation counts seem to have a higher variance of reviewer ratings (Wang et al. 2021). Indeed, the typical member of the AI peer reviewer pool may be so underequipped at evaluating transformative research that forcing them to do so prior to proper training and vetting invites poor reasoning, and this may in turn skew the pool of funded projects in a way that is *actively harmful* to Kuhnian normal science (and thus, paradoxically, to transformative advances).

Thus, the mere presence of transformative potential as an evaluation criterion, *in funding programs, which are also meant to support normal science*, is harmful. But although evaluation on transformative potential should not be a universally utilized criterion, we recognize that there is a danger of moving the needle too far in the other direction. Kuhn's essential tension requires that transformative research be *explicitly supported alongside* normal science. Therefore, we must clarify: We do not believe funds available for transformative research should be reduced. A culture in which truly transformative ideas are shunned is quite the opposite of what we advocate. Instead, we believe that the current approach to encouraging transformativity needs refinement; namely, that potential for transformation should not be a criterion that is required

(or even allowed) for *all* funding programs indiscriminately. Instead, relegate it to programs specifically tailored to transformative work. In short, funding mechanisms designed for normal science should: (1) be separate from those designed exclusively to support transformative science; (2) be completely stripped of evaluation criteria designed for evaluating transformative work; and (3) constitute a much larger percentage of funding opportunities than those for transformative science.

Recommendation 2: Where potential transformative value must be assessed, (a) require better education and guidance of reviewers, and (b) require reviewers to justify their assessments using practices known to reduce bias

Funding mechanisms designed exclusively to support potentially transformative ideas must exist, though we advocate for their strict separation from those which are not. But there is an important caveat: programs designed to support transformative ideas must be implemented with a serious recognition of the flaws in the assessment process and aggressive combating of the biased reasoning to which all of us—yes, even highly educated academics—can so easily default. Reviewers for transformative research should be specifically trained⁸ (perhaps even qualified on their ability) to recognize inference patterns that are subject to bias, and produce strong justifications for all of their assessments of transformative potential, such that those justifications can be examined. Measures that are by now well-known to reduce bias in assessment should be aggressively employed: double-blind reviews, rebuttal periods for authors to clarify points to reviewers, requiring justifications of reviewer statements that can then be subject to scrutiny, and so forth.

At present, reviewers are asked to assess transformative value, and it is uncommon practice to question the reasoning behind their assessments. This is highly problematic as it invites assessments that, unbeknownst to the assessor, may have been influenced by unfair biases. Most researchers (in a survey to influential Ecologists (Gravem et al. 2017)) agreed that they cannot reliably assess or predict transformative research in proposals, but most “somewhat agreed” that such statements were still useful. There are, fortunately, some techniques that have shown promise in combating the potential influence of biased reasoning, and may decrease overconfidence effects. Requiring reviewers to state their justifications in causal, external statements has been shown to reduce the illusion of explanatory depth (Sewell et al. 2017). Furthermore, requiring reviewers to list unknowns prior to writing their reviews (e.g., by forcing reviewers to reason about what evidence would be needed to support their concluded assessments, and whether that evidence is present or missing) has also been shown to reduce unwarranted

overconfidence (Walters et al. 2016). But such practices do not appear to be common in the reviewing mechanisms of funding agencies.

Recommendation 3: Demand more transparency and data availability about everything related to public funding of academic research

Science in general suffers when findings cannot be replicated due to inaccessible data. Metascience, unsurprisingly, is no exception. The 2007 NSF report, which recommended a new transformative initiative stated that “[i]f NSF were to implement [this] initiative, it would need to be viewed as an experiment and thus assessed appropriately.” But now, almost 15 years later, the success of that experiment is unclear, at least to the wider scientific community.

But other publicly funded agencies, which decide on science funding are perhaps even more obscure about their practices. To our knowledge, no DoD-affiliated scientific funding office provides detailed knowledge about the individual projects funded, reviews used to select those projects for funding, or post-funding analyses of project success. High-level overviews are sometimes seen, but data are not available at a level of granularity sufficient for a third party to carry out meaningful analyses. Privacy and national security may be invoked as arguments for this status quo, and perhaps a more extended discussion on this topic is necessary. But the consequences of the lack of publicly available data as it relates to the focus of the present paper are clear: we have no concrete data about how effective evaluations of transformative research are, how effective various initiatives are at selecting transformative research, which attributes of peer reviewers or program officers enable them to better predict advances in science, and so on. It is reasonable to expect that some form of these studies are being carried out internally within funding agencies, but they are neither open to inspection, re-analysis, scrutiny, or debate by the broader scientific community—a practice that is inherently antiscientific.

Of course, we would be remiss not to acknowledge the enormous practical difficulty of completely releasing all data related to reviews, rejected proposals, proposal text, and the like. But public funding agencies by now must have detailed data relating to projects whose funding ended years ago—such data might be effectively provided as part of a grand challenge. Datasets of this type would likely serve as a boon to the science of science, and may lead to the very sort of transformative change that funding agencies are seeking.

In summation

It is a necessary precondition of revolutionary discoveries that researchers have detailed knowledge of the existing



theories and the experimental results, which justify them. It is only through such knowledge that problems emerge and develop clear characteristics. Absent this detailed knowledge of problems, it is difficult to come up with adequate solutions. As Kuhn argues, “The scientist requires a thoroughgoing commitment to the tradition with which, if he is fully successful, he will break” (Kuhn 1977, 235). It is a paradox in scientific research that in order to create novel advances in theory, it is necessary to embrace and explore theories to which the experts in the field are committed. Scientific progress is driven by an essential tension created between the drive for novelty and a commitment to disciplinary conventions. If funders and publishers of AI research are serious about advancing transformative change, *both sides* of this tension must be actively maintained.

CONFLICT OF INTEREST

The authors declare that there is no conflict.

ORCID

Michael Cooper  <https://orcid.org/0000-0002-4714-3186>

ENDNOTES

¹In much of his work, Kuhn famously refers to these broad theories as scientific “paradigms,” and this technical term was adopted into the popular lexicon. However, later criticisms, notably by Masterman (Masterman 1970), have plausibly argued that the term is unacceptably vague. As a result, we avoid using it in this article. Instead, we use the generic “theory” and attempt to describe the meaning we ascribe to it in each usage.

²www.nsf.gov/about/transformative_research/merit_review_criteria.jsp.

³grants.nih.gov/grants/guide/notice-files/NOT-OD-18-149.html.

⁴www.onr.navy.mil/-/media/Files/Research/ONR-Instruction-39661A.ashx?la=en, p. 9.

⁵In the reviewer instructions for AAAI 2021, for example, one criterion awards the highest rating to ideas that “are ground-breaking,” and the lowest possible score if “[t]he main ideas of the paper are not novel or represent incremental advances.”

⁶Available at youtu.be/VsnQf7exv5I?t=1380.

⁷www.nsf.gov/pubs/2020/nsf20612/nsf20612.htm.

⁸The idea of an expert panel specifically able to assess transformative research was suggested in the 2012 NSF panel (Frodeman and Holbrook 2012), but it is not clear whether it was followed up on.

REFERENCES

- Air Force Office of Scientific Research. 2014. *AFOSR 2014 Technical Strategic Plan*, 1–26.
- Army Research Office. 2012. *ARO Year in Review 2020*.
- Bird, A. 2000. *Thomas Kuhn*. Chesham: Acumen.
- Bressan, D. 2012. “Darwin the Geologist.” *Scientific American*. February.

- Bricker, P. 2016. “Ontological Commitment.” In *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.), edited by E. N. Zalta. Stanford, CA: Metaphysics Research Lab, Stanford University.
- Bromham, L., R. Dinnage, and X. Hua. 2016. “Interdisciplinary Research has Consistently Lower Funding Success.” *Nature* 534(7609): 684–87.
- Committee on Science and Technology. 2009. *House Subcommittee on Space, Science, and Technology*. October.
- Conant, J. B. 1947. *On Understanding Science: An Historical Approach*. Terry Lectures. New Haven: Yale University Press.
- Cortes, C., and N. D. Lawrence. 2021. Inconsistency in Conference Peer Review: Revisiting the 2014 Neurips Experiment.
- Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favored Races in the Struggle for Life*. London: Murray.
- Darwin, C., E. Forbes, W. Lonsdale, and G. B. Sowerby. 1844. *Geological Observations on the Volcanic Islands and Parts of South America Visited During the Voyage of H.M.S. ‘Beagle’*. New York: D. Appleton and Company. <https://www.biodiversitylibrary.org/bibliography/61452>
- Defense Advanced Research Projects Agency. Broad Agency Announcement HR001121S0010.
- Frodeman, R., and J. B. Holbrook. 2012. “Introduction.” In *Transformative Research: Ethical and Societal Implications*. Arlington, VA: The Promise and Perils of Transformative Research.
- Gao, Y., S. Eger, I. Kuznetsov, I. Gurevych, and Y. Miyao. 2019. “Does My Rebuttal Matter? Insights from a Major NLP Conference.” In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Long and Short Papers)*, Volume 1, 1274–90. Minneapolis, Minnesota: Association for Computational Linguistics.
- Ginther, D. K., J. Basner, U. Jensen, J. Schnell, R. Kington, and W. T. Schaffer. 2018. “Publications as Predictors of Racial and Ethnic Differences in NIH Research Awards.” *PLoS One* 13(11): e0205929.
- Ginther, D. K., W. T. Schaffer, J. Schnell, B. Masimore, F. Liu, L. L. Haak, and R. Kington. 2011. “Race, ethnicity, and Nih Research Awards.” *Science* 333(6045): 1015–9.
- Gravem, S. A., S. M. Bachhuber, H. K. Fulton-Bennett, Z. H. Randell, A. J. Rickborn, J. M. Sullivan, and B. A. Menge. 2017. “Transformative Research is not Easily Predicted.” *Trends in Ecology & Evolution* 32(11): 825–34.
- Hein, C. J., J. E. Ten Hoeve, S. Gopalakrishnan, B. Livneh, H. D. Adams, E. K. Marino, and C. Susan Weiler. 2018. “Overcoming Early Career Barriers to Interdisciplinary Climate Change Research.” *WIREs Climate Change* 9(5): e530.
- Houriham, M., and D. Parkes. 2016. *Federal R&D Budget Trends: A Summary*. Washington, DC: American Association for the Advancement of Science.
- Institute of Medicine. 2005. “Facilitating Interdisciplinary Research.” Technical report, The National Academies Press.
- Jacobs, S. 2010. “Fall. J. B. Conant’s Other Assistant: Science as Depicted by Leonard K. Nash, Including Reference to Thomas Kuhn.” *Perspectives on Science* 18(3): 328–51.
- Kotseruba, I. and J. K. Tsotsos. 2018. “40 Years of Cognitive Architectures: Core Cognitive Abilities and Practical Applications.” *Artificial Intelligence Review* 53: 17–94.



- Kuhn, T. S. 1957. *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought*. Cambridge, MA: Harvard University Press.
- Kuhn, T. S. 1962. *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.
- Kuhn, T. S. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago, IL: University of Chicago Press.
- Masterman, M. 1970. "The Nature of a Paradigm." In *Criticism and the Growth of Knowledge: Volume 4: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965*, Volume 4, 59. Cambridge University Press.
- Minsky, M., and S. Papert. 1969. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: MIT Press.
- Politi, V. 2018. "Scientific Revolutions, Specialization and the Discovery of the Structure of DNA: Toward a new Picture of the Development of the Sciences." *Synthese* 195(5): 2267–93.
- Sewell, D. K., P. J. Rayner, D. B. Shank, S. Guy, S. D. Lilburn, S. Saber, and Y. Kashima. 2017. "Causal Knowledge Promotes Behavioral Self-Regulation: An Example Using Climate Change Dynamics." *PLOS ONE* 12(9): 1–19.
- Smith, R. 2010. "Classical Peer Review: An Empty Gun." *Breast Cancer Research* 12: S13.
- Sun, R. 2004, September. "Desiderata for Cognitive Architectures." *Philosophical Psychology* 17(3): 341–73.
- Task Force on Transformative Research. 2007. "Enhancing Support of Transformative Research at the National Science Foundation." Technical report, National Science Board.
- Tran, D., A. Valtchanov, K. Ganapathy, R. Feng, E. Slud, M. Goldblum, and T. Goldstein. 2020. "An Open Review of Openreview: A Critical Analysis of the Machine Learning Conference Review Process." *CoRR abs/2010.05137*.
- Walters, D. J., P. M. Fernbach, C. R. Fox, and S. A. Sloman. 2016. "Known Unknowns: A Critical Determinant of Confidence and Calibration." *Management Science* 63(12): 4298–307.
- Wang, G., Q. Peng, Y. Zhang, and M. Zhang. 2021. "What have We Learned from Openreview?" *CoRR abs/2103.05885*.

AUTHOR BIOGRAPHIES

Michael Cooper recently completed his Ph.D. in philosophy at the University of South Florida in the Summer of '22. His work focuses on the philosophy of science, with special emphasis on interdisciplinary studies. He has been working with the Advancing Machine and Human Reasoning (AMHR) lab for 2 years on analogical arguments, among other topics. He has previously earned an M.A. in philosophy and an M.S. in ethics and public policy.

John Licato is an Assistant Professor of Computer Science and Engineering at the University of South Florida, and Director of the Advancing Machine and Human Reasoning (AMHR) Lab. His expertise is in natural language processing (NLP), and the application of advances in artificial intelligence to modeling and understanding cognitive-level reasoning and argumentation. His recent work centers on a variety of subareas in which AI can potentially help people reason better, including: providing fair justifications of judgments, interpretations of open-textured language, decision-making in domains that are highly prone to biased reasoning, and others.

How to cite this article: Cooper, M., and J. Licato. 2022. "Transformative research focus considered harmful." *AI Magazine* 43: 273–81. <https://doi.org/10.1002/aaai.12063>