|  |  |
| --- | --- |
| **ALFRED P. SLOAN FOUNDATION**  [sloan.org](https://sloan.org/) | [proposal guidelines](https://sloan.org/grants/apply#tab-grant-proposal-guidelines) | PROPOSAL COVER SHEET |

**Does Government Funding Change What You Do? The Effects of Funding on the Direction and Impact of Academic Energy Research**

|  |  |  |
| --- | --- | --- |
| **Project Information** | | |
| **Principal Investigator** | **Grantee Organization:** | Syracuse University |
| David Popp, Caroline Rapking Faculty Scholar in Public Administration and Policy, Maxwell School, Syracuse University  (315) 443-2482  [dcpopp@maxwell.syr.edu](mailto:dcpopp@maxwell.syr.edu)  Daniel Acuna, Assistant Professor, Science of Science Lab, School of Information Studies, Syracuse University  (315) 443-4473  [deacuna@syr.edu](mailto:deacuna@syr.edu) | **Amount Requested:** | $349,380 |
| **Requested Start Date:** | June 1, 2020 |
| **Requested End Date:** | May 31, 2023 |
|  | **Project URL (if any):** | n/a |
| **Project Goal** | | |

Our interdisciplinary team of researchers will examine how government funding shapes the direction of clean energy research, both by asking whether targeted funding programs encourage new entrants into the field and by evaluating the relative research success of new entrants. Do increases in government spending attract more researchers to the field of clean energy or simply substitute for other sources of funding?

|  |
| --- |
| **Objectives** |

Our project uses both machine learning and regression analysis to answer three research questions: (1) Do scientists change the focus of their research in response to targeted government funding opportunities? (2) If so, what types of calls for funding best attract new researchers? (3) Do researchers new to a field contribute novel ideas? Do they produce more highly cited research?

|  |
| --- |
| **Proposed Activities** |

Our project combines text analysis of both publications and requests for proposals with regression analysis to answer the three questions above. We use machine learning analysis of text to track changes in research portfolios and measure novelty. We also apply machine learning to construct a control group of researchers with similar research interests and demographic characteristics to our funded researchers. We apply regression analysis to the resulting data set to test, for example, if movements towards clean energy research are truly a result of receiving funding, rather than a reaction to general trends in the broader research community.

|  |
| --- |
| **Expected Products** |

We expect this project to generate at least three manuscripts targeted at academic journals: (1) A paper on funding affects the direction of research and, if so, do targeted funding programs move researchers further than open funding programs. (2) A paper asking whether researchers new to a topic generate more novel and more impactful research. (3) A paper comparing the results of machine learning to other commonly used matching techniques such as propensity score matching, so as to inform future research using matching techniques. While the publication databases used in our analysis are proprietary, we will make the code generated by our research publicly available, so that other users with access to these databases can replicate and extend our work.

|  |
| --- |
| **Expected Outcomes** |

Our research will contribute to both energy policy and to the broader scientific policy community. It evaluates whether ARPA-E successfully achieved their goal of bringing new researchers into the field. It will help guide the decisions of all funding agencies by identifying whether targeting funding at new researchers provides greater impact, and if so, identify strategies to best target these researchers. By combining machine learning and regression analysis, we create new methods for causally identifying the effect of R&D policy.

**Does Government Funding Change What You Do? The Effects of Funding on the Direction and Impact of Academic Energy Research**

**Principal Investigators: David Popp and Daniel Acuña**

**1. What is the research question and why is it important?**

Government-funded energy R&D remains an important part of policy initiatives to address climate change. Meeting currently proposed climate goals requires replacing much of the world’s current fossil fuels-based energy capacity with alternative, carbon-free energy sources. While clean energy technologies improved dramatically in recent years, meeting these goals requires continued innovation. Recognizing this need, during the December 2015 Paris climate meetings, a coalition of governments pledged to double their renewable energy R&D budgets to over $32 billion over the next five years as part of “Mission Innovation.” To be truly impactful, the resulting increases in government spending should attract more researchers to the field of clean energy and not simply substitute for other sources of funding. Recognizing this, the Department of Energy’s Advanced Research Projects Agency-Energy (ARPA-E) program explicitly lists creating new communities of researchers, including those “not traditionally... involved in the topic area” (National Academies of Science, 2017, p. 122) as a goal.

Most academic literature evaluating government research asks whether funding increases research output, but does not study the mechanisms by which research increases. Does it come from moving existing researchers into the field? Does it encourage junior researchers to enter the field? Notably few papers in the scientific policy literature address the direction of research, and fewer still attempt to make causal links between funding and research direction. Exceptions include Azoulay *et al*. (2011) and Myers (2019), both of which focus on health sciences. Our interdisciplinary team of researchers will examine how government funding shapes the direction of clean energy research, both by asking whether targeted funding programs encourage new entrants into the field and by evaluating the relative research success of new entrants.

Our proposed study examines clean energy research funding from the US Department of Energy (DOE) and National Science Foundation (NSF). In particular, by including the ARPA-E program, which began in 2009, we will evaluate whether a new infusion of R&D funds available for clean energy R&D encouraged researchers receiving funding to do new research on clean energy. We will combine data on both calls for proposals and funding from multiple sources with publication data taken from both Web of Science and Scopus. Using these data, our proposed research addresses the following research questions:

1. *Do scientists change the focus of their research in response to targeted government funding opportunities?*  Many Department of Energy programs are mission driven. Funding Opportunity Announcements (FOAs) include specific goals for funded research. Researchers can choose the approach they will take, but have less choice over the questions to be addressed. These programs will have greater impact if they influence the direction of research, and not just support research likely to have been carried out anyway. We use text analysis of both publications and requests for proposals to test whether funded scientists change the direction of their research more than a control group with similar research portfolios prior to the funding decision. Do these programs fund research unlikely to be otherwise carried out? Do they bring in new researchers to the field, as stated in ARPA-E’s goals, or do funds primarily go to researchers already active on similar topics?
2. *What types of calls for funding best attract new researchers?* Given ARPA-E’s goal to attract new scientists to energy research, we ask how funding agencies can best attract new researchers. By comparing the research of both funded scientists and their control group to the proposed research topics, we will assess what types of research solicitations are most likely to influence the direction of energy research. We will compare the effect of the mission-driven targeted programs (e.g., focused ARPA-E programs targeting a specific technology) to programs featuring open-ended calls for proposals, such as DOE’s Office of Science and ARPA-E’s tri-annual OPEN FOAs, as well as related National Science Foundation programs. Open-ended programs allow for a broad range of proposed ideas, typically bounded by a specific discipline (e.g. NSF’s Cellular and Biochemical Engineering) or broad research area (e.g. NSF’s Science and Innovation Policy), seemingly lowering the incentive to switch research topics. Therefore, are targeted funding opportunities more likely to encourage researchers to change topic, or do they only attract researchers already established in the field? Are open-ended programs in an agency with a single focus (e.g. DOE) more or less effective than similar programs at NSF, which funds scientists in many fields?
3. *Do researchers new to a field contribute novel ideas? Do they produce more highly cited research?* Should funding agencies even desire to attract new researchers to a field? Might a researcher working in materials science contribute new ideas to improve the efficiency of solar panels? Are research teams combining both experienced members and researchers new to the field more effective? Using text analysis on a broad set of renewable energy publications, we compare the similarity of these publications to each author’s previous work. We ask whether papers less similar to an author’s previous work are (1) more novel and (2) more highly cited than other papers in the same field.

This work has impacts for both energy policy and the broader science policy community. The technological challenges of further reducing greenhouse gas emissions will be much greater than the challenges overcome so far (ITIF 2018). Breakthrough innovations are imperative to reduce carbon emissions to near zero in the long term. Such innovations will have large spillover benefits and will require public R&D support. To be effective, new investments in government R&D must both expand the pool of scientists in the field and move energy research in new directions needed to achieve nearly complete decarbonization. Our research evaluates whether existing energy R&D programs achieve these goals, paying particular attention to whether differences in the design of funding programs affects these outcomes.

For science policy, our work uses new data and develops new methods to evaluate causal links between funding and future research directions. Nearly all existing work on research pattern analysis uses the PubMed related articles (*pmra*) algorithm. *Pmra* uses a probabilistic model of curated keywords (MeSH terms), which does not account for keyword correlation and may limit their ability to adapt to rapidly changing fields. Instead, our proposed method uses topic modeling, which produces compact yet representative research descriptors that tend to agree more with human-based similarity measures (Achakulvisut *et al*. 2016; Blei *et al.* 2003). Rather than using a predefined or hand-curated set of keywords, our method relies on machine learning analysis of text to track changes in research portfolios and measure novelty, in effect discovering the most informative keywords automatically. We also apply machine learning to construct a control group of researchers with similar research interests and demographic characteristics to our funded researchers, but that do not receive funding from the targeted programs in our study. These controls allow us to ascertain, for example, if movements towards clean energy research are truly a result of receiving funding, rather than a reaction to general trends in the broader research community. Finally, our results will have broad implications for almost all funding agencies. By asking whether scientists new to a field contribute more novel and impactful research, we consider whether funding agencies should aim to attract new researchers to a field or support those with established records on the funding agencies desired topics. If they do wish to attract new researchers, our cross-agency comparison allows us to study the impact of different types of programs on scientists’ research choice.

**2. What is the state of the research on this question?**

Our research contributes to three different strands of academic research: (1) energy R&D policy research, (2) research on the direction of scientific activity, and (3) studies on the impact of novel research. Regarding energy R&D policy, few papers provide causal evidence of public clean energy R&D effectiveness. Most studies simply include public R&D expenditures as one of several variables in more general studies of the drivers of clean energy patenting (e.g. Johnstone *et al.*, 2010; Verdolini and Gaelotti, 2011; Peters *et al*., 2012; Dechezleprêtre and Glachant, 2014; and Nesta *et al.*, 2014). To more directly focus on the effectiveness of public energy research, Popp (2016) links data on scientific publications to public energy R&D funding, finding that $1 million in additional government R&D funding leads to 1-2 additional publications, but with lags as long as ten years between initial funding and publication. Specific to ARPA-E, a few recent studies provide initial evaluations (e.g. Goldstein and Narayanamurti, 2018), including a 2017 National Academies of Science report (National Academies of Science, 2017). However, given the long-term goals of ARPA-E, these early studies acknowledge that they can only provide very preliminary analysis that is not causal in nature. Now having ten years of data for ARPA-E enable us to provide a more thorough causal analysis.

Similarly, few papers study novelty in energy research. Most closely related is Nemet (2012), who uses patent citations to study inter-technology knowledge flows, finding that energy patents citing patents from multiple technological domains receive more citations themselves. Popp (2017) uses citations between journal articles and patents to assess the value of knowledge from different institutions, including universities, private sector, and government laboratories. Research performed at government institutions appears to play an important translational role linking basic and applied research. Moreover, both scientific articles and patents with authors from multiple types of institutions (e.g., university and corporations) are cited more frequently, suggesting that collaborations may have positive impacts on research quality. However, the collaborations in Popp (2017) include researchers with different professional backgrounds, not researchers coming from different scientific communities.

Our first two research questions contribute to the literature on the direction of scientific activity. A large literature addresses why scientists choose the topics they do (Stephan 2010), including scientific priority (e.g. Merton 1957), reputation, and financial rewards (e.g. Stephan 1996). Here, we focus on the role of government funding. Governments may use government funding to encourage research on specific topics considered important to society (e.g. mission research). While several recent studies ask whether funding increases the level of research activity (e.g. Jacob and Lefgren, 2011 and Rosenbloom *et al*. 2015), this is not the appropriate question if government funding aims to redirect research. Notably, while the above studies find mixed results, studies focusing on narrow fields of research typically find a positive relationship between funding and research outputs (e.g. Blume-Kohout, 2012 on NIH and Popp, 2016 on energy R&D). Thus, even if researchers are able to substitute funding from multiple source to maintain a given level of research activity, government funding may influence research by affecting the topics chosen.

Fewer papers in the scientific policy literature address the direction of research. Myers (2016) reviews related work in the field of biomedicine. Much of the work here looks at how changes in stem cell funding by the U.S. government affected researchers. The closest published work to our proposed research is Azoulay *et al* (2011), who document how the nature of different research grants affect the riskiness of topics chosen. However, they do not consider how funding affects the choice of research field itself. Corredoira *et al*. (2018) show that federally funded patents have greater influence (e.g. are part of more active and diverse citation trees) and appear more frequently in different technological classes than corporate patents, providing suggestive evidence of federal funds affecting the direction of research. However, they cannot identify whether private inventors *would* have worked in these same areas had there been less federal support. In addition, two recent working papers address topic choice. Myers (2019) compares applications to open-ended and targeted NIH competitions. Researchers are indifferent between a one standard deviation redirection to less similar science and about $1 million additional grant funds. In an event that significantly lowered the cost to access technology, in late 2010 hackers demonstrated how to use Microsoft Kinect video game sensors for other applications, making it an affordable tool for motion sensing research. Furman and Teodoridis (2018) use text-based analysis to show that this lowering of cost of motion sensing research both increases the quantity of research on motion-sensing and encourages new entrants into the field. Thus, while there is promising but scarce evidence about how incentives change what researchers do, more systematic studies are needed.

Our research contributes to this nascent literature in several ways. First, by comparing the effect of funding from multiple government agencies and different funding programs, we can learn how different program funding goals affect research direction. Second, while most existing research on the topics of academic research uses the PubMed related articles (*pmra*) algorithm, our work develops new methods to extend such analysis into new fields of study. For instance, while Furman and Teodoridis (2018) use machine learning to construct measures of diversity of research profiles, they do not apply machine learning to the construction of their matched sample. Our machine learning methods will find such matched samples to construct a control group of researchers. Our proposed methods could be applied to other areas of science policy as well, such as using patent records to study the effectiveness of policies designed to support small businesses. Third, ours will be the first study asking whether funding moves additional researchers into clean energy, an area of growing policy interest around the world.

Our third research question contributes to the literature asking what types of research generate high-impact results. While no existing work explicitly uses the similarities between a scientist’s previous research to their current work when assessing the impact of a publication, a few recent papers use journal references to assess novelty. Papers combining references in new or unusual ways are considered more novel. Novel works are cited more frequently (Uzzi *et al*., 2013), but may take longer to be cited (Wang *et al*., 2017). Research teams with expertise from more diverse fields and performing multiple research tasks produce more novel research (Lee *et al*., 2015). Wu *et al*. (2019) add a cautionary tale, however, as they find larger research teams are more likely to build on existing work, and that smaller teams are more disruptive. In contrast to these papers, our measures of novelty focus on characteristics of individual researchers, rather than the citations in published papers. For funding agencies, the focus on researchers, rather than research output, is important, as similarity to a researcher’s existing portfolio can be observed *prior to the research being completed*. If researchers new to a field do contribute more valuable ideas, funding agencies can consider differences between an investigator’s proposed research and her prior research record when evaluating funding proposals.

**3. Why is the proposer qualified to address the research question for which funds are being sought?**

Our interdisciplinary team is well equipped to successfully finish this study. While both scholars have training in traditional disciplines, they work in interdisciplinary departments and publish articles in a range of academic journals. David Popp is an economist in the Department of Public Administration and International Affairs at Syracuse University’s Maxwell School. He has published extensively in both economics and policy journals on the intersection of energy and science policy, including papers examining the relationship between both energy prices and policies on innovation and on the role of R&D subsidies for climate policy. His 2002 publication in the *American Economic Review,* “Induced Innovation and Energy Prices,” was selected for the 2017 Association of Environmental and Resource Economists Publication of Enduring Quality Award. In addition to original scholarly research on energy innovation, he has published several reviews of literature on energy innovation (e.g. Popp *et al*. 2010, Popp 2010, Popp 2015, Popp 2019) and presents frequently to practitioner audiences on the topic of energy innovation.

Acuña is a computer scientist with expertise in big data analytics and the study of productivity and impact in science. He leads the Science of Science and Computational Discovery Laboratory and is an affiliate member of the Center for Computational and Data Science. His most prominent work has appeared in *Nature* (Acuna *et al*. 2012), *Nature Communications* (Liénard *et al*. 2018), and *Research Policy* (Teplitskiy *et al*., 2018), and his results have been featured in *Nature Podcast*, *The Chronicle of Higher Education*, *NPR*, and *the Scientist*. He is part of the inaugural editorial board of the *Journal of Social Computing*. Acuña also builds systems to automatically detect scientific falsification (Acuna *et al*., 2018) and to automatically recommend grants to junior faculty (https://eileen.io). He has multiple concurrent grants from NSF, the Department of Human and Health Services, and DARPA.

**4. What is the research methodology?**

***Programs to Study***

We will compare programs relevant for clean energy at two agencies: the U.S. Department of Energy and the National Science Foundation. We consider two dimensions of funding agency strategy: *focus* and *risk-taking*. Focus refers to whether a program provides targeted or open-ended funding opportunities. *Risk-taking* indicates whether a program was set up specifically to encourage interdisciplinary and/or high-risk/high-reward research—as opposed to other research programs, which we label as “traditional”. While all funding programs hope to fund novel, transformative research, we hypothesize that both interdisciplinary programs and those with a focus on untested, high-risk ideas will encourage research less similar to a scientist’s existing portfolio. We also ask whether targeted high-risk programs move researchers more than high-risk open solicitations, although we offer no *a priori* hypothesis as to which effect is larger. Figure 1 summarizes our categorization of the programs described below. Methods Appendix 1 provides more details on these programs.

|  |  |  |
| --- | --- | --- |
|  | **Open** | **Targeted** |
| **traditional** | NSF, standard grants  DOE Office of Science | DOE EERE |
| **high-risk/ high-reward** | DOE ARPA-E Open  NSF Cross-cutting  NSF EAGER | DOE ARPA-E focused |

**Figure 1. Categorization of research programs included in our study.**

We consider three programs at the DOE that span a range of funding options. ARPA-E aims to identify and fund high-risk, potentially high return energy research. It solicits research proposals through both *Focused* Funding Opportunity Announcements (FOA) targeting specific technology areas and tri-annual *Open* FOAs to ensure they do not miss potential innovative opportunities in other areas. Because ARPA-E was designed to bridge the gap between basic and applied research at DOE (Goldstein and Narayanamurti 2018), we include additional DOE programs on either side of this spectrum. The DOE’s Office of Science Basic Energy Sciences program provides an example of a traditional program supporting basic research using open solicitations. DOE’s Office of Energy Efficiency and Renewable Energy (EERE), which supports applied clean energy research, provides an example of a traditional targeted program.

NSF standard grants includes multiple directorates overseeing programs supporting research in a broad range of science and engineering fields through open calls for proposals. We will study programs most commonly used as funding sources by researchers also funded by DOE, so that we are comparing NSF programs that our DOE funded researchers are likely to consider. As our examples at the end of Methods Appendix 1 show, these NSF programs are not necessarily focused on energy, suggesting our results will inform how researchers traditionally outside the domain of energy research may be affected by changing funding priorities. While supporting transformative research is a goal of all NSF programs, specific NSF initiatives pay particular attention to novel and interdisciplinary research, as listed in Figure 1. We hypothesize that these programs, which emphasize interdisciplinary and/or high-risk, untested ideas, should promote greater changes in research direction than other open-ended programs.

***Data***

We will construct two similar but unique data sets to answer our questions. Asking whether scientists change the focus of their research in response to targeted government funding opportunities and, if so, what types of funding calls best attract new researchers require us to (i) identify researchers funded by relevant DOE and NSF programs and (ii) compare the research profiles of these researchers to a control group of similar researchers. Thus, the unit of observation in our first data set is an individual scientist.

To know whether attracting novel researchers *should* be a goal of funders, we then ask whether the publications of scientists who are new to the clean energy field contribute novel, high impact ideas. Measuring novelty to the field requires a set of articles reasonably representing the state of the field at the time of publication. Therefore, for this research question we use a larger sample of publications in biofuels, energy storage, solar energy, and wind energy from 1991-2017, based on previous work in Popp (2016, 2017). This larger sample both provides more comprehensive coverage of the state of research in these fields and sufficient time for the articles under consideration to be cited. Thus, while we use author data to compare the topics of each article to its authors’ prior research portfolio, for this research question our unit of observation is an article, rather than a scientist.

For both data sets, we will obtain publication data from two databases: Scopus and the Web of Science. Unlike Web of Science, Scopus includes full author names and adds a unique identifier for each author. Azoulay *et al*. (2017) conclude that their “systematic comparisons led (them) to believe that the Scopus system provides an accurate set of career publications.” Thus, Scopus will be our initial source of publication records, allowing us to obtain complete publication records back to 1970 for each researcher. However, while Scopus provides very limited information on funding acknowledgements, Web of Science began consistently collecting and providing funding acknowledgements in 2008. We will link the Web of Science and Scopus data when identifying publications from research programs included in our study.

Asking whether funding programs encourage researchers to change direction also require us to identify (1) scientists funded by the programs described above and (2) the proposal solicitations of these programs. We will compile funding data from multiple publicly available sources, including the Federal Reporter (for NSF funding) and Department of Energy websites. In Methods Appendix 2, we describe how these publicly available data will be assembled to allow mapping to the proprietary Scopus and Web of Science data for researchers with access to these data. Finally, while we can obtain data on awards made up to the present day, we will not be able to include the most recent awards, to allow time for the awards to generate publications. Preliminary analysis suggests a stopping point around 2014, allowing for five years of subsequent records by the end of 2019.

We use publications to track scientists’ research activity because academic scientists have large records of publications that will serve as inputs to constructing measures of similarity. We acknowledge that publications are not the only output of academic researchers in applied fields such as clean energy. However, we can compare the similarity of the proposed research to a scientist’s previous publications even if the grant in question does not produce any publications itself. Moreover, Goldstein and Narayanamurti (2018) show that (i) publications are an important output of both ARPA-E and Office of Science research and that (ii) the EERE projects in their sample are more likely to have a publication than a patent.

***Measuring Research Similarity***

For each research question, we require measures of similarity to (i) measure the similarity between funded research and those scientists’ previous work (Q1 & 2), (ii) identify the control group of similar researchers (Q1 & Q2), and (iii) measure the similarity between a publication and an author’s previous work (Q3). Thus, we begin by introducing the measure of similarity that we use in our research. We then discuss separately the methods used to address the three research questions above.

We can analyze whether funding affects research patterns by detecting changes between sets of documents published before, during, and after a grant's funding period. We can use the same methods to compare any one publication to an author’s previous research portfolio, allowing us to identify authors moving to new fields. In both cases, these changes can be detected using similarity measures. For example, a scientist’s research moves towards the goals of funding agencies if her research becomes more similar to the grant's Funding Opportunity Announcement after funding is received. The procedure we will describe is agnostic to the sets considered, and therefore enables both comparisons within a scientist or even across a field. Change detection using similarities is therefore appropriate for our goals.

There are multiple ways to compute similarities. To assess whether scientists change their research focus in response to funding, we will use the text of the grant's FOA (or equivalent) and publications as the primary data to estimate similarities. Some researchers have used other signals to estimate similarities such as the citations between publications (e.g., Wesley-Smith & West, 2016; West *et al*., 2016), references (e.g., Kessler, 1963) and user-provided keywords (e.g., Hong *et al*., 2013). Citations, however, take significant time to accrue (Wang *et al*., 2013), references may require full-text versions of articles, and keyword similarities rely on authors' personal opinions, producing poor agreement with expert classification (Achakulvisut *et al*., 2016). Therefore, we use Natural Language Processing (NLP) techniques to find *semantic* similarity between sets of documents based on their raw text. This is a faster and potentially more accurate way to measure similarities (Achakulvisut *et al*., 2016; Manning *et al*., 2008).

Measuring similarities using NLP involves computing the distances between documents—either abstracts of publications or grant text. We represent documents numerically using topic modeling (Methods Appendix 3). These topics allow us to consider correlations between words or set of words (e.g., "clean energy" vs "renewable energy" are similar concepts) while simultaneously differentiating contextual word usage (e.g., "solar power" vs "statistical power" are different concepts). For the distance computation, we use the *cosine similarity*, which is robust to varying document sizes (e.g., distance between a publication abstract and the full-text of a grant).

To illustrate how our similarity measure works, we performed a preliminary analysis of 16 scientists funded by two separate ARPA-E programs and examined how the similarity of the principal investigators' publications to the FOA changed before and after receiving funding. Using the Scopus API, we obtained publications of ten researchers from the “Electrofuels” program and six from the “Green Electricity Network Integration (GENI)” program. These researchers were the primary contact faculty associated with the grant, which typically lasted between 3 to 5 years starting from 2010. We analyzed a total of 3,331 publications published before or after the grant. We defined before as the four-year window before the grant start and after as the four-year window after the grant end. The publication's abstracts and FOA of the grants were modeled using a 10-dimensional topic model (see Methods Appendix 3).



**Figure 2. Preliminary analysis of 16 grants and the publications of their principal investigators.**

Figure 2 shows the similarity measure before and after receiving funding. The publications of funded scientists moved significantly closer to the grant topic after funding (blue bars, two-sample t-test, *t*(15) = 2.07, *p* = 0.02). As a validation exercise, we also compute the similarity scores between each researcher’s publications and the other ARPA-E program. The two sample programs attracted very different sets of scientists. Electrofuels funded a range of biologists, chemists, and chemical engineers whose work could be applied to alternative fuels, whereas GENI primarily attracted engineers working on electricity transmission. Our validation exercise shows that a researcher’s publication record is less similar to the other FOA, and that research moves no closer to the topic of the other FOA (orange bars, *t*(15) = 0.74, *p* = 0.46).

*This preliminary analysis shows promising avenues for further exploration*. While the sample set was small and does not include a control group for comparison, we are able to detect changes in the direction of research. In its full form, the project contemplates an analysis that gets at the causality of such effects.

***Does Funding Affect the Direction of Research? Which Types of Programs Influence Direction More?***

Our first two research questions ask whether funding influences the direction of research. To make causal inferences we must control for other external factors affecting the choices of both funded and non-funded scientists. For example, as environmental concerns such as climate change become more pressing, scientists may choose to enter the field of clean energy even if targeted funding opportunities are not available. Our goal is to identify a control group of authors who are similar to funded scientists in research interest, career stage, and other features. Our methods will identify such scientists *up to the time in which our treated scientists start receiving funding*. We can then ask whether our treated scientists were more likely to publish papers on clean energy than other scientists with similar profiles.

Our method therefore will construct a synthetic experimental dataset to get at the causality of why scientists shift their interests. Our proposed method will find synthetic controls that, unlike other methods, fulfill several criteria simultaneously. The controls will be topically close, have close matches on observed characteristics, and do not overuse the same individual as controls. Because of the large size of publication and author datasets, we will follow three steps to efficiently identify potential control scientists. *First*, using the Scopus publication database, which provides unique identifiers for each author, we will use subject, topic, and date filters to narrow the set of potential controls. *Second*, since some potential controls will be unlikely matches for our funded researchers, we use machine learning to estimate the likelihood of a match given a potential control–funded scientist pair. Using machine learning for matching, rather than traditional matching methods such as propensity score matching or nearest neighbor matching, allows us to evaluate the quality of the overlap for multiple features, such as the similarities between overall research topics, research specific to the FOA, number of publications, number of citations, number of co-authors, topic breadth, years of experience, and the quality of an author’s research institution. *Third*, we use linear programming to find the top *k* controls for each treatment while trying to maximize the total likelihood of the matches. These steps will keep our dataset construction computationally tractable while ensuring good matching quality. Methods Appendix 4 provides greater detail on each of these three steps, as well as potential robustness checks to ensure the quality of our matches compared to standard matching methods in the literature.

Once we identify our control group, we propose two sets of regressions to isolate the effect of program funding from other researcher-specific and external forces influencing research direction. Figure 3 summarizes these regressions, which are described in greater detail in Methods Appendix 5. Our first set (R1) includes only recipients of targeted DOE funded opportunities and their controls. Here, we are asking whether targeted research programs move researchers closer to the goals of the FOA (Q1). We will allow for different effects between ARPA-E and EERE, to test whether ARPA-E truly is more effective attracting new researchers to the field.

|  |  |  |
| --- | --- | --- |
|  | R1: **Similarity to the targeted FOA** | R2: **Similarity to the proposed abstract** |
| Sample | all researchers (and their controls) funded by targeted ARPA-E & EERE programs | all researchers (and their controls) funded by our DOE and NSF programs |
| Research question: | Q1: Do scientists change research focus in response to targeted funding opportunities? | Q2: What types of calls for funding best attract new researchers? |
| Dependent variable: *option 1* | Count of publications sufficiently “close” to the FOA | Count of publications sufficiently “close” to the research proposal abstract |
| Dependent variable: *option 2* | Continuous measure of similarity between a researcher’s publications in year *t* and the funding announcement | Continuous measure of similarity between a researcher’s publications in year *t* and the research proposal abstract |

**Figure 3: Summary of Regressions on Funding and the Direction of Research**

Our second set of regressions (R2) includes researchers (and their controls) funded by all the DOE and NSF programs described in Figure 1. Here, we are asking whether targeted programs and/or programs encouraging research with interdisciplinary and novel perspectives move researchers further away from their existing research than other programs. We allow for separate effects from targeted and open calls for proposals, as well as for programs specifically indicating preference for high-risk/high-reward research.

For each set of regressions, we propose two separate dependent variables to measure the similarity between the researcher’s research in a given year and the FOA (R1) or research abstract (R2). In option 1, our dependent variable is a count of publications in each year that are sufficiently “close” to either the FOA or the proposed research abstract, obtained using machine learning to identify topics related to the funded research.[[1]](#footnote-1) While this method uses counts of publications that are straightforward to interpret, it requires an arbitrary decision of how similar related work must be to be counted. Thus, as a robustness check, option 2 simply uses the continuous measure of similarity between a researcher’s publications in each year and the funding announcement (or funded proposal), as defined in the previous section. If other researchers are moving in similar directions (e.g. it is not funding, but other external forces changing the direction of research), we will observe little correlation between funding and similarity.

For identification, our main specification uses a differences-in-differences research design, using researchers receiving funding from program *p* as the treated group. Here being “funded” means receiving a grant from a specific program. It does not mean receiving any research funding. It may be the case that non-treated scientists obtain other funding sources. The difference-in-difference regression allows us to control for general trends affecting all scientists, with the interaction of funded researchers and the years post-funding capturing the additional effects of funding from program *p* on funded scientists. We will allow for lagged effects multiple years post-funding, as Popp (2016) finds that the effect of public funding on publication counts may persist for six years or more.

Assuming our control group is well-matched, the difference-in-differences specifications cleanly identify the marginal effect of receiving funding from a specific government program. However, as it relies on multiple fixed effects, it does not allow us to observe how specific researcher characteristics affect research direction. For example, are junior scientists influenced more by funding (e.g. are we building human capital) or less (e.g. junior scientists work in a lab and then move on to areas of their own interest)? Thus, we will also estimate a richer model that includes various researcher controls in place of these fixed effects. For the treatment effect in such a model to be causal, there must be no additional unobserved variables that affect research choice. A comparison between the results from these models and the difference-in-differences specification allows us to verify whether our controls are sufficient. Potential controls include the skewness of a scientist’s portfolio (to ask whether researchers with more variety in their publication record are more likely to be influenced by funding opportunities) and years of experience.

***Do researchers new to a field contribute novel ideas? Do they produce more highly cited research?***

Our last set of research questions ask whether funding agencies *should* encourage new scientists to move to a field. We consider two potential benefits from new researchers: (1) Do these scientists bring in ideas that are *novel* to the field? (2) Do the publications of these researchers have higher *impact*, as measured through citations? To compare a given research article to the field as a whole, here we use the broader set of articles published from 1991-2017 described earlier. Because our analysis focuses on U.S. funding agencies, we will restrict our analysis to papers with at least one U.S. author.

Our proposed research is the first to consider how the distance from a scientist’s previous work affects the impact and novelty of a publication. Using text-based analysis of a scientist’s publication record, we consider both the *similarity* of author *a*’s publication *p* to the previous publications of author *a* and the *breadth* of this researcher’s prior work. Among two researchers with comparable similarity measures to a given paper topic, a researcher with additional experience in other areas may be able to combine the knowledge from each field in novel ways. Our breath measure, described in greater detail in Methods Appendix 3, allows us to ask whether researchers working on a wider range of topics contribute more valuable knowledge than other researchers. More specifically, we can use a body of work (e.g., set of publications or grants, or both) to construct a topic distribution of it. This topic distribution can then be used to estimate the research breadth of such body of work using, for example, the distribution's entropy. This entropy quantifies how far away the body of work is from a very topically-concentrated distribution, an intuitive proxy for research breadth.

One challenge is that most publications include multiple authors. Thus, the breadth of research experiences encompassed by the entire research team also offers opportunities for new ideas to emerge. We will include multiple measures to account for differences among research team member, performing sensitivity analysis on multiple combinations of similarity and research breadth measures. These include:

* Similarity and breadth measure(s) of the lead and last authors
* Similarity and breadth measure(s) of the most senior author
* Lowest and highest similarity measure of the team: is there a new and/or established researcher?
* Range of similarity measure(s) for the team: how diverse is the research team?

Our regression analysis focuses on two key outcome measures: (1) *noveltyp*, the novelty of publication *p* compared to the previous literature in the field, and (2) *impactp*, the impact of publication *p* measured by total citations received. Because our interest is in how new research builds upon previous work, we include self-citations in our analysis. To measure the *novelty* of a publication compared to the previous literature in the field, we will apply machine learning techniques that detect outliers. For any given year and research field, we will build an outlier detection algorithm that learns what "normal" research looks like, using documents from a range of past years. The detection method estimates how much of an outlier each new publication is compared to other research published that year. Because the novelty of a paper will change depending on the window used to define the existing field, we will experiment with different windows (e.g. articles from the previous five years, ten years, or the full set of articles since 1991) to explore sensitivity to the appropriate window size for defining novelty. Ideally, any window used should show that novel papers become harder to publish as time progresses and a field becomes more crowded.

We propose regressions for each dependent variable using the modeling strategy first used by one of the PI’s in earlier research on energy funding (Popp 2016, 2017). Methods Appendix 5 provides more detail. These regressions control for other factors affecting novelty or impact. For example, the number of citations received depends not only on article impact, but also on the number of opportunities for citation. For energy publications, we would expect the raw number of citations to be higher during periods of intense research activity, such as when energy prices are highest. Similarly, novelty varies over time as creating novel research becomes more difficult in a more crowded field. Year fixed effects control for these changes over time. Controls for other article characteristics include number of authors, as papers with larger research teams are cited by other articles more frequently (Wuchty *et al*. 2007) and are more novel (Lee *et al.* 2015).[[2]](#footnote-2) We include a dummy variable for papers with authors from multiple countries, which may be exposed to a broader set of potentially citing papers. Finally, some specifications will include journal impact factors. While this limits the comparison of impact and novelty to other similar journals, it will be important if different sub-fields publish in journals with very different impact factors. Finally, we will estimate some models with author fixed effects, allowing us to control for unobserved differences across authors, such as author prestige, that affect novelty or citations. However, these models offer a narrower interpretation, asking whether researchers working on new topics generate more highly cited (or more novel) work than their other research. Thus, we will include models both with and without author fixed effects.

Because the work of these researchers may receive more citations because it has higher impact on future work, or simply because the work is exposed to other researchers in a broader set of fields, for this dependent variable we will also perform an additional set of regressions that also include our measures of similarity, breadth, and novelty as explanatory variables. By conditioning on each, we can see, for example, whether less similar researchers have greater impact even when controlling for article novelty. If that is the case, that suggests that these researchers not only contribute new ideas to the field, but also expose their research to a broader set of scientists.[[3]](#footnote-3) For example, Theodoris (2018) finds that generalists play an important role facilitating collaborations. As average novelty and citation rates should vary across fields, we will estimate separate models for each of our four technologies.

**5. What is the work plan?**

Our proposed research would begin in the summer of 2020. During year 1, Dr. Acuña will lead (i) data collection, including developing a set of control researchers, and (ii) development of the synthetic control matching methods. Dr. Popp and his doctoral student will provide guidance on appropriate matches for the controls, and will test the validity of the controls. Additional data refinement will continue in year 2, led by Dr. Acuña, including development of the data set required for our third research question. Dr. Acuña will also lead the development of a methodological paper on our synthetic control method. Dr. Popp will lead the initial data analysis and dissemination of these results. In year 3 we will complete data analysis and prepare final manuscripts.

**6. What will be the output from the research project?**

We expect this project to generate at least three manuscripts that will contribute to both the academic literatures on both energy and science policy, as well as to related practitioners. One manuscript will address whether funding affects the direction of research and, if so, do targeted funding programs move researchers further than open funding programs. A second manuscript will explore whether researchers new to a topic generate more novel and more impactful research. This paper will help guide the decisions of funding agencies by identifying whether targeting funding at new researchers provides greater impact. In addition, a third manuscript will compare the results of machine learning to other commonly used matching techniques such as propensity score matching, so as to inform future research using matching techniques. To disseminate the research findings to the appropriate policy communities, we will submit at least one manuscript from this research to an interdisciplinary journal read by policymakers, such as the *Journal of Policy Analysis and Management* or *Research Policy*. In addition, both PI regularly present to audiences that include both academics and practitioners.

Our project will also generate code and data for use by others in the scientific community. Data analysis will be carried out in Stata and Python. The software will be made available through a public Github repository. While the Scopus and Web of Science data cannot be made publicly available, we will share the code used to download the data necessary to re-run our analysis. Our project will train multiple graduate students as described in the next section.

**7. What is the (summary) justification for the amount of money requested?**

Our budget supports the Principal Investigators and a team of graduate students who will assist with data collection and analysis. Our interdisciplinary team will support graduate students working in both Syracuse University’s Maxwell School and the School of Information Studies. Dr. Acuña will supervise a student at the School of Information Studies who will build the queries to access the Scopus API and process Web of Science data. This student will also assist writing reports, papers, and software documentation to disseminate the results. Dr. Popp will supervise an advanced Ph.D. student in Public Administration, who will assist Dr. Popp with regression analysis and manuscript preparation. This student will also help with preparation of the data sets, including identifying funded researchers at the agencies studied and assisting the evaluation of training data sets in year 1 of the project, and helping to perform statistical analysis and prepare the results for publication in years 2 and 3. We include funding for both the academic year and two summer months for each graduate student. Our budget also includes funding for Dr. Acuña and a graduate student to participate in one professional conference per year. Dr. Popp has access to other university funds that will support his travel related to this project. To encourage wide dissemination of the results, we request $2,000 per year for years two and three of the project for publishing two articles in open access journals.

**8. What other sources of research support does the proposer have in hand or has applied for?**

While we have no other funding directly supporting this research, Dr. Acuña has an NSF-SciSIP grant entitled “Collaborative Research: Social Dynamics of Knowledge Transfer Through Scientific Mentorship and Publication” that has synergistic activities with the current application. The grant, which runs through September 2021, will produce datasets about publications of individual scientists using the same data sources and will train graduate students to produce similar text analysis.

**9. What is the status and output of current and/or previous Sloan grants?**

Popp is currently co-PI on a Sloan grant through the National Bureau of Economic Research, “Economics of Innovation in the Energy Sector”. This project supports fourteen new research studies on energy innovation, to be presented at a series of workshops organized by Popp and his co-PI, Ashley Langer of the University of Arizona. As this project runs from July 1, 2019- June 30, 2022, there is little to report at this time. Submissions for the first workshop were due October 31. We received many more good submissions than expected (37), and are now deciding among them to select 6 papers for the first workshop.

**Appendix A: Budget and Budget Justification**



*Salaries and Wages – Senior Personnel*

The Principal Investigator, Dr. David Popp, will devote 1.0 summer month per year throughout the three-year project period. He will be responsible for designing and carrying out the regression analysis, and will collaborate with Dr. Acuna co-PI on construction of the data set and similarity measures and on manuscript preparation. He will also supervise graduate students working on data analysis. As Dr. Popp’s academic program, Public Administration and International Affairs, is a 12-month degree program, this support will enable Dr. Popp to minimize commitments to academic programs during this summer time.

The Co-Principal Investigator, Dr. Daniel Acuna, will devote 1.0 summer month per year throughout the three-year project. He will be responsible for construction of the data set and similarity measures, and will collaborate with Dr. Popp on the regression analysis, on manuscript preparation, and dissemination of the work. He will also supervise graduate students to further refine and expand the data set as needed.

Effort spent on the project but not compensated for within the academic year is deemed to be included within the faculty member’s regular organizational duties. Syracuse University faculty appointments are for 8.5 months. Salaries are escalated by 3.0% annually for budget preparation purposes; actual salaries in place during the time of the award are charged.

*Salaries and Wages – Other Personnel*

Funding for one graduate student, 9 AY mo. and 2 summer months (0.5 FTE per month), is requested for all three years of the grant to work with Dr. Popp. Academic graduate assistantships support 20 hours per week during the academic year. The graduate student will be an advanced Ph.D. student in Public Administration. The student assist Dr. Popp with regression analysis and manuscript preparation. This student will also help with preparation of the data sets, including identifying funded researchers at the agencies studied and assisting the evaluation of training data sets in year 1 of the project, and helping to perform statistical analysis and prepare the results for publication in years 2 and 3. Salaries are escalated by 3.0% annually for budget preparation purposes.

Funding for one graduate student, 9 AY months and 2 summer months (0.5 FTE per month), is requested for years one and two to work with Dr. Acuna. The graduate student will build the queries to access the Scopus API and process Web of Science data. This GA will also assist writing reports, papers, and software documentation to disseminate the results.

*Fringe Benefits*

Fringe Benefits are calculated as direct costs in accordance with Syracuse University’s indirect cost rate agreement (Department of Health and Human Services, dated July 10, 2019, 16% for faculty during the summer; 29.9% for faculty during the academic year and full time staff, 15.4% for graduate students and 7.9% for undergraduate students and temporary staff). Actual rates in place during the time of the award would be charged.

*Travel*

A total of $3,600 is requested for Dr. Acuna and the graduate student to participate in one professional conference per year. Dr. Popp has access to other university funds that will support his travel related to this project. Estimated annual costs per single traveler are: $550 for round trip air-fare, registration fee of $500, lodging of $150/day for three days, and $75 for meals and incidentals.

*Publication Costs/Documentation/Dissemination*

We request $2,000 per year for years two and three of the project for publishing two articles in open access journals (i.e., 'gold' open access).

**Appendix B: List of Citations**

Achakulvisut, T., Acuna, D. E., Ruangrong, T., & Kording, K. 2016. Science Concierge: A fast content-based recommendation system for scientific publications. *arXiv preprint arXiv:1604.01070*.

Acuna, D. E., Achakulvisut, T., & Kording, K. P. 2015. *Automated reviewer-manuscript matching and article scoring*. Paper presented at the Science of Team Science (SciTS) Conference.

Acuna, D. E., Allesina, S., & Kording, K. P. 2012. Future impact: Predicting scientific success. *Nature, 489*(7415), 201-202.

Acuna, D. E., Brookes, P. S., & Kording, K. P. 2018. Bioscience-scale automated detection of figure element reuse. *bioRxiv.*

Aggarwal, C. C., Hinneburg, A., & Keim, D. A. 2001. *On the surprising behavior of distance metrics in high dimensional spaces.* Paper presented at the ICDT.

Azoulay, P., Bonatti,A. & Krieger, J.L. 2017. The career effects of scandal: Evidence from scientific retractions. *Research Policy* 46: 1552-1569.

Azoulay, P., Graff Zivin, J.S.., & Manso, G. 2011. Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics*, 42(3):527–554.

Blei, D. M., Ng, A. Y., & Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

Blume-Kohout, M.E. 2012. Does targeted, disease-specific public research funding influence pharmaceutical innovation? *Journal of Policy Analysis and Management* 31(3), 641-660.

Caruana, R., Karampatziakis, N., & Yessenalina, A. 2008. *An empirical evaluation of supervised learning in high dimensions.* Paper presented at the Proceedings of the 25th international conference on Machine learning.

Corredoira, Rafael A., Brent D. Goldfarb, and Yuan Shi. 2018. Federal funding and the rate and direction of inventive activity. *Research Policy* 47: 1777-1800.

Dasgupta, S., & Langford, J. 2009. *A tutorial on active learning.* Paper presented at the International conference on machine learning.

Dechezleprêtre, A. and M. Glachant. 2014. “Does Foreign Environmental Policy Influence Domestic Innovation? Evidence from the Wind Industry.” *Environmental and Resource Economics*. 58 (3): 391-413.

Furman, J. & Teodoridis, F. 2018. The cost of research tools and the direction of innovation: Evidence from computer science and electrical engineering. Working Paper. <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3060599>

Gelman, A., & Hill, J. 2007. *Data analysis using regression and multilevelhierarchical models* (Vol. 1): Cambridge University Press New York, NY, USA.

Goldstein, A.P. and Narayanamurti, V. 2018. Simultaneous pursuit of discovery and invention in the US Department of Energy. *Reserch Policy* 47(8): 1505-1512.

Hong, K., Jeon, H., & Jeon, C. 2013. Personalized research paper recommendation system using keyword extraction based on userprofile. *Journal of Convergence Information Technology, 8*(16), 106.

ITIF 2018. “An Innovation Agenda for Deep Decarbonization: Bridging Gaps in the Federal Energy R&D&D Portfolio.” November 2018.

Jacob, B.A., Lefgren, L. 2011. The impact of research grant funding on scientific productivity. *Journal of Public Economics* 95: 1168-1177.

Johnstone, N. I. Hascic and D. Popp. 2010. “Renewable Energy Policies and Technological Innovation: Evidence Based on Patent Counts.” *Environmental and Resource Economics.* 45(1): 133-155.

Kallus, N. 2016. Generalized optimal matching methods for causal inference. *arXiv preprint arXiv:1612.08321*.

Kessler, M. M. 1963. Bibliographic coupling between scientific papers. *Journal of the Association for Information Science and Technology, 14*(1), 10-25.

Lancichinetti, A., Sirer, M. I., Wang, J. X., Acuna, D., Körding, K., & Amaral, L. A. N. 2015. High-reproducibility and high-accuracy method for automated topic classification. *Physical Review X, 5*(1), 011007.

Lee, Y.N., Walsh, J.P. & Wang, J. 2015. Creativity in scientific teams: Unpacking novelty and impact. *Research Policy* 49: 684-697.

Liénard, Jean F., Titipat Achakulvisut, Daniel E. Acuna, and Stephen V. David. 2018. "Intellectual synthesis in mentorship determines success in academic careers." *Nature communications* 9(1): 4840.

Manning, C. D., Raghavan, P., & Schütze, H. 2008. *Introduction to information retrieval* (Vol. 1): Cambridge university press Cambridge.

Merton, R.K. 1957. Priorities in scientific discoveries: A chapter in the sociology of science. *American Sociological Review* 22: 635-659.

Murphy, K. P. 2012. *Machine learning: a probabilistic perspective*: MIT press.

Myers, K.R. 2016. The Direction of Biomedical Science,” prepared for the NBER-IFS International Network on the Value of Medical Research, available at <https://docs.wixstatic.com/ugd/73b908_af32174a9d70426ea684e495c6f34b58.pdf>.

Myers, K.R. 2019. “The Elasticity of Science”, forthcoming in *American Economic Journal: Applied* Economics, working paper available at <https://www.kylemyers.org/research>.

National Academies of Sciences, Engineering, and Medicine. 2017. An Assessment of ARPA-E. Washington, DC: The National Academies Press. doi: <https://doi.org/10.17226/24778>.

National Science Foundation 2018. *Proposal & Award Policies and Procedures Guide*, NSF 18-1, <https://www.nsf.gov/pubs/policydocs/pappg18_1/nsf18_1.pdf>.

Nemet, G.F., 2012. “Knowledge Spillovers from Learning by Doing in Wind Power.” *Journal of Policy Analysis and Management*. 31(3): 600-621.

Nesta, L., F. Vona, and F. Nicolli. 2014. “Environmental Policies, Competition, and Innovation in Renewable Energy.” *Journal of Environmental Economics and Management*. 67: 396-411.

Office of Energy Efficiency and Renewable Energy. 2015. *2016-2020 Strategic Plan and Implementing Framework*, DOE/EE-2061, <https://energy.gov/sites/prod/files/2015/12/f27/EERE_Strategic_Plan_12.16.15.pdf>.

Peters, M., M. Schneider, T. Griesshaber and V.H. Hoffman. 2012. “The impact of technology-push and demand-pull policies on technical change–Does the locus of policies matter?” *Research Policy*. 41 (8):1296-1308.

Popp, D. 2010. “Innovation and Climate Policy,” *Annual Review of Resource Economics*, *vol. 2.* Gordon C. Rausser, V. Kerry Smith and David Zilberman eds., Annual Reviews, Palo Alto, CA, 275-298 (<http://dx.doi.org/10.1146/annurev.resource.012809.103929>).

Popp, David. 2015. “Climate-Friendly Technological Change for Developing Countries,” in *Oxford Handbook of Macroeconomics of Global Warming*, edited by Lucas Bernard and Willi Semmler, Oxford University Press, pp. 321-348.

Popp, D. 2016. Economic analysis of scientific publications and implications for energy research and development. Nature Energy 1(4): 1–8.

Popp, D. 2017. From Science to Technology: The Value of Knowledge From Different Energy Research Institutions. *Research Policy* 46(9): 1580-1594.

Popp, David. 2019. Environmental Policy and Innovation: A Decade of Research. *International Review of Environmental and Resource Economics*, 2019, 13(3-4), 265-337.

Popp, D., R. Newell and A.B. Jaffe 2010. “Energy, the Environment, and Technological Change.” In *Handbook of the Economics of Innovation: vol. 2*, Bronwyn Hall and Nathan Rosenberg, eds., Academic Press/Elsevier, 873-937.

Rosenbaum, P. R. 1989. Optimal matching for observational studies. *Journal of the American Statistical Association, 84*(408), 1024-1032.

Rosenbloom J.L., Ginther D.K., Juhl T., & Heppert J.A. The Effects of Research & Development Funding on Scientific Productivity: Academic Chemistry, 1990-2009. *PLoS ONE* 10(9): e0138176. doi:10.1371/journal.pone.0138176 (2015).

Stephan, P.E. 2010. The Economics of Science. *Handbook of the Economics of Innovation: vol. 1*, Bronwyn Hall and Nathan Rosenberg, eds., Academic Press/Elsevier, 217-273.

Stephan, P.E. 1996. The economics of science*. Journal of Economic Literatur*e, 34(3):1199–1235.

Taylor, C. J. 2008. On the optimal assignment of conference papers to reviewers. Technical Report MS-CIS-08-30, University of Pennsylvania, 2008.

Teplitskiy, M., D. Acuna, A. Elamrani-Raoult, K. Körding, J. Evans. 2018, The sociology of scientific validity: How professional networks shape judgement in peer review, *Research Policy*, (in press).

Theodoris, Florenta 2018. Understanding Team Knowledge Production: The Interrelated Roles of Technology and Expertise. *Management Science* 64(8):3625-3648.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. 2013. Atypical Combinations and Scientific Impact.  *Science* 342(6157: 468-472.

Verdolini, E. and M. Galeotti. 2011. “At Home and Abroad: An Empirical Analysis of Innovation and Diffusion in Energy Technologies.”  *Journal of Environmental Economics and Management*. 61: 119–134.

Wang, D., Song, C., & Barabási, A.-L. 2013. Quantifying long-term scientific impact. *Science, 342*(6154), 127-132.

Wang, J, Veugelers, R. & Stephan, P.. 2017. Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy* 46: 1416-1436.

Wesley-Smith, I., & West, J. D. (2016). *Babel: a platform for facilitating research in scholarly article discovery.* Paper presented at the Proceedings of the 25th international conference companion on world wide web.

West, J. D., Wesley-Smith, I., & Bergstrom, C. T. (2016). A recommendation system based on hierarchical clustering of an article-level citation network. *IEEE Transactions on Big Data, 2*(2), 113-123.

Wu, Lingfei, Dashun Wang and James Evans. 2019. Large teams develop and small teams disrupt science and technology. *Nature* 566: 378-382.

Wuchty, S., Jones, B.F., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. *Science* 316, 1036-1039.

**Appendix C: Curricula Vitae**

**David Popp**

**Caroline Rapking Faculty Scholar in Public Administration and Policy**

Center for Policy Research

Syracuse University Ph: 315-443-2482

The Maxwell School Fax: 315-443-1081

426 Eggers Hall E-mail: dcpopp@maxwell.syr.edu

Syracuse,NY 13244-1020  
Home page: <https://dcpopp.expressions.syr.edu/>

**Date of Birth:** October 30, 1970 **Citizenship:** U.S.A.

**Education:** Ph.D. – Yale University, Department of Economics, 1997

*Fields of Concentration:* Economics of Natural Resources and the Environment

Public Finance

*Dissertation Title*: *Induced Innovation, Energy Prices, and the Environment*

*Dissertation Committee*: William Nordhaus, Robert Evenson, Joel Waldfogel

BA – Williams College, 1992 Major: Political Economy

**Professional Experience:**

Professor of Public Administration and International Affairs, Syracuse University, May 2014- (Associate Professor May 2006-May 14; Assistant Professor August 2000-May 06)

Research Associate, National Bureau of Economic Research (NBER) Environmental and Energy Economics Program, May 2007-present

Research Associate, NBER Productivity Program, September 2006 – present

Research Network Member, Energy & Climate Economics Research Group, CESifo, Feb. 2009-

Senior Research Associate, Center for Policy Research, Syracuse University, January 2006-

Senior Research Associate, Center for Environmental Policy Administration, SU, August 2000-

Senior Research Associate, Center for Technology and Information Policy, SU, August 2000-15

Consultant, World Bank, 2011-2012

Consultant, OECD Directorate, 2005-11

Faculty Research Fellow, NBER Productivity Program, March 2001 – Aug. 2006

Assistant Professor of Economics, The University of Kansas, August 1997– August 2000

**Selected Published Papers:**

**Popp, David**, “Promoting Innovation for Low-Carbon Technologies,” Policy Proposal 2019-14, The Hamilton Project, Brookings Institution, Washington, DC.

**Popp, David**, “Environmental Policy and Innovation: A Decade of Research,” *International Review of Environmental and Resource Economics*, 2019, 13(3-4), 265-337.

Vona, Francesco, Giovanni Marin, Davide Consoli, and **David Popp**, “Environmental Regulation and Green Skills: an empirical exploration,” *Journal of the Association of Environmental and Resource Economists*, October 2018, 5(4), 713-753.

Hepburn, Cameron, Jacquelyn Pless, and **David Popp**, “Encouraging Innovation that Protects Environmental Systems: Five Policy Proposals,” *Review of Environmental Economics and Policy*, Winter 2018, 12(1), 154-169.

**Popp, David**, “From Science to Technology: The Value of Knowledge From Different Energy Research Institutions,” *Research Policy*, November 2017, 46(9), 1580-1594.

Webster, Mort D., Karen Fisher-Vanden, **David Popp**, and Nidhi R. Santen, “Should We Give Up After Solyndra? Optimal Technology R&D Portfolios under Uncertainty,” *Journal of the Association of Environmental and Resource Economists*, September 2017, 4(S1), S123-S151.

**Popp, David**, “Economic Analysis of Scientific Publications and Implications for Energy Research and Development,” *Nature Energy*, April 2016, 1(4), 1-8, DOI: 10.1038/nenergy.2016.20.

Tang, Tian and **David Popp**, “The Learning Process and Technological Change in Wind Power: Evidence from China’s CDM Wind Projects,” *Journal of Policy Analysis and Management*, Winter 2016, 35(1), 195-222.

Miao, Qing and **David Popp**,“Necessity as the Mother of Invention: Innovative Responses to Natural Disasters,” *Journal of Environmental Economics and Management*, September 2014, 68(2), 280-295.

**Popp, David**, Nidhi Santen, Karen Fisher-Vanden and Mort Webster, “Technology Variation vs. R&D Uncertainty: What Matters Most for Energy Patent Success?” *Resources and Energy Economics*, November 2013, 35(4), 505-533.

**Popp, David** and Richard Newell, “Where Does Energy R&D Come From? Examining Crowding out from energy R&D,” *Energy Economics*, July 2012, 34(4), 980-991.

**Popp, David**, Tamara Hafner, and Nick Johnstone, “Environmental Policy vs. Public Pressure: Innovation and Diffusion of Alternative Bleaching Technologies in the Pulp Industry,” *Research Policy*, November 2011, 40(9), 1253-1268.

**Popp, David**, Ivan Hascic and Neelakshi Medhi, “Technology and the Diffusion of Renewable Energy,” *Energy Economics*, July 2011, 33(4), 648-662.

Lovely, Mary and **David Popp**, “Trade, Technology, and the Environment: Does Access to Technology Promote Environmental Regulation,” *Journal of Environmental Economics and Management*, January 2011, 61(1), 16-35.

**Popp, David**, “Exploring the Links Between Innovation and Diffusion: Adoption of NOX Control Technologies at U.S. Coal-fired Power Plants,” *Environmental and Resource Economics*, March 2010, 45(3), 319-352.

Johnstone, Nick, Ivan Hascic and **David Popp**, “Renewable Energy Policies and Technological Innovation: Evidence Based on Patent Counts,” *Environmental and Resource Economics,* January 2010, 45(1), 133-155.

**Popp, David**, “Innovation in Climate Policy Models: Implementing Lessons from the Economics of R&D,” *Energy Economics*, 28(5-6), November 2006, 596-609.

**Popp, David**, “R&D Subsidies and Climate Policy: Is there a ‘Free Lunch’?” *Climatic Change*, 77(3-4), August 2006, 311-341.

**Popp, David**, “International Innovation and Diffusion of Air Pollution Control Technologies: The Effects of NOX and SO2 Regulation in the U.S., Japan, and Germany,” *Journal of Environmental Economics and Management*, 51(1), January 2006, 46-71.

**Popp, David**, “ENTICE: Endogenous Technological Change in the DICE Model of Global Warming,” *Journal of Environmental Economics and Management*, 48(1), July 2004, 742-768.

**Popp, David**, “Pollution Control Innovations and the Clean Air Act of 1990,” *Journal of Policy Analysis and Management*, 22(4), Fall 2003, 641-660.

**Popp, David**, “Induced Innovation and Energy Prices,” *American Economic Review*, 92(1), March 2002, 160-180.

**Selected Book Chapters**

Dechezleprêtre , Antoine and **David Popp** 2017. “Fiscal and Regulatory Instruments for Clean Technology Development in the European Union,” Chapter 6 in Energy Tax and Regulatory Policy in Europe: Reform Priorities, edited by Ian Parry, Karen Pittel, and Herman Vollebergh, MIT Press, pp. 167-213.

**Popp, David**, “Climate-Friendly Technological Change for Developing Countries,” 2015, in *Oxford Handbook of Macroeconomics of Global Warming*, edited by Lucas Bernard and Willi Semmler, Oxford University Press, pp. 321-348.

**Popp, David**, “Innovation and Climate Policy,” *Annual Review of Resource Economics*, vol. 2., 2010, Gordon C. Rausser, V. Kerry Smith and David Zilberman eds., Annual Reviews, Palo Alto, CA, 275-298.

**Popp, David**, Richard Newell and Adam Jaffe, “Energy, the Environment, and Technological Change,” *Handbook of the Economics of Innovation: vol. 2*, Bronwyn Hall and Nathan Rosenberg, eds., Academic Press/Elsevier, 2010, 873-937.

**Selected Grants, Fellowships, Honors, and Awards:**

2017 Association of Environmental and Resource Economists Publication of Enduring Quality

National Science Foundation grant # SMA-1064161: “Using Scientific Publications to Evaluate Government R&D Spending: The Case of Energy,” June 2011-May 2014.

US Department of Energy, “Integrated Assessment Model Development, Comparison, and Diagnostics Project,” subcontract with Penn State University, September 2010-August 2013.

US Environmental Protection Agency, “Environmental & Resource Economics Workshop: ‘Climate Policy Without Cost? Can Technology Solve the Climate Problem?’,” 2005-2006 (with Peter Wilcoxen).

US Department of Energy grant #DE-FG02-04ER63927: “Knowledge Spillovers and the Opportunity Cost of Climate Mitigation R&D” September 2004 - August 2007 (with Richard Newell).

US Department of Energy grant DE-FG02-ER63467: “International Innovation and Diffusion of Environmental Technologies: The Case of NOX,” September 2002-August 2004.

National Science Foundation grant SES-0001679: “Induced Innovation in the DICE model of Global Warming,” August 2000-July 2002.

Selected Professional Service:

*Editorial:*

Co- Editor, Journal of the Association of Environmental and Resource Economists 2014-

Co-Editor, *Environmental and Resource Economics* 2010-

Co-Editor, *Journal of Environmental Economics and Management* 2011-2013

Associate Editor, *Energy Economics* 2004-

*Expert panel/review boards:*

National Science Foundation review panel, 2012

Porter Hypothesis Meta-analysis Expert Panel, Ottawa, ON, November 2012

Member, Advisory Committee of the Green Growth Knowledge Platform, 2012-2016

GAO Expert Panel on Climate Change Economics, 2007

Member, Environmental Protection Agency’s Advisory Council on Clean Air Compliance Analysis, November 2006-2009

**Daniel Ernesto Acuña**

Assistant Professor

Science of Science and Computational Discovery Laboratory ([https://scienceofscience.org](https://scienceofscience.org/))

312 Hinds Hall

School of Information Studies

Syracuse University

Syracuse, NY 13244-1190

**Appointments**

Syracuse University Syracuse, NY 2016-present

Assistant Professor, School of Information Studies

RIC & Northwestern University Chicago, IL 2011-2016

Postdoctoral Researcher in the Sensory Motor Performance Program at RIC

**Education**

University of Minnesota Minneapolis, MN 2011

Ph.D. in Computer Science

Thesis: Rational analysis of sequential decision-making in humans and machines

University of Santiago Santiago, Chile 2004

Master of Engineering Sciences in Information Technology

*Thesis*: An algorithm for the Traveling Salesperson Problem based on players of a computer game

Bachelor in Engineering Sciences (Highest Honors)

**Fellowships and awards**

* Best poster award Metascience 2019 conference (with Han Zhuang)
* DARPA: Systematizing Confidence in Open Research and Evidence (SCORE) (Subcontractor, $ 7,672,188 [$ 129,552]), 02/2019 - 12/2021
* ORI-DHHS-Office of Research Integrity: Methods and tools for scalable figure reuse detection with statistical certainty reporting (sole PI, $150,000), 08/2018 – 07/2019
* NSF-SCISIP: Optimizing Scientific Peer Review (PI, $531,339 [$ 214,144]), 05/2018 – 04/2021
* NSF EAGER: Improving scientific innovation by linking funding and scholarly literature (Sole PI, $ 168,711), 09/2016 – 08/2018
* Microsoft Azure Research Award (US$ 20,000), 2015-2016
* University of Chicago’s Knowledge Lab Grant (co-I) “Optimizing scientific peer review”, 2014-2016
* Amazon AWS Educational Grant “Automatic detection of figure element reuse in biological sciences” (US$ 19,850), 2014-2015
* NIH Neuro-physical-computational Sciences (NPCS) Graduate Training Fellowship (1R90 DK71500-04, full tuition, stipend, conference travel expenses), 2008-2010
* International Graduate Student Fellowship of the Chilean Council of Scientiﬁc and Technological Research and the World Bank (stipend, books), 2006-2010

**Publications**

*Journal articles*

**Acuna, DE**, Teplitskiy, M, Evans, J, Kording, K “Should journals allow authors to suggest reviewers?” (Resubmitted)

**Acuna, DE**, Brookes, P, Kording, K “Automatic detection of figure element reuse in biological science articles” (2018) , BioArxiv

Zeng, T, **Acuna, DE**, "Dead science: most resources linked in scientific articles disappear in eight years", iConference 2019 (to appear in Lecture Notes of Computer Science)

Líenard, JF, Achakulvisut, T, **Acuna, DE**, David, SV (2018) "Intellectual Synthesis in Mentorship Determines Success in Academic Careers", Nature Communications

Teplitskiy, M, **Acuna, DE**, Elamrani-Raoult, A, Körding, K, Evans, J, (2018) The Social Structure of Consensus in Scientific Review, Research Policy

Taraz G. Lee, **Acuna, DE**, K. P., Grafton, S. T. (2018) "Limiting motor skill knowledge via incidental training protects against choking under pressure", Psychonomic Bulletin & Review

Shema, A, **Acuna, DE**, Show Me Your App Usage and I Will Tell Who Your Close Friends Are: Predicting User’s Context from Simple Cellphone Activity, CHI 2017, Pages 2929-2935, Denver, Colorado

Ramkumar, P, **Acuna, DE**, Berniker, M, Grafton, S, Turner, RS, Kording, K (2016) “Chunking as an integral strategy for effective motor learning”, Nature Communications

Achakulvisut, T, **Acuna, DE**, Ruangrong, T and Kording, K (2016). "Science Concierge: A Fast Content-Based Recommendation System for Scientific Publications." PLoS One 11(7): e0158423.

Ethier, C, **Acuna, DE**, Solla, S, Kording, K, Miller, L “Adaptive Neuron-to-Muscle Decoder Training for FES Neuroprostheses”, Journal of Electrophysiology

**Acuna, DE**, Berniker, M, Fernandes, H, Kording, K (2015) “Using psychophysics to ask if the brain samples or maximizes”, Journal of Vision 15(3): 7

Lancichinetti, A, Sirer, MI., Wang, J. X, **Acuna, DE**, Kording, K., Amaral, LAN, (2015) “A high-reproducibility and high-accuracy method for automated topic classification”, Phys. Rev. X 5, 011007

**Acuna, DE**, Wymbs, Nicholas F, Reynolds, Chelsea A., Picard, N, Turner, RS, Strick, PL, Grafton, ST, Kording, KP (2014) “Multi-faceted aspects of chunking enable robust algorithms”, Journal of Neurophysiology Vol. 112 no. 8, 1849-1856

**Acuna, DE**, Penner, O, Orton CG, (2013) “The future h-index is an excellent way to predict scientists' future impact”, Med. Phys. 40, 110601

**Acuna, DE**, Allesina, S, Kording, KP (2012) “Future impact: Predicting scientific success”, Nature, Volume 489, Number 7415, 201-202

Avraham, G, Nisky, I, Fernandes HL, **Acuna, DE**, Kording, KP, Loeb, GE, Karniel A. (2011) “Towards perceiving robots as humans – Three handshake models face the Turing-like handshake test”, IEEE Transactions on Haptics

**Acuna, DE**, Schrater, P. (2010) “Structure learning in human sequential decision-making”, PLoS Computational Biology

**Acuna, DE**, Parada, V. (2010) “People efficiently explore the solution space of the computationally intractable traveling salesman problem to find near-optimal tours”, PLoS ONE 5(7)

*Conference publications*

Sheima, A. **Acuna, DE** (accepted) “Show me your app usage and I will tell who your close friends are: Predicting user's context from simple cellphone activity”, CHI 2017 Late-Breaking Work

**Acuna, DE**, Green, CS, Schrater, P (2010) “The rational control of aspiration in learning”, COSYNE 2010 (Abstract and poster presentation)

**Acuna, DE**, Green, CS, Schrater, P (2010) “Decision-making in unbounded environments using nonparametric Bayesian Reinforcement Learning”, NIPS 2010 Workshop on Bounded-rational analyses of human cognition: Bayesian models, approximate inference, and the brain (Poster presentation)

**Acuna, DE**, Schrater, P. (2009) “Improving Bayesian reinforcement learning using transition abstraction”, ICML/UAI/CLT Workshop on Abstraction in Reinforcement Learning 2009

**Acuna, DE**, Parada, V, Schrater, P (2009) “Skill acquisition and performance on the Traveling Salesman Problem”, Center for Cognitive Science, Spring Research Day (Poster presentation)

**Acuna, DE**, Schrater P.(2009) “Structure learning in human sequential decision-making”, COSYNE 2009

**Acuna, DE**, Schrater, P. (2009) “Structure learning in human sequential decision-making”, NIPS 2008

Acuna, DE, Schrater, P. (2008) “Bayesian modeling of human sequential decision-making on the Multi-Armed Bandit Problem”, COGSCI 2008

*Books*

**Acuna, DE,** (2011) Rational Bayesian analysis of sequential decision-making under uncertainty in humans and machines, Ph.D. Thesis, University of Minnesota-Twin Cities

**Patents**

**Daniel E. Acuna**, Konrad Kording, "System and method for automated detection of figure element reuse", U.S. Provision Patent Application

Konrad Kording, **Daniel E. Acuna**, Titipat Achakulvisut. “Data Butler”. U.S. Provisional Patent Application No. 62/218,998, filed September 15, 2015 (assignee Rehabilitation Institution of Chicago)

**Academic service**

* Inaugural Editorial Board Member of the Journal of Social Computing (to start 2020)
* Associate Chair: Late-Breaking Work CHI 2017
* Reviewer for: IEEE Transactions on Human-Machine Systems, Journal of the Royal Society Interface, Research Evaluation, Operations Research, PLoS Computational Biology, PLoS ONE, Scientometrics, NIPS 2009, NIPS 2010, CogSci 2009
* Training Committee member of the Center for Cognitive Science, University of Minnesota, organized panel discussion on “Job hunting, hiring process and setting up a new lab in academia”

**Memberships**

University of Chicago's Knowledge Lab, Disambiguation Working Group

**Media**

Nature News (2018) "Researchers have finally created a tool to spot duplicated images across thousands of papers", author: Declan Butler

Interviews: The Daily Orange (Syracuse University, 2016), Nature Podcast (2012, The Chronicle of Higher Education (2012), NPR Science Friday (Spanish, 2012), The Scientists (2012)

**Appendix D: Conflicts of Interest**

The PIs have no conflicts of interest related to this project.

**Appendix E: Attention to diversity**

Both Dr. Acuña and Dr. Popp will strive to maintain racial and gender diversity throughout the project. Dr. Acuña is a native of Chile and he actively attempts to connect with the Spanish-speaking scientific community in the U.S. Additionally, Dr. Acuña supervises three Ph.D. students, two from China and one from Bangladesh. He has worked and published with four other Ph.D. students, one from Rwanda and two females. He has worked with 23 international master students from China and India, 6 of whom are female. Dr. Popp will employ two different PhD students during the three years of this project. Both are international. One is female. He has supervised or co-supervised 12 PhD students, of whom 7 were female and 9 were international students.

**Appendix F: Empirical Research Methods**

**Methods Appendix 1: Details on DOE and NSF Programs**

We consider three programs at the DOE that span a range of funding options. The DOE’s Advanced Research Projects Agency-Energy (ARPA-E) began operating in 2009. ARPA-E aims to identify and fund high-risk, potentially high return research in technologies that will “reduce imports of fossil fuels, reduce energy-related emissions, improve energy efficiency ... and ensure that the United States maintains a technological lead in the development and deployment of advanced energy technologies” (National Academies of Science, p. 25). Among ARPA-E’s goals is creating new communities of researchers, including those “not traditionally... involved in the topic area” (National Academies of Science, 2017, p. 122). Thus, changing the direction of research is an explicit goal of the ARPA-E program.

ARPA-E solicits research proposals through two types of Funding Opportunity Announcements (FOA). Focused FOAs *target* specific technology areas. These FOAs include clear, well-defined project goals for potential applicants. For example, the FOA for ARPA-E’s Green Electricity Network Integration program (DE-FOA-0000473) includes four pages of technical performance targets that successful applications must demonstrate an ability to meet. As of July 2018, the ARPA-E website lists 44 archived and current focused programs. *Open* FOAs solicit proposals for potentially disruptive technologies across a wide range of energy applications. Their purpose is to ensure that ARPA-E does not miss potential innovative opportunities in areas outside of the focused FOAs. Open FOAs account for about one-third of ARPA-E’s funding support, which were awarded during four open solicitations (2009, 2012, 2015 & 2018).

ARPA-E was designed to bridge the gap between basic and applied research at DOE (Goldstein and Narayanamurti 2018). For comparison, we consider DOE programs on either side of this spectrum. DOE’s Office of Science supports basic research (Goldstein and Narayanamurti 2018). We will focus on research funded by the Basic Energy Sciences program.[[4]](#footnote-4) Illustrating that these programs are open to a range of ideas, in recent years the Office of Science has issued a single FOA entitled “Continuation of Solicitation for the Office of Science Financial Assistance Program”, which provides brief one or two paragraph descriptions of each program. DOE’s Office of Energy Efficiency and Renewable Energy (EERE) is the applied research program supporting clean energy projects most comparable to those funded by ARPA-E. Like ARPA-E, EERE uses targeted FOAs to solicit research proposals that specify detailed technological improvements to be achieved. Its goal is to address gaps in technology development where the private sector will not invest (Office of Energy Efficiency and Renewable Energy, 2015). EERE provides an example of targeted programs not classified as high-risk/high-reward (i.e., traditional).

NSF includes multiple directorates that oversee programs supporting research in a broad range of science and engineering fields. While each program has different goals and specifies topics of interest on its website, the programs are open in nature. Individual researchers have broad discretion for proposing research to meet program goals – no specific technical guidelines are specified to be met.

Given the large number of programs in NSF, we will study programs most commonly used as funding sources by researchers also funded by DOE, so that we are comparing NSF programs that our DOE funded researchers are likely to consider. For example, Dr. Scott Banta of Columbia University received an ARPA-E award from the Electrofuels program: “Biofuels from CO2 Using Ammonia or Iron-Oxidizing Bacteria in Reverse Microbial Fuel Cells.” He has received NSF awards from four programs: the Cellular and Biochemical Engineering, Process Separations, and Energy for Sustainability programs in the Division of Chemical, Bioengineering, Environmental and Transport Systems, and the Biomaterials program in the Division of Material Research. As the range of programs from which Dr. Banta has received funding exemplifies, researchers apply for funding from different programs depending on the goals of a specific research topic. This motivates the question of whether researchers move their work in the direction of opportunities offered by NSF, or whether they instead are able to find programs that fit their research portfolio. In contrast, Gregory Stephanopoulos, an MIT researcher also funded by the Electrofuels program, has received no NSF funding since 2008. He has received multiple NIH grants related to the same general topic (microbe engineering) as his ARPA-E award, suggesting he may be a researcher brought into the alternative energy field through ARPA-E.

While supporting transformative research is a goal of all NSF programs, specific NSF initiatives pay particular attention to novel and interdisciplinary research. First, NSF offers various cross-cutting solicitations in which multiple programs participate. Relevant for this proposal is 2011’s “Sustainable Energy Pathways” program, which involved five different NSF directorates and required interdisciplinary teams of researchers. It is an open solicitation, explicitly mentioning interest in a “broad scope of sustainable energy interest areas.” The solicitation does not provide specific technological goals. Second, all NSF programs consider Early-Concept Grants for Exploratory Research (EAGER) proposals to promote exploratory work on untested, but potentially high risk-high payoff research that “involves radically different approaches, applies new expertise, or engages novel disciplinary or interdisciplinary perspectives” (NSF 2018, p. II-33). We hypothesize that these programs, which emphasize interdisciplinary and/or high-risk, untested ideas, should promote greater changes in research direction than other open-ended programs.

**Methods Appendix 2: Mapping data for public availability**

The primary data sources for this project are the Scopus and Web of Science databases. Both databases are available to the researchers through Elsevier’s API and Claritive Analytics' API (also known as Web of Science). We also require data on (1) scientists funded by our targeted programs and (2) the proposal solicitations of these programs. Both will be complied from publicly available sources. The Federal Reporter includes information on grants from multiple government agencies, including NSF. For NSF, these data begin in 2008. However, the Federal Reporter does not provide information on DOE awards. Office of Science awards and FOAs are available from their website. While available through the 2000s, project abstracts are only available in their online database beginning in 2014. However, recent published reports of project summaries allow us to use projects funded by 2011. Following National Academies of Sciences (2017), we use the government website USAspending.gov to obtain award numbers of grants from ARPA-E and EERE. We can use the award numbers to identify publications resulting from these awards, and thus the authors whose work was funded. We will complement these data with additional information from the Department of Energy website, which includes project descriptions and the principal investigator for each ARPA-E award.[[5]](#footnote-5)

To compare researcher records to the initial grant solicitations, we will obtain DOE FOAs from the DOE’s Funding Opportunity Exchange website. Note that ARPA-E programs begin in 2009, and the availability of EERE FOA’s on the Funding Opportunity Exchange begin in 2011. When possible, we will complement these data with additional FOA data from the grants.gov website, which includes funding opportunities from 2009 onward. NSF program solicitations are available from the NSF website.

To enable researchers with access to Scopus and Web of Science to link the award data to the proprietary publication data, we will share the Scopus Author ID, the Web of Science's ResearcherID, or ORCID ID, or all of them when possible. These identifiers would allow access to the entire profile of authors used in our analysis. Additionally, we will share an in-house entity disambiguation software to cross-link authors across these datasets. This in-house entity disambiguation software will produce a unique author identifier across our entire data analysis pipeline and 1-to-1 links to Scopus and Web of Science identifiers, if available.

**Methods Appendix 3: Measuring similarity and breadth**

We can analyze whether funding affects research patterns by detecting changes between sets of documents published before, during, and after a grant's funding period. These changes can be detected using similarity measures. For example, a research change occurs if research becomes more similar to the FOA after funding is received. The procedure we will describe is agnostic to the sets considered, and therefore enables both comparisons within a scientist or even across a field. Change detection using similarities is therefore appropriate for our goals.

There are multiple ways to compute similarities. In this project, we will use the text of the grant's Funding Opportunity Announcement (or equivalent) and publications as the primary data to estimate similarities. Some researchers have used other signals to estimate similarities such as the citations between publications (e.g., Wesley-Smith & West, 2016; West, Wesley-Smith, & Bergstrom, 2016), references (e.g., Kessler, 1963) and user-provided keywords (e.g., Hong, Jeon, & Jeon, 2013). Citations, however, take significant time to accrue (Wang, Song, & Barabási, 2013), references may require full-text versions of articles, and keyword similarities rely on authors' personal opinions, producing poor agreement with expert classification (Achakulvisut, Acuna, Ruangrong, & Kording, 2016). Therefore, we use Natural Language Processing (NLP) techniques to find *semantic* similarity between sets of documents based on their raw text. This is a faster and more accurate way to measure similarities (Achakulvisut et al., 2016; Manning, Raghavan, & Schütze, 2008).

Measuring similarities using NLP involves two steps: (1) transforming the sets of documents into a vector representation, and (2) using an appropriate distance function for estimating their similarities. Ideally, the vector representation should aim to capture the semantics of the documents using a low dimensional space. Some representations such as vectors holding the counts of unigrams or *n*-grams are very high dimensional, causing several difficulties (Aggarwal, Hinneburg, & Keim, 2001). For this reason, dimensionality reduction techniques are performed first (Manning et al., 2008). As for the choice of the distance function, the cosine similarity is preferred because it allows comparison of documents of different lengths. Using dimensionality reduction and cosine similarity, documents can be compared effectively.

The dimensionality reduction used in this project will be as follows. First, a matrix *X* containing the term frequency–inverse document frequency (tf-idf) of all documents is created. This is, for the document *i* and term *j*, the matrix will be constructed as

where tf*ij* is the frequency of term *j* in document *i*, *n* is the number of documents, and df*j* is the number of documents with term *j*. Assuming that there are *m* terms, the matrix *X* will be an *n* by *m* highly-sparse matrix, where *m* typically ranges in the hundreds of thousands.

After this step, a non-negative matrix factorization (NMF) (Murphy 2012) is performed on matrix *X*. NMF tends to capture a more semantically meaningful representation of a document compared to using term frequencies or keywords. In order to drastically reduce the dimensions, the matrix *X* is represented by a set of *topics* *T* which are in turn mapped back into the full *vocabulary* through *V*. The matrix *V* therefore captures correlations between terms. This cannot be achieved by methods that rely on simple term counting, such as the representation used by the PubMed related articles features (*pmra*). We perform the factorization as follows

where , making *T* a "tall-and-skinny" matrix, and *V* a "wide-and-short" matrix. The matrices *T* and *V* (*T ≥* 0, *V ≥* 0) are estimated by solving the following optimization problem: min*T,V* ||*X* – *TV*||2. By representing the documents using *T* instead of the full *X*, we reduce the dimensionality of the documents from *m* to *k* dimensions. The dimension *k* is typically found by cross validation (Manning et al., 2008).

The representation provided by NMF is closely related to Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA). It makes a more correct assumption (i.e., non-negativity of *X*) compared to LSA and it is much faster and accurate to compute than LDA (Lancichinetti et al., 2015). Because the matrix *T* can be casted as a probabilistic representation of all topics present on a document, it can represent the topics of a corpus by simply averaging the topics of each document in the corpus as follows:

where *L* is the number of documents in the corpus and *tz* is the vector representation of the z-th document in the set. Because the matrix *T* can be casted as a probabilistic representation of all topics present on a document, it can also represent the topics of a corpus by simply averaging the topics of each document.

Finally, we use the cosine similarity between two corpora *c*1 and *c*2:

Our similarity measure will vary from zero (the two sets are completely different) to one (the two sets are exactly the same).

In addition to these measures of similarity, we also require measures of the breadth of a researcher’s portfolio for use in research question 3. Among two researchers with comparable similarity measures to a given paper topic, a researcher with additional experience in other areas may be able to combine the knowledge from each field in novel ways. Our breath measure allows us to ask whether researchers working on a wider range of topics contribute more valuable knowledge than other researchers. We first construct a topic distribution of the body of a researcher’s work (e.g., set of publications or grants, or both) to construct a topic distribution of it. We can use the normalized vector representation of a set of documents (e.g., using Eq. 3) to estimate the breath of the topics represented in such vector. More specifically, each element of the vector representation *s* is positive and therefore by normalizing it, we transform the topic assignment into a probability distribution. We then use the entropy of such probability distribution to measure the research breadth. This entropy quantifies how far away the body of work is from a very topically-concentrated distribution, an intuitive proxy for research breadth.

**Methods Appendix 4: Identifying the control groups**

To establish causal links between funding and the direction of research we must control for other external factors affecting the choices of both funded and non-funded scientists. Our goal is to identify a control group of authors who are similar to funded scientists in research interest, career stage, and other features. Our methods will identify such scientists *up to the time in which our treated scientists start receiving funding*. We can then ask whether our treated scientists were more likely to publish papers on clean energy than other scientists with similar profiles.

Our method therefore will construct a synthetic experimental dataset to get at the causality of why scientists shift their interests. Because of the large size of publication and author datasets, we will follow three steps to efficiently identify potential control scientists. Our first step is to construct search queries that capture a relatively small pool of “potential” control scientists. Since some potential controls may be unlikely matches for our funded researchers, our second step is to use machine learning to estimate the likelihood of a match given a potential control–funded scientist pair. Our third step is to use linear programming to find the top *k* controls for each of our funded researchers while maximizing the global likelihood of the matches. These steps will keep our dataset construction computationally tractable while ensuring good matching quality.

*Step 1) Constructing a dataset of "potential" controls:* We will use the Scopus publication database, which provides unique identifiers for each author. Azoulay *et al*. (2017) do comparisons that conclude that “the Scopus system provides an accurate set of career publications.” In our first step, we will use three filters provided by the Scopus API to narrow the set of potential controls. The first filter will narrow by Scopus' broad search subject categories. These categories are developed based on the journals in which a funded researcher publishes. While individual categories are broad, using combination of categories provides a first step to limiting the pool of researchers to those working on related topics.[[6]](#footnote-6) For example, Dr. David Popp, is an environmental economist. The first three Subject Areas in his Scopus profile are “Economics, Econometrics and Finance,” “Environmental Science,” and “Energy”. Thus, while the first Subject Area would yield a broad set of nearly every economist, adding the second two subject areas limits the potential set of controls to other economists working on energy and environmental issues. The second filter will narrow by terms that are particularly unique to the funded researcher compared to other researchers. We will extract such unique terms by selecting the top 95% of the tf-idf terms in matrix *X* of equation 1. The third filter will narrow the search to only a limited number of years before the grant start date. Thus, Scopus provides an effective information retrieval interface to identify a pool of potential control authors.

*Step 2) Estimating pair-wise match likelihood using machine learning:* We will use standard machine learning techniques to learn a function that predicts the likelihood of a potential control being a match for a target funded scientist. In particular, we will use Random Forest (RF) to learn such a function based on a broad set of features. RF has shown a remarkably consistent ability to generalize and deal with non-linear relationships between features (Caruana, Karampatziakis, & Yessenalina, 2008). The features we will include are the similarities between overall research topics, research specific to the FOA, number of publications, number of citations, number of co-authors, topic breadth, years of experience,[[7]](#footnote-7) and the quality of an author’s research institution.[[8]](#footnote-8) We will create the training data by evaluating suggested matches manually. At first, we will use nearest neighbor matching to create an initial list of the top matches based on the features previously described. Then, we will iteratively fit a sequence of RFs using an active learning mechanism to further refine pairs (Dasgupta & Langford, 2009). The advantage of using machine learning for matching, rather than traditional matching methods such as propensity score matching and nearest neighbor matching, is that it allows us to evaluate the quality of the overlap for the specific features above. For example, we can ensure that our matched pairs both have similar research interests relevant for the grant, and be more lenient on both having similar years of experience. To evaluate the quality of the RF prediction and generalization across NSF and DOE programs, we will use standard machine learning cross validation. In particular, we will construct datasets for three significantly different programs. We will use one dataset for fitting the several RFs with different parameters, another dataset for estimating the generalization quality of each RF, and a final dataset to estimate the generalization performance of the best RF from the previous step. These three datasets are typically known as training, validation, and testing datasets in ML (Murphy, 2012). Therefore, machine learning can be used effectively to estimate the likelihood of matches.

*Step 3) Global optimal matching with linear programming*: With the function estimated in Step 2, we can construct a matrix with the probabilities of all potential controls and funded researchers being good matches. However, we need to find an optimized set of potential controls. Using too few controls (i.e., "greedy" matching) may provide treatment estimates with too much variance whereas using too many controls may provide estimates with too much bias (Kallus, 2016; Rosenbaum, 1989). We can find a balance by finding a set of the top *k* controls for each treatment while trying to maximize the total likelihood of the matches. While this may appear to be a computational expensive integer programming problem, it can be casted as linear programming after some variable modification and therefore made efficient (Taylor, 2008). The problem of global optimal matching can thus be solved relatively effectively using standard off-the-shelf software.

After these three steps, further checks will be made for the resulting match. These checks will include balance and overlap. Assuming proper statistical estimation of errors when the match is unbalanced, the most difficult problem is the lack of overlap (e.g., large discrepancies in the distribution of number of citations between controls and funded researchers). We will take standard fixes to this potential problem by restricting the maximum distance between covariates before Step 3—using methods described in (Gelman & Hill, 2007, Ch. 10). It could be the case that the problem is the training data (Step 2). In this case, we will evaluate other methods that need less intervention at the cost of matching quality. In particular, we will evaluate Coarsened Exact Matching (CEM) or Propensity Score Matching (PSM). However, our previous experience with optimal matching of research portfolios (e.g., Acuna, Achakulvisut, and Kording, 2015) gives the proposed three-step matching process a high chance of success.

**Methods Appendix 5: Estimation**

***Does Funding Affect the Direction of Research? Which Types of Programs Influence Direction More?***

Once we identify our control group, we propose two sets of regressions (described as R1 and R2 in the main text) to isolate the effect of program funding from other researcher-specific and external forces influencing research direction. Using the measures of similarity described in Methods Appendix 3, our dependent variable for the analysis for research questions 1 and 2 follow is the similarity between the funded research and a researcher’s previous body of work. By observing researchers both before and after funding, we can assess whether funding changes the direction of research. As described in the main text of the proposal, for each regression, we measure similarity between previous research and funded research in two ways. In option 1, our dependent variable is a count of publications in each year that are sufficiently “close” to either the FOA or the proposed research abstract, obtained using machine learning to identify topics related to the funded research.[[9]](#footnote-9) Let *i* represent an individual scientist, *p* represent a research program, and *t* represent time. Our primary dependent variable is a count of publications sufficiently “close” to either the goals of the funding agency or of the proposed research.[[10]](#footnote-10) To obtain this count, we use the machine learning techniques described above to identify topics related to our target funding programs. This will vary for each program, *p*. To provide an initial guideline as to how “similar” research must be to be counted as related, we will identify publications acknowledging the funding programs we study and compare the similarity of these publications to the funding announcements themselves. We will use these similarity scores as guidelines for our cutoffs. For both scientists receiving funding and those in our control group, we can thus observe their work on related research both before and after the year of funding. Thus, we can ask both whether the funded scientist produces more published work on the funded topic *and* whether any changes observed are greater for funded scientists than other similar researchers. Since, by definition, funded scientists will have some publications directly related to the grant, we can also run regressions omitting publications that acknowledge the grant. This enables us to ask whether research on the funded topic continues after the initial round of funding.

While this method uses counts of publications that are straightforward to interpret, it requires an arbitrary decision of how similar related work must be to be counted. Thus, as a robustness check, option 2 simply uses the continuous measure of similarity between a researcher’s publications in year *t* and the funding announcement (or funded proposal), *p*. This provides a continuous measure, and thus obviates the need to arbitrarily decide how narrowly we define a topic. If other researchers are moving in similar directions (e.g. it is not funding, but other external forces changing the direction of research), we will observe little correlation between funding and similarity.

The advantage of this method is that counts of publications are straightforward to interpret. However, it requires an arbitrary decision of how similar related work must be to be counted. Thus, as a robustness check, we will simply use the measure of similarity between a researcher’s publications in year *t* and the funding announcement (or funded proposal), *p*, as defined above. This provides a continuous measure, so obviates the need to arbitrarily decide how narrowly we define a topic. If other researchers are moving in similar directions (e.g. it is not funding, but other external forces changing the direction of research), we will observe little correlation between funding and similarity.

With these measures in hand, our simplest specification uses a differences-in-differences research design, using researchers receiving funding from program *p* as the treated group. Let *yi,p,t* represent our alternative dependent variables described above. *FYp,t* represent a dummy variable equal to 1 in the year in which program *p* awards funding.[[11]](#footnote-11) Our treatment variable, *Fundedi,p,* is a dummy variable equal to 1 for scientists receiving funding from program *p*. Here “funded” means receiving a grant from a specific program. It does not mean receiving any research funding. It may be the case that non-treated scientists obtain other funding sources.[[12]](#footnote-12) Thus, *FYp,t* controls for general trends affecting all scientists, and the interaction of *FYp,t* and *Fundedi,p* captures the additional effects of funding from program *p* on funded scientists. For both *FYp,t* and the interaction, we allow for lagged effects multiple years post-funding. Using macro-level energy R&D and publication data, Popp (2016) finds that the effect of public funding on publication counts may persist for six years or more. *L* represents the maximum lag length.

Our first regressions includes only recipients of targeted DOE funded opportunities and their controls, using similarity between a researcher’s output and the FOA for the dependent variable. Here, we are asking whether targeted research programs move researchers closer to the goals of the FOA.

(6)

*p* and *t* represent program and year fixed effects, respectively. Recall that goals vary across programs. ARPA-E specifically cites attracting new researchers as a goal. In contrast, EERE, may look for grant recipients with research profiles close to program goals. As we only observe funding decision outcomes, we cannot separate these two effects. However, we will include robustness checks allowing for interactions with an ARPA-E dummy variable, to see if the results differ for ARPA-E, where attracting new researchers and encouraging high-risk/high reward research are explicit goals.

Our second set of regressions allows for separate effects for targeted and open calls for proposals and includes researchers (and their controls) funded by all the DOE and NSF programs. We include additional interactions with the dummy variables *Targeted*, which equals one for proposals funded by funding programs with specific research targets, and *RiskReward*, which equals one for programs specifically indicating preference for high-risk/high-reward research in Figure 1 of the main proposal.[[13]](#footnote-13) In these regressions, our dependent variable uses similarity measures comparing research output to the grant proposal, *p*, and thus asks whether funding programs using targeted FOAs and/or encouraging research with interdisciplinary and novel perspectives move researchers further away from their existing research than other programs.[[14]](#footnote-14)

(7)

Assuming our control group is well-matched, the above specifications cleanly identify the marginal effect of receiving funding from a specific government program. However, because it relies on multiple fixed effects, it does not allow us to observe how specific researcher characteristics affect research direction. Thus, as described in the main text we will also estimate a richer model that includes various researcher controls in place of **5*Fundedi,p*.

***Do researchers new to a field contribute novel ideas? Do they produce more highly cited research?***

Our final regressions ask whether research from researches new to the field creates more novel and more impactful research. The data for this analysis will use an updated version of publication data first created by one of the PIs for research on the effect of government energy R&D on publication counts. Popp (2016, 2017) use a series of keyword searches of article titles, abstracts, and keywords, to identify journal publications for each of three technologies: biofuels, solar energy, and wind energy. Energy storage publications were collected as part of that project, but do not appear in the final papers due to limited availability of control variables across the countries used in those papers. The article data come from the Clarivate Web of Science Core Collection database. The original data set runs from 1991-2011, as complete records of titles, abstracts, and keywords are first available in 1991.We will use the same search strategies to extend these data through 2017. We focus on publications in scientific journals by dropping articles such as reviews, editorials, or news items. We do include proceedings papers included in the Web of Science Core Collection database.

Using these data, we focus on two key outcome measures: (1) *noveltyp*, the novelty of publication *p* compared to the previous literature in the field, and (2) *impactp*, the impact of publication *p* measured by total citations received. Because our interest is in how new research builds upon previous work, we include self-citations in our analysis. The key explanatory variables are **similarityp**, a vector of the various measures of the similarity and research breadth of each paper’s authors as described in the main body of the proposal.[[15]](#footnote-15) Following the analysis used by one of the PI’s in earlier NSF-funded research on energy funding (Popp 2016, 2017), we propose two sets of regressions using the basic formulation below. As average novelty and citation rates should vary across fields, we will estimate separate models for each of our four technologies. Letting **Zp** represent a matrix of controls and *a* and *t* represent author and year fixed effects, we propose the following regressions:

*yp* = *f*(**similarityp**, **Zp**, *a,t*) (8)

In our first set of regressions, our dependent variable, *y*, is *noveltyp*. This regression asks whether papers that are less similar to the authors’ research portfolio make more novel contributions to the field as a whole. Our second set of regressions use *impactp* as the dependent variable, to assess whether papers less similar to the authors’ research portfolio have greater impact on future researchers in the field. Because the work of these researchers may receive more citations because it has higher impact on future work, or simply because the work is exposed to other researchers in a broader set of fields, we will also include a set of regressions where both **similarityp** and *noveltyp* are explanatory variables. By conditioning on each, we can see, for example, whether less similar researchers have greater impact even when controlling for article novelty. If that is the case, that suggests that these researchers not only contribute new ideas to the field, but also expose their research to a broader set of scientists.[[16]](#footnote-16)

In all regressions, we must control for other factors affecting novelty or impact. For example, the number of citations received depends not only on article impact, but also on the number of opportunities for citation. For energy publications, we would expect the raw number of citations to be higher during periods of intense research activity, such as when energy prices are highest. Similarly, novelty varies over time as creating novel research becomes more difficult in a more crowded field. Year fixed effects, *t* control for such variation over time. **Zp** represents a matrix of control variables related to article characteristics other than our measures of interest. These include the number of authors, as papers with larger research teams are cited by other articles more frequently (Wuchty *et al*. 2007) and are more novel (Lee *et al.* 2015).[[17]](#footnote-17) We include a dummy variable for papers with authors from multiple countries, which may be exposed to a broader set of potentially citing papers. Finally, some specifications will include journal impact factors. While this limits the comparison of impact and novelty to other similar journals, it will be important if different sub-fields publish in journals with very different impact factors. *a* represents author fixed effects. Models with author fixed effects control for unobserved differences across authors, such as author prestige, that affect novelty or the likelihood of citation.[[18]](#footnote-18) However, these models offer a narrower interpretation, asking whether researchers working on new topics generate more highly cited (or more novel) work than their other research. Thus, we will include models both with and without author fixed effects.

**Appendix G: Information Products**

To reach a broad audience, we will target both top scientific and policy journals when submitting manuscripts. Moreover, all working papers will be made available in the highly visible NBER Working Paper series. For the technical papers describing new matching strategies for causal inference, the manuscript will be made available in the pre-print service ArXiv.org We anticipate at least three publications from this project:

1. A paper using our treated and control groups to ask whether funding affects the direction of research. Given the interdisciplinary nature of this research, we plan to target a top interdisciplinary journal with this paper, such as *Science* or *Nature*.
2. A paper exploring whether researchers new to a topic generate more novel and more impactful research. A likely target for this manuscript is *Research Policy*. Alternative venues include *Nature Human Behavior*.
3. A paper comparing the results of machine learning to other commonly used matching techniques such as propensity score matching, so as to inform future research using matching techniques. A target journal for this paper is the *Journal of the American Statistical Association*. Alternative venues include *Nature Machine Intelligence*.

As noted in the main body of the text, while our publication data come from proprietary sources, we will make the code used to collect these data publicly available, so that other researchers with access to Scopus and the Web of Science can replicate our results.

1. To provide an initial guideline as to how “similar” research must be to be counted as related, we will identify publications acknowledging the funding programs we study and compare the similarity of these publications to the funding announcements themselves. We will use these similarity scores as guidelines for our cutoffs. [↑](#footnote-ref-1)
2. We will also consider models including a squared term for the number of authors, as Lee *et al*. (2015) find that the effect of team size on novelty diminishes as team size grows. [↑](#footnote-ref-2)
3. We will also include interactions of these variables in some specifications. [↑](#footnote-ref-3)
4. Note that publications are an important output of both ARPA-E and Office of Science research. Goldstein and Narayanamurti (2018) show that ARPA-E researchers are more likely to publish than researchers from either EERE or Office of Science. While EERE research is more applied, and often results in technical reports and patents, the EERE projects studied by Goldstein and Narayanamurti are more likely to have a publication than a patent, and we can compare the similarity of the proposed research to a scientist’s previous publications even if the EERE grant in question does not produce any publications itself. [↑](#footnote-ref-4)
5. Less information is available for EERE awards on their website. Complete lists of funded projects are available from the solar and wind energy programs but not for other programs. Thus, we are only able to obtain proposal abstracts for these two programs. [↑](#footnote-ref-5)
6. Alternatively, one could simply select as controls other researchers working on the same topics funded by the government programs in our study. However, if some of the funded researchers are new to the field, it is important that our controls also include people who are new to the field being funded. [↑](#footnote-ref-6)
7. Calculated based on the date of the author’s first publication in Scopus. [↑](#footnote-ref-7)
8. The resources institutions provide affect research productivity. While we don’t want to compare only researchers from the same university, we can select as controls researchers working in similar environments. We will experiment with multiple measures here, including the Carnegie research classification of a researcher’s university affiliation and the university’s ranking in world university rankings such as the Times Higher Education rankings. We will also include an indicator for researchers at non-university institutions. [↑](#footnote-ref-8)
9. To provide an initial guideline as to how “similar” research must be to be counted as related, we will identify publications acknowledging the funding programs we study and compare the similarity of these publications to the funding announcements themselves. We will use these similarity scores as guidelines for our cutoffs. [↑](#footnote-ref-9)
10. For clarity, we describe our procedure below based on our first measure of similarity, which compares a researcher’s portfolio to the FOA. As noted earlier, for comparing targeted and open funding requests, we will also compare the similarity of each researcher’s existing work to their own research proposal. While the comparison group is different, the methodology is otherwise the same. [↑](#footnote-ref-10)
11. For many of the targeted DOE programs, there is simply a single year of funding. For more general programs, *FYp,t* represents the year in which funding was given to researcher *i*. In both cases, *FYp,t* is turned on for both funded scientists and their matched controls. [↑](#footnote-ref-11)
12. In principle, one could control for the size of the grant. Myers (2019) does this in his working paper on the elasticity of science. However, as Myers’ pool of researchers is all scientists who apply for NIH funding, he knows the funding amounts (or funding applied for) for all scientists in his sample. We do not. [↑](#footnote-ref-12)
13. *Targetedp* or *RiskRewardp* do not appear separately, as they are encompassed by the program fixed effects. [↑](#footnote-ref-13)
14. Proposal fixed effects are the same for funded researchers and control scientists that match each funded scientist for proposal *p*. In our larger sample, we expect our dataset will include researchers who receive multiple awards from either NSF or DOE. Thus, we can also run regressions adding individual researcher fixed effects. Such a regression would compare the effects of targeted and open funding opportunities on the same researcher. [↑](#footnote-ref-14)
15. All similarity measures compare papers to what has been previously published. While the novelty measure of a single paper will not vary over time, the similarity and breadth of an author’s work to her previous work uses only the author’s work published prior to the paper in question. Thus, these measures are not constant for an author, but will differ for each paper considered. [↑](#footnote-ref-15)
16. We will also include interactions of these variables in some specifications. [↑](#footnote-ref-16)
17. We will also consider models including a squared term for the number of authors, as Lee *et al*. (2015) find that the effect of team size on novelty diminishes as team size grows. [↑](#footnote-ref-17)
18. Co-authored papers will have author fixed-effects for each individual author. [↑](#footnote-ref-18)