Pork Barrel or Barrel of Gold? Examining the performance implications of earmarking in public R&D grants

Kok, Holmer; Faems, Dries; de Faria, Pedro

Published in:
Research Policy

DOI:
10.1016/j.respol.2022.104514

Research Policy 51 (2022) 104514

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Pork Barrel or Barrel of Gold? Examining the performance implications of earmarking in public R&D grants

Holmer Kok [a,*], Dries Faems [b], Pedro de Faria [c]

[a] Stockholm School of Economics, House of Innovation, Saltmätargatan 13-17, Stockholm, SE-113 83, Sweden
[b] WHU - Otto Beisheim School of Management, Burgplatz 2, 56179 Vallendar, Germany
[c] University of Groningen, Nettelbosje 2, Groningen, 9747AE, the Netherlands

A R T I C L E I N F O

Keywords:
Competitive selection
Innovation policy, Public funding, Earmarks
Peer review, R&D grants

A B S T R A C T

Scholars tend to assume that publicly funded R&D projects, which are competitively selected, outperform projects, which receive funding through a political selection process. In this paper, we empirically explore this assumption, examining the outcomes of 321 R&D projects that were funded by the U.S. Department of Energy's Hydrogen Program. Between 2003 and 2011, projects in this program could not only receive funding by means of a competitive selection process, but also by being earmarked by a U.S. member of Congress. We find that, whereas earmarked projects receive considerably lower peer review evaluation scores than non-earmarked projects, they do not consistently underperform in terms of the productivity, spillovers, and novelty of research- and science-based outcomes. Post-hoc analyses provide indications that this misalignment is driven by the existence of a bias of peer reviewers toward earmarked projects. Jointly, our findings challenge the dominant assumption that competitively selected projects always outperform politically selected ones in the setting of public R&D grants. In this way, we provide academics and policy makers with a richer perspective on the advantages and liabilities of earmarks.

1. Introduction

Governments use different instruments to provide financial support to a wide variety of actors with the objective of mitigating R&D-related market inefficiencies (Aschhoff & Sofka 2009; Becker, 2015; Guerzoni & Raiteri, 2015). Next to providing R&D tax incentives (Bloom et al., 2002; Czarnitzki et al., 2011; Knoll et al., 2021) and initiating public technology procurement (Aschhoff & Sofka 2009; Czarnitzki et al., 2018; Raiteri, 2018), governments can make use of public R&D grants, where government agencies provide R&D funding to particular actors within particular programs (Goldstein & Narayanamurti, 2018; Howell, 2017; Jacob & Lefgren, 2011). Public R&D grants programs allow governments to choose sets of criteria, such as the creation of spillover effects (Feldman & Kelley, 2006; Jaffe 1998), and select particular projects to receive funding. In other words, public R&D grants allow governments to influence the R&D activities of actors in a specific field by selecting specific projects to reach policy objectives (Ebersberger, 2005).

The allocation of funding in public R&D grants therefore is a crucial dimension of public R&D funding. This allocation process can be organized in different ways (Heinze, 2008; Ioannidis, 2011; Wang et al., 2018). Governmental agencies can organize competitive selection processes, relying on peer review panels to evaluate submitted projects and subsequently select projects based on their rankings. However, selection processes can also have a more political nature, where politicians exercise discretion regarding which projects should receive public R&D funding. In general, scholars assume that competitive selection processes outperform their political counterparts in terms of identifying high-quality projects (Boyle & Matheson, 2009; Brach & Wachs, 2005). They argue that funding decisions, which are based on competitive selection processes, tend to be less biased and less likely to be influenced by factors that are unrelated to merit (de Figueiredo & Silverman 2006).

However, empirical evidence, supporting this core assumption, is lacking (de Figueiredo & Silverman, 2007; Frisch & Kelly, 2011). Addressing this empirical gap is important since indications are present that competitive selection processes might also face challenges. Specifically, some studies provide evidence that peer review panels can also be biased toward certain types of applicants or projects which can lead to suboptimal project selection (Bornmann, 2011; Fang & Casadevall, 2016; Lee et al., 2013). Moreover, some scholars point to potential advantages of political selection processes. For example, political selection

* Corresponding author.
E-mail address: holmer.kok@hhs.se (H. Kok).

https://doi.org/10.1016/j.respol.2022.104514
Received 28 February 2021; Received in revised form 6 February 2022; Accepted 3 March 2022
Available online 16 March 2022

can lead to the funding of more unconventional and impactful projects (Frisch & Kelly, 2011; Silber, 2002).

The core objective of this paper is to empirically test whether competitively selected projects produce better quality outcomes than politically selected projects in R&D grants. To do so, we examine the outcomes of projects from the U.S. Department of Energy's Hydrogen Program. An important characteristic of this program is that, before 2011, projects could receive funding in two ways: (1) by being selected through a competitive selection process or (2) by receiving funds that were earmarked by a U.S. member of Congress. An earmark is a provision placed into a discretionary spending bill, stipulating that a certain amount of public funds should be transferred to a specific recipient, thereby circumventing conventional competitive selection processes (Boyle & Matheson, 2009). Earmarks, therefore, represent a political selection approach that is used as an alternative to conventional competitive selection. Given the presence of both competitive and political selection processes in the Hydrogen Program, we leveraged this empirical setting and collected detailed information on 321 projects within this program between 2003 and 2011. For our main analyses, we collected two types of data. First, we obtained the evaluation scores provided to ongoing projects by peer reviewers. Second, we collected data on the productivity, spillovers, and novelty of projects' scientific and research outputs (i.e. patents and publications). In this way, we were able to systematically compare the performance outcomes of competitively selected projects and projects that were politically selected (i.e. by means of earmarks).

Our analyses show that, whereas peer reviewers consistently give lower evaluation scores to earmarked projects than to non-earmarked projects, this underperformance is not consistently present when comparing the productivity, novelty and spillovers of earmarked and non-earmarked projects. For scientific productivity and research novelty, we even find some evidence that earmarked projects can outperform non-earmarked projects. To get a better understanding of this misalignment between how earmarked projects are evaluated by reviewers and how they perform in terms of productivity, novelty and spillovers, we conducted several post-hoc analyses. These analyses suggest the existence of a bias of peer reviewers toward earmarked projects that might influence their evaluation scores.

Our findings have important implications for research on public R&D grants. First, our results challenge the dominant assumption that competitively selected projects always outperform politically selected ones. Second, whereas prior research has already generated evidence for the existence of peer review bias during the initial selection of projects for funding (Bornmann, 2011; Lee et al., 2013), we provide indications for the presence of peer review bias when reviewers evaluate ongoing projects that have already been selected for funding. Third, our findings contribute to the ongoing policy debate on the advantages and disadvantages of earmarking of public R&D funding (Brunner, 2021; Dickerson, 2021; The New York Times, 2020). Our findings indicate that policy makers cannot assume that non-earmarked projects always outperform earmarked ones. This means that earmarking critics will have to rely on a different and more nuanced set of arguments to justify banning this type of political selection approach.

2. Allocation of funding in public R&D grants: competitive versus political processes

2.1. Public R&D grants

Governments rely on different funding instruments, such as R&D tax incentives, public technology procurement, and R&D grants to stimulate organizations to invest in R&D activities that contribute to certain policy objectives and help society (Aschhoff & Sofka 2009; Becker, 2015; Bloom et al., 2019). Empirical evidence shows that the overall effect of R&D tax incentives on R&D investments and innovation outputs appears to be positive (Bloom et al., 2002, 2019; Czarnitzki et al., 2011), but also suggests that R&D tax incentives just lead to a reallocation of R&D expenditures from one region to another (Knoll et al., 2021). Public technology procurement allows for more targeted policy implementation. In this case, governments engage in contracts with private sector partners to purchase novel technologies that need to meet pre-defined functional characteristics (Aschoff & Sofka, 2009; Guerzoni & Raiteri, 2015). Studies find that public technology procurement can stimulate demand in targeted sectors (e.g Ghisetti, 2017) and can positively influence innovation outputs when certain conditions are met (e.g Aschoff & Sofka, 2009; Czarnitzki et al., 2018; Raiteri, 2018).

Another funding instrument, which allows governments to steer R&D activities in specific directions (Ebersberger, 2005), is R&D grants. Most R&D grants are coordinated by specialized governmental agencies and are part of large programs that focus on specific sectors or technologies. For example, within the U.S. Department of Energy, there are specific programs in the domains of wind energy, hydrogen and fuel cell, and bioenergy technologies. Within each of these programs, organizations (e.g. firms, universities, and research institutes) can receive funding for projects that are aimed at contributing to the improvement of the performance and efficiency of particular technologies in these specific domains. To realize the policy objectives of a particular R&D grant program, it is essential that funding is allocated to those projects that have the highest likelihood to generate novel solutions to specific technological challenges (Goldstein & Narayanamurti, 2018; Grodal & O'Mahony, 2017) and/or have the highest estimated ability to cause significant knowledge spillovers (Feldman & Kelley, 2006; Jaffe, 1998). Therefore, the selection of projects is a crucial decision-making dimension within the context public R&D grants. In this study, we highlight that selection processes in public R&D grants can be competitive or political. Below, we systematically describe these two different selection processes, explaining their potential advantages and disadvantages.

2.2. Competitive selection for public R&D grants

In a competitive selection process, governmental agencies attract organizations to participate in a particular research program, usually by releasing calls for proposals on specific topics. Organizations that want to receive funding from such a program draft a detailed proposal, describing the intended activities and expected outcomes. Peer reviewers, who are experts in the field, subsequently evaluate and rank these proposals on different criteria, such as their feasibility and ability to contribute to technological and scientific progress. Finally, program managers rely on these evaluations and rankings to allocate funding to specific projects.

The competitive selection process has become the default mechanism for allocating public funds through R&D grants (Fang & Casadevall, 2016; Ioannidis, 2011). The core advantage of this process is that peer review panels have the knowledge and expertise to filter out poor research and identify high quality research (Ginther & Heggeness, 2020; Li & Agha, 2015; Reinhart, 2009). Moreover, competitive selection processes rely on standardized and transparent procedures, which reduce the likelihood of favoritism or nepotism (Bornmann et al., 2008; Marsh et al., 2008). Competitive selection processes may also increase research productivity in more indirect ways by fostering collaborative networks among co-applicants (Ayoubi et al., 2019) and by providing detailed feedback to applicants regarding the content of their proposals (Brach & Wachs, 2005).

Although scholars mainly highlight the advantages of competitive selection of projects in the context of public R&D grants, they also acknowledge potential challenges and limitations (Braun 1998; Fang & Casadevall, 2016; Ioannidis, 2011). In particular, studies point to the risk that peer review panels can be biased toward certain funding applicants and projects, resulting in suboptimal funding allocation decisions by government agencies (Bornmann, 2011; Fang & Casadevall, 2016; Lee et al., 2013). Extant research points to four major sources of bias when expert panels evaluate proposals. First, peer review panels tend to provide higher evaluation scores to applicants that are connected to them (Marsh et al., 2007; Sandström & Hällsten, 2008). Second, studies report that reviewers can be biased toward applicants of a particular gender (Bornmann et al., 2007; Sandström & Hällsten, 2008; Wennerås & Wold, 1997) and race/ethnicity (e.g Ginther et al., 2011; Hayden, 2015). Third, studies report that peer reviewers can be biased toward already-successful applicants that have a history of receiving grant funding (Bol et al., 2018; Langfeldt, 2001). Fourth, peer reviewers can be positively biased toward conservative research proposals (Grodal & O'Mahony, 2017; Heinze, 2008; Nicholson & Ioannidis, 2012) that are intellectually proximate to them (Boudreau et al., 2016; Braun, 1998; Li, 2017).

The issue of peer review bias is exacerbated by the fact that peer reviewers are often explicitly instructed to take applicants' characteristics into account in their evaluations (Bol et al., 2018; Langfeldt, 2001; Stephan et al., 2017). Moreover, peer review processes for public R&D grants are frequently single-blind (i.e. reviewees are known but reviewers are anonymous) rather than double-blind (i.e. reviewers and reviewees are both anonymous), further amplifying potential biases (Lee et al., 2013).

2.3. Political selection for public R&D grants

Competitive selection processes are not the only possible approach to allocate public funds to organizations through R&D grants (Heinze, 2008; Ioannidis, 2011; Wang et al., 2018). Such decisions can also be based on a political selection process, where politicians exercise discretion regarding the allocation of public R&D funding to certain projects (de Figueiredo & Silverman, 2007). For instance, up until 2011, members of U.S. Congress could significantly influence the allocation of public R&D funds by means of so-called earmarks. In particular, they could include certain lines in a discretionary spending bill, specifying that a certain amount of funds should be allocated to a specific recipient, thereby circumventing conventional competitive selection processes.

The earmark process is typically initiated by an organization that is seeking external funding to support ongoing or novel R&D activities (Frisch & Kelly, 2011; Lazarus, 2010). The organization drafts a proposal for a project and then submits it to the office of their respective member of Congress. Organizations also frequently spend resources on dedicated third-party services, such as lobbying firms, to manage this process (de Figueiredo & Silverman, 2006). It is important to emphasize that members of Congress and their team do not automatically accept all earmark requests (Frisch & Kelly, 2011). Instead, they actively decide which projects they want to further pursue for inclusion in a discretionary spending bill. Selection processes for earmarks are not very transparent, however, which has resulted in limited insights into the exact factors that influence the decision-making process of members of Congress when they select the projects to be earmarked. Nevertheless, studies provide some indications regarding what those factors might be (e.g Frisch & Kelly, 2011; Lazarus, 2010), such as the extent to which the project can advance the member of Congress' own political objectives and, relatedly, the local economic benefits that members of Congress' constituents will obtain from the project. Once earmark requests make it into a discretionary spending bill, governmental agencies are expected to fulfill these requests and provide funds to the recipients (Law et al., 2008).

Several scholars have argued that politically selected projects are likely to be of inferior quality when compared to competitively selected projects (e.g Doyle, 2011; Lawler, 2000; Payne, 2002). Two core arguments are used to support this assumption. First, scholars argue that, in comparison with field experts, who are responsible for the evaluation of projects in competitive selection processes, politicians are worse in identifying projects that have the potential to produce high-quality outcomes because they lack the necessary scientific and research background (Finnigan, 2007; Payne, 2002; Sciara, 2012). Second, politically selected projects are often chosen based on reasons that are largely unrelated to their scientific or technological merits. For example, studies report that organizations with more extensive lobbying activities are more likely to receive earmarked funds (de Figueiredo & Silverman, 2006). Moreover, strong indications are present that politicians might give earmarks to please their constituents instead of nurturing technological progress. Members of Congress often request earmarks that provide funds to projects that are highly visible and appreciated by their constituents. For example, the first earmarks at the USDA were related to cotton – a commodity which was highly politicized and scrutinized (Law et al., 2008). In this way, members of Congress were seen as "bringing home the bacon", positively swaying public opinion. In support of this notion, studies find that earmarks were often requested by and provided to members of Congress that were electorally vulnerable (e.g Engstrom & Vanberg, 2010).

Although scholars have mainly highlighted the disadvantages of political selection processes, some also point to potential advantages of this alternative selection approach (Frisch & Kelly, 2011; Kunz & O'Leary, 2012; Silber, 2002). First, in comparison with competitive selection processes, political discretion makes it possible to fund projects that deviate from the norm. Government agencies typically formulate funding programs on rather narrowly defined topics (Azoulay et al., 2011; Grodal & O'Mahony, 2017) and peer review panels tend to prefer more conservative and familiar applicants and proposals (Grodal & O'Mahony, 2017; Heinze, 2008; Nicholson & Ioannidis, 2012). In political selection processes, however, politicians have more freedom in selecting projects. In the context of earmarks, for instance, members of Congress did not have to strongly consider the strict funding criteria which were defined by the responsible government agencies. Moreover, researchers in competitively selected projects face the risk of reduced funding or even discontinuation when they do not execute the project according to predefined plans (Goldstein & Kearney, 2020). In contrast, since program managers have considerably less influence on funding and continuation decisions for earmarked project, such projects can diverge more from the initially planned R&D activities when unexpected challenges or opportunities emerge (Azoulay et al., 2011; Manso, 2011). In other words, political selection of projects might allow directing funding to unconventional projects and permit more flexibility in their implementation. In explorative settings such as R&D projects, this might help to generate high-quality outcomes (Iansiti, 1995). Second, competitive selection processes tend to favor already-successful applicants that have strong research capabilities (Bol et al., 2018; Langfeldt, 2001). When one organization already has developed research capabilities through, for example, a prior grant, this increases the chance that it will receive a subsequent grant (Bol et al., 2018; Feldman & Kelley, 2006). Such path-dependent trajectories contribute to concentration of funding in a few organizations over time (Ma et al., 2015). Political selection of funds, in contrast, can result in allocating public funding to organizations that do not (yet) have a strong research track record, or historically have not been successful in competitive funding processes. Since the marginal effects of providing grants to funding-deficient applicants are higher (Ganguli, 2017; Hyytinen & Toivanen, 2005), political selection processes can help to "unlock" the research potential of organizations, leading to high-quality outputs.

In sum, although most scholars assume that competitive selection of projects for public R&D funding is superior to political selection, both approaches seem to have advantages as well as challenges. Therefore, we need empirical research to test whether competitively selected projects indeed produce higher quality than politically selected projects. In the following section, we explain our empirical approach to address this empirical question.

3. Methods

3.1. Empirical setting

We examine Research, Development, and Deployment projects that were funded by the U.S. Department of Energy (DOE) within the Hydrogen Program. The Hydrogen Program aims to fund projects that can address major issues such as fuel cell system cost efficiency, hydrogen production cost efficiency, and hydrogen storage tank durability[1]. The Hydrogen Program is an ideal setting for comparing the competitive and political selection of projects in public R&D grants for two main reasons. First, next to having a competitive selection of projects, earmarks were frequently used to fund projects within the program. For FY2004, for instance, it is reported in the DOE's Congressional Budget Requests that approximately 50% of appropriated funds for the Hydrogen Program were earmarked (for an overview: Sissine, 2006). Second, we were able to identify and track peer review evaluation scores for ongoing projects in a standardized way within this program. These scores allowed us to observe how experts evaluated the quality of ongoing earmarked and non-earmarked projects.

3.2. Sampling strategy

Our sample includes all projects that, as they were ongoing, received at least one review within the scope of the Hydrogen Program. We identify these projects by examining the Hydrogen Program's Annual Merit Review and Peer Evaluation reports. We apply two exclusion criteria to arrive at our final sample of projects. First, we excluded projects with a missing Federal Award Identification Number (FAIN). These unique identifiers allow tracking the patent and publication output of projects as well as obtaining relevant information about their duration and funding specifications. The majority of FAINs were identified in the Hydrogen Program's Annual Progress Reports. Missing FAINs were found in the USA Spending database and the 'Government Spending' section of the National Archives and Records Administration (NARA) by searching for keywords in the project title and/or the name of the recipient. This procedure excluded projects led by government laboratories, which are funded via large organization-wide grants instead of individual grants.

Second, to ensure that our focal sample was comparable and representative, we only included projects with a starting date between FY2003 and FY2011 (i.e. the 1st of October 2002 and the 30th of September 2011). The lower bound is determined by the availability of peer review evaluation scores. The upper bound is determined by the moratorium on earmarks that was decided in 2011. In our data, we indeed find that earmarks were no longer provided after 2011, with the latest starting date of an earmarked project in our sample being the 1st of June 2011. Projects with an ending date after the FY2011 are included[2].

3.3. Dependent variables

We collect data on multiple dependent variables to systematically compare the performance of earmarked and non-earmarked projects along dimensions that are relevant in the context of R&D grants. First, we create a dependent variable that represents a broad assessment of the project's performance, as evaluated by expert peer reviewers. This outcome variable is informative as it captures experts' opinion on the added value of the ongoing project. Governments want to fund projects that can generate multiple solutions to extant problems. Moreover, in the context of public R&D grants, they typically aim to fund projects that generate novel solutions with high spillover potential. To capture the tangible outcomes of projects, we therefore wanted to capture three different performance dimensions: productivity, spillovers, and novelty. For each of these three dimensions, we generate a research-based measure (i.e. using patent data) and a science-based measure (i.e. using publication data).

The dependent variables are examined at different levels of analysis. We measure project evaluation scores and the productivity of projects at the project-level, while spillovers and novelty are measured at the patent- and publication-level. We have three reasons to follow this approach. First, by measuring spillovers and novelty at a different level we avoid confounding the effect of earmarks on spillovers and novelty with the effect of earmarks on productivity (i.e. the sheer quantity of patents and publications resulting from projects). Second, by conducting the analyses at the patent- and publication-level, we can include a more dedicated set of control variables that are important to assess spillovers and novelty. Finally, it would be difficult to correctly interpret the values of our dependent variables for spillovers and novelty at the project-level. For example, a score of 0 on spillovers or novelty would not necessarily imply that no patents or publications were generated from the project, it could also be that the project generated patents and publications but that they were not cited or they were not considered novel.

3.3.1. Project evaluation scores

We use peer review scores to observe how experts evaluate the quality of ongoing projects. We emphasize that we only consider peer review scores provided to projects that are already ongoing, and not pre-funding peer review scores that are used to select which projects to include in a program. Pre-funding peer review scores can be used to better understand why certain projects, going through a competitive process, are selected in the first place, while post-funding scores can be used to understand the actual, rather than anticipated, outcomes of funded projects (Goldstein & Kearney, 2020). Since we are interested in the performance of projects, the latter type of data is more suitable in our context. We collected the peer review scores from the Hydrogen Program Annual Merit Review and Peer Evaluation reports. The EERE Peer Review Guide, which the Hydrogen Program uses as a guideline for its peer review activities[3], explicitly states that: "Earmarks will be included in the review and treated on the same basis as other activities (Office of Energy Efficiency and Renewable Energy, 2004, p. 10)." Moreover, it states that: "The EERE minimum requirement is that all programs and key projects be assessed, on average, every two years. In general, all projects in a given topical portfolio will be considered for review, regardless of their stage of maturity, with the primary focus on the key projects, typically comprising 80-90% of the program budget, and earmarks (Office of Energy Efficiency and Renewable Energy, 2004, p. 14)." This means that the earmarked projects in our sample were reviewed in ways that are similar to non-earmarked projects. Moreover, it implies that the majority of projects are reviewed at least once.

Every year, during the Annual Merit Review (AMR) meeting of the Hydrogen Program, project participants present their recent activities and results, allowing peer reviewers to evaluate the progress and outcomes of each ongoing project. According to the EERE guidelines, peer reviewers are to be provided with preparatory materials ahead of time, such as copies of the project summaries (Office of Energy Efficiency and

[1] Within the DOE, the Office of Energy Efficiency and Renewable Energy (EERE) manages the Hydrogen Program and funds the majority of projects. Other projects are funded by the DOE's Office of Fossil Energy, Office of Science, and Office of Nuclear Energy. Hydrogen and fuel cell technologies represent an important part of the U.S. strategy to advance clean energy technologies and increase energy independence from oil. Yet, hydrogen and fuel cell technologies are still in the early stage of deployment and commercialization, facing numerous technical challenges (Sharaf & Orhan, 2014).

[2] Out of the 321 projects, a sizeable number (i.e. 131 projects) had an ending date after FY2011. Moreover, we noticed that several projects selected via earmarks also had an ending date after FY2011. Therefore, we had no reason to exclude projects with an ending date after FY2011.

[3] The evaluation reports explicitly state that the EERE peer review guidelines were followed between FY2004 and FY2017. However, we could not find explicit evidence that this was the case for FY2003. In robustness checks that are available upon request, we exclude the scores from peer reviews provided during FY2003, and the results remain stable.

Renewable Energy, 2004). Moreover, these guidelines stipulate that the quality, objectivity, and impartiality of reviewers can be further ensured through only selecting reviewers with relevant experience in the field (based on, e.g. publication record, relevant degrees), taking steps to increase the anonymity of peer reviewers, and not assigning reviewers to projects for which they might have a conflict of interest. The peer reviewers are asked to rate a project on five aspects related to: (1) relevance to overall DOE objectives, (2) approach to performing research, development, and deployment, (3) technical accomplishments and progress toward project and DOE goals, (4) technology transfer/collaborations with industry, universities, and other laboratories, (5) approach to and relevance of proposed future research. Each aspect was rated on a four-point scale, with a score of one being the lowest and four the highest. An overall score is derived by taking a weighted average of these five aspects. Since our data are at the project-level, we took the average of the overall scores that the projects received across different evaluations[4,5].

3.3.2. Patents and scientific publications

In line with prior research (e.g Gittelman & Kogut, 2003; Goldstein & Narayanamurti 2018; Wang et al., 2018), we measure research outputs using patents and scientific outputs using publications. Patents and publications are seen as highly relevant outputs in the context of the Hydrogen Program. The DOE extensively disseminates information about the patents and publications resulting from funded projects. For example, the DOE explicitly mentions patent and publication outputs on the website of the Office of Scientific and Technical Information (OSTI). Moreover, in the Hydrogen Program's annual progress reports, project participants are explicitly asked to describe the patents and publications resulting from their projects. In our sampling approach, we consistently exclude patents granted and articles published after the 31st of December 2017. We use this date as the cut-off point because we initiated our data collection efforts in late 2018, meaning that not all articles published in 2018 were identified.

Patents that result from government funding need to acknowledge such funding in the patent text (Corredoira et al. 2018, Fleming et al. 2019). For example, patent US7829652 states that "This invention was made with Government support under contract number DE-FG36-06GO16034 awarded by the Department of Energy. The Government has certain rights in the invention." The FAIN DE-FG36-06GO16034 is linked to a project that was led by General Electric in the Hydrogen Program. We collect data on the listing of government funding acknowledgment on patents from two public databases: PatentsView, a database that provides access to various information regarding granted USPTO patents from 1976 and onwards, and DOEpatents, a database that keeps track of granted USPTO patents resulting from research funded by the DOE.

To measure research-based productivity, we consider the quantity of patents, which is measured at the project-level and is calculated as the total number of granted USPTO patents acknowledging funding from a specific project. To measure research-based spillovers, we follow an established line of research (Jaffe & De Rassenfosse, 2017) and count the number of citations a patent receives from other granted USPTO patents within the first five years after (and including) the year in which it was filed. The latter data are mainly retrieved through PatentsView. To measure research-based novelty, we use two alternative variables based on Arts et al. (2021). The measures created by Arts et al. (2021) are based on text in the patent title, abstract, and claims. The first measure calculates the overlap in individual words between the patent text of the focal patent and other USPTO patents filed in the past five years. This measure is continuous and bounded with a minimum value of 0 and a maximum value of 1. We recoded this variable such that high values on it indicate low overlap with prior patents' text and thus high novelty, while low values indicate high overlap with prior patents' text and thus low novelty. The second variable is dichotomous and indicates the presence of a bigram (i.e. sequence of two words) in the patent text that has not appeared on a previous granted USPTO patent. The variable takes a value of 0 when no new bigram is detected in the text of the focal patent, and a value of 1 if at least one new bigram is detected.

For our science-based outcome variables, we identify publications by searching for FAINs in the Web of Science database (Goldstein & Narayanamurti, 2018; De Rassenfosse et al., 2019; Wang et al., 2018). To complement this, we also search for the FAINs in Google Scholar. This is a necessary step as the Web of Science database only reports reliable funding information for papers published from 2008 onward. During this procedure, we apply three exclusion criteria. First, we open each identified publication to examine where it mentions the FAIN. This step helps to ensure that the publication is indeed acknowledging funding from a particular government grant. If this is not the case, such as when a FAIN only appears in the references list, the publication is not considered. Second, since we focus on articles published in scientific journals, we exclude outputs such as book chapters, dissertations, and conference proceedings/abstracts that mention the FAIN. Third, a small number of articles is excluded because the journals in which they are published are not indexed in the Web of Science.

Science-based productivity is measured at the project-level and is calculated as the total number of Web of Science-listed publications that acknowledge funding from a specific project. Like research-based spillovers, we calculate science-based spillovers by using forward citation data (e.g Abramo et al., 2020). Specifically, we count the number of citations a publication receives within the first five years after (and including) the year in which the publication was published. Citation data come from the Web of Science database. For science-based novelty, we use two alternative variables. The first one measures the presence of a new keyword in the publication based on the Keywords Plus data from the Web of Science. This approach follows prior studies that also consider the introduction of a new keyword in a publication as a reflection of novelty (e.g Azoulay et al., 2011; Bonaccorsi & Vargas, 2010). Keywords in the Keywords Plus data are provided by the Web of

[4] For example, a project titled 'High Throughput Combinatorial Chemistry Development of Complex Hydrides' and led by Intematix Corporation was evaluated in FY2006 and FY2007. In FY2006, it received a score of 3.1 for the relevance aspect (20% weight in overall score), 2.3 for the approach aspect (20% weight in overall score), 2.3 for the accomplishments aspect (35% weight in overall score), 2.7 for the collaboration aspect (10% weight in overall score), and 2.6 for the future research aspect (15% weight in overall score). The overall score in FY2006 was 2.545. In FY2007, it received a score of 3.7 for the relevance aspect (20% weight in overall score), 3.2 for the approach aspect (20% weight in overall score), 2.7 for the accomplishments aspect (35% weight in overall score), 3.1 for the collaboration aspect (10% weight in overall score), and 2.9 for the future research aspect (15% weight in overall score). The overall score in FY2007 was 3.07. Aggregating these two overall scores at the project level and taking the average, this results in a value of 2.8075 which we use as the dependent variable for this project.

[5] Three temporal variations in peer review scores need to be noted. First, projects within the American Recovery and Reinvestment Act (ARRA) division were never evaluated on the future research aspect, and were not evaluated on the relevance aspect in FY2010. Second, the weight of each project aspect in the total score changed over the years. Notably, the aspect related to 'technological accomplishments and progress' had a weight in the overall score of 20% in FY2003 and FY2004, 35% between FY2005 and FY2007, 40% between FY2008 and FY2012, and 45% between FY2013 and FY2017. Third, in FY2009, the reviews for projects within the divisions Education; Safety, Codes and Standards; and Technology Validation, were part of the Annual Merit Review report for the Vehicle Technologies Office (VTO). However, the five aspects on which these projects were evaluated were the same as those used for other projects within the Hydrogen Program during that fiscal year.

Science and not by the authors of a publication[6]. They are generated through an algorithm that inspects the titles of the references of a publication and looks for frequently occurring terms. For this dependent variable, a value of 1 means that at least one new keyword is present in the publication that has not appeared on a publication in previous years, whereas a value of 0 means that no new keywords are present. The second variable measures the age of keywords associated to the publication, which is an alternative way of measuring novelty (e.g Azoulay et al., 2011). The assumption is that keywords that were introduced a long time ago are less novel, as they represent familiar concepts that are already widely diffused. To calculate this variable, we look at the year in which a keyword first appeared in a publication. Then, we subtract that year from the year of publication of the focal publication, giving us a value that reflects the age of a keyword. Finally, we take the average age of all the keywords associated to the focal publication.

3.4. Independent variable: Earmarked projects

To the best of our knowledge, there is no existing database that records the specific FAINs associated with projects that are funded via earmarks. However, within the Hydrogen Program, we found several ways to find out which projects were funded via earmarks and which were not. First, within the Annual Progress Reports of the Hydrogen Program, earmarked projects were marked with an asterisk, stating that the project was "congressionally directed", which is another way of referring to the presence of earmarked funds. Second, we also investigated the DOE's Congressional Budget Requests, where congressionally directed projects are listed in a separate section. Third, the presence of earmarked funds in a project is frequently mentioned in the Annual Peer Review Evaluation Report of the Hydrogen Program and/or the AMR Proceedings. The proceedings include all slides that the project participants used when presenting their research during AMR meetings. Based on these data[7], we compute a variable (*Earmarked project*) that takes a value of 1 if the project had received earmarked funds and 0 if this was not the case.

In our analyses, we control for a substantial number of dimensions at the project-, patent- and publication-level. For the sake of brevity, we report a summary of the control variables in Table 1 and include a more extensive explanation for each variable in the Online Appendix.

Table 1
Concise description of control variables at project-, patent-, and publication-level

	Variable name	Description of variable
Project-level	For-profit lead	Dummy variable indicating that project is led by for-profit firm
	Prior experience	Dummy variable indicating that project lead has prior experience with Hydrogen Program
	Resource munificence	GDP per capita of home state of project lead in start year of project
	Project duration	Duration of project in days
	Project funding	Total amount of federal funds allocated to project
	Cooperative agreement	Dummy variable indicating that project is funded through a cooperative agreement
	EERE-funded	Dummy variable indicating that project is funded within core part of EERE
	Project similarity	Degree of similarity of project relative to all other ones in the sample
Patent-level	Patent grant lag	Number of days between filing date and grant date of patent
	Patent team size	Number of inventors listed on patent
	Patent independent claims	Number of independent claims on patent
	Patent scope	Number of CPC codes on patent (at main group level)
	Patent backward citations	Number of backward citations on patent
	Patent NPL citations	Number of NPL citations on patent
	For-profit lead	Dummy variable indicating that for-profit firm lead a project to which patent belongs
Publication-level	High-impact journal	Dummy variable indicating that publication is in a high-impact journal
	Publication team size	Number of authors listed on publication
	Publication interdisciplinarity	Number of Web of Science categories associated to journal of publication
	Publication keyword count	Number of keywords associated to publication (based on Keywords Plus data)
	Publication team internationalization	Dummy variable indicating that authors on publication are not all from the same country
	Publication references	Number of references listed on publication
	Publication title length	Number of alphanumeric characters in title of publication
	For-profit lead	Dummy variable indicating that for-profit firm lead a project to which publication belongs

3.5. Analytical approach

For project-level outcomes, we follow prior research and use a cross-sectional approach where data are aggregated at the project-level and each observation represents one project (e.g Du et al., 2014; Goldstein & Narayanamurti, 2018). We use OLS regressions to test the relationship between earmarks and project evaluation scores. To examine the relationship between earmarked projects and the quantity of patents and publications, we use count models since these two dependent variables only contain non-negative integer values. Poisson models are typically used to analyze count data. However, when the dependent variable is over-dispersed, such models can produce inconsistent estimates. In our sample, we find that the dependent variables, representing the quantity of patents and publications are over-dispersed. When this is the case, negative binomial models are usually preferred because they accommodate over-dispersed dependent variables through the inclusion of an additional dispersion parameter. Moreover, when we assess goodness of fit by looking at the log-likelihood values of Poisson and negative binomial models, we conclude that negative binomial models consistently yield the highest model fit[8].

For the patent- and publication-level analyses, we follow prior studies (e.g Fleming, 2001) and use negative binomial regressions to examine the effect of earmarks on spillovers. This choice is also motivated by the fact that forward citations of patents and publications are over-dispersed. We use OLS regressions to test the relationship between earmarks and patent-level novelty (as measured by the extent of overlap in the patent text between the focal patent and prior patents). The

[6] Keywords Plus data are only reliably available from 1991 and onwards. Hence, when we search for previously occurring keywords in the Keywords Plus data field, we restrict the selection of publications to those with a publication year after (and including) 1991.

[7] We also investigate external data sources, such as the Office of Management and Budget (OMB) earmarks datasets, the Taxpayers for Common Sense (TCS) earmarks datasets, and the Citizens Against Government Waste (CAGW) earmarks database. The latter datasets helped us validating previously collected information.

[8] Nevertheless, we run sensitivity checks and obtain very stable results for the patent quantity, publication quantity, patent forward citations, and publication forward citations models when estimating them with Poisson models. The results are available from the authors upon request.

presence of a new bigram in the patent text is examined using logistic regressions, since the outcome variable is dichotomous. We examine publication novelty, as reflected through the presence of a new keyword, using logistic regressions. Finally, for the alternative publication novelty measure, keyword age, we use OLS regressions.

3.6. Results

Tables 2-4 display the descriptive statistics and correlation matrix for the project-, patent-, and publication-level analyses[9]. On average, the 321 projects in our sample produce 0.48 patents and 3.71 publications. Out of the 321 projects, 74 projects are earmarked. Whereas prior studies on public R&D funding tend to focus on university-led projects (e.g Sandström & Hällsten, 2008; Wang et al., 2018), 171 projects in our sample are led by for-profit firms.

The regression results at the project-level for project evaluation scores, patent quantity, and publication quantity are shown in Table 5. In Models 1 and 2, we examine the relationship between earmarked projects and project evaluation scores. We introduce the variable for earmarked projects in Model 2 and find that it has a negative and statistically significantly relationship with project evaluation scores (Model 2: $\beta_{\text{Earmarked project}} = -0.336$, $p = 0.000$). Subsequently, we examine the relationship between earmarked projects and the quantity of patents (Models 3 and 4) and publications (Models 5 and 6) resulting from projects. We find that there is no statistically significant relationship between earmarked projects and the quantity of patents (Model 4: $\beta_{\text{Earmarked project}} = 0.518$, $p = 0.162$). At the same time, we find a positive and statistically significant relationship between earmarked projects and the quantity of publications (Model 6: $\beta_{\text{Earmarked project}} = 0.563$, $p = 0.016$).

The patent- and publication-level analyses of spillovers and novelty are shown in Tables 6 and 7 respectively. Model 2 in Table 6 and Model 2 in Table 7 indicate that, while spillovers between patents from earmarked and non-earmarked projects do not statistically significantly differ (Model 2: $\beta_{\text{Earmarked project}} = -0.426$, $p = 0.090$), the publications that result from earmarked projects are associated to statistically significantly fewer spillovers than those resulting from non-earmarked projects (Model 2: $\beta_{\text{Earmarked project}} = -0.493$, $p = 0.000$). We also find indications that, in comparison to non-earmarked projects, patents from earmarked projects are more novel. Specifically, in Model 4 in Table 6, we find that earmarked projects produce patents that are more dissimilar from existing patents and are, thus, more novel (Model 4: $\beta_{\text{Earmarked project}} = 0.004$, $p = 0.010$). However, patents from earmarked and non-earmarked projects are not different in terms of the inclusion of new bigrams in the patent text (Model 6: $\beta_{\text{Earmarked project}} = 0.321$, $p = 0.550$). Third, in Models 4 and 6 in Table 7, we find no indications that the publications of earmarked and non-earmarked projects differ in terms of novelty. Specifically, the relationships between earmarks and the presence of a new keyword (Model 4: $\beta_{\text{Earmarked project}} = 0.255$, $p = 0.283$) and the age of keywords (Model 6: $\beta_{\text{Earmarked project}} = 0.103$, $p = 0.486$) are not statistically significant. Sensitivity checks, in which we examine various operationalizations of the dependent variables for research and scientific productivity and spillovers, are reported in the Online Appendix.

In sum, we find a strong negative effect of earmarks on project evaluation scores. When examining the scientific and research output of projects, however, this strong negative effect is not clearly present. We can only observe a negative effect for one dependent variable, i.e.

[9] We note that some patents and publications that are used to measure the quantity of patents and publications resulting from the projects in our sample are excluded from the patent- and publication-level analyses because of missing data (i.e. for patents, one patent had missing data on the patent text and CPC field while for publications, Keyword Plus data was not available for all publications).

Table 2
Correlation matrix and descriptive statistics for project-level analyses

	Variable	Mean	S.D.	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12
1	Project evaluation scores	2.915	0.383	1.630	3.920	1.000											
2	Patent quantity	0.483	1.255	0	9	0.075	1.000										
3	Publication quantity	3.707	7.313	0	61	0.003	0.151	1.000									
4	Earmarked project	0.231	0.422	0	1	-0.451	-0.010	0.055	1.000								
5	For-profit lead	0.533	0.500	0	1	0.175	0.087	-0.317	-0.214	1.000							
6	Prior experience	0.495	0.501	0	1	0.050	-0.024	0.219	-0.143	-0.009	1.000						
7	Resource munificence	54.021	13.145	33.726	184.073	0.162	-0.064	-0.095	-0.174	0.228	0.055	1.000					
8	Project duration	1602.732	705.042	198	3590	0.218	0.100	0.184	-0.308	-0.100	0.031	-0.053	1.000				
9	Project funding	2.613	4.508	0	37.999	0.190	0.121	0.023	-0.064	0.173	-0.035	-0.065	0.352	1.000			
10	Cooperative agreement	0.389	0.488	0	1	0.046	0.136	-0.025	-0.103	-0.020	-0.063	-0.101	0.168	0.194	1.000		
11	EERE-funded	0.844	0.363	0	1	-0.089	0.015	0.123	0.072	0.011	0.030	-0.134	0.081	-0.027	-0.256	1.000	
12	Project similarity	71.733	4.761	53.882	80.040	-0.057	0.108	0.066	-0.015	0.042	0.096	-0.066	-0.050	-0.025	0.136	0.003	1.000

N = 321

Table 3
Correlation matrix and descriptive statistics for patent-level analyses

	Variable	Mean	S.D.	Min	Max	1	2	3	4	5	6	7	8	9	10	11
1	Patent forward citations	4.604	8.943	0	74	1.000										
2	Patent text dissimilarity	0.977	0.008	0.950	0.994	-0.001	1.000									
3	Patent new bigram	0.448	0.499	0	1	0.083	0.074	1.000								
4	Earmarked project	0.214	0.412	0	1	-0.027	0.195	0.007	1.000							
5	Patent grant lag	1309.942	616.659	138	3269	0.069	-0.083	0.329	-0.011	1.000						
6	Patent team size	3.331	1.879	1	13	-0.024	0.015	0.001	-0.109	0.065	1.000					
7	Patent independent claims	2.110	1.218	1	8	0.057	-0.101	0.176	-0.074	0.102	0.084	1.000				
8	Patent scope	3.091	2.094	1	10	0.070	0.118	0.030	-0.061	-0.116	-0.171	-0.032	1.000			
9	Patent backward citations	22.435	29.489	0	175	0.195	-0.188	-0.171	0.098	-0.105	-0.087	-0.004	0.017	1.000		
10	Patent NPL citations	15.448	30.377	0	247	0.199	0.118	0.104	-0.130	-0.008	0.088	-0.062	0.192	0.243	1.000	
11	For-profit lead	0.643	0.481	0	1	-0.060	-0.289	-0.173	-0.238	0.016	0.002	0.112	-0.234	0.306	-0.260	1.000

N = 154

Table 4
Correlation matrix and descriptive statistics for publication-level analyses

	Variable	Mean	S.D.	Min	Max	1	2	3	4	5	6	7	8	9	10	11	12
1	Publication forward citations	44.859	123.981	0	2272	1.000											
2	Publication keyword novelty	0.109	0.312	0	1	0.147	1.000										
3	Publication keyword age	17.476	3.313	2	26	-0.130	-0.333	1.000									
4	Earmarked project	0.276	0.447	0	1	-0.107	-0.004	-0.061	1.000								
5	High-impact journal	0.296	0.457	0	1	0.240	0.024	-0.011	-0.129	1.000							
6	Publication team size	4.725	2.489	1	20	0.100	-0.074	0.071	-0.055	0.090	1.000						
7	Publication interdisciplinarity	2.342	1.136	1	6	-0.057	0.029	-0.064	0.025	0.097	-0.055	1.000					
8	Publication keyword count	7.531	2.882	1	10	0.094	0.158	0.019	-0.161	0.117	0.062	-0.040	1.000				
9	Publication team internationalization	0.233	0.423	0	1	0.024	-0.043	0.066	0.062	0.036	0.338	-0.028	-0.026	1.000			
10	Publication references	41.355	26.999	6	350	0.257	0.169	-0.048	-0.118	0.091	0.040	-0.075	0.487	0.044	1.000		
11	Publication title length	84.650	26.818	18	182	-0.059	-0.066	0.123	-0.072	0.076	0.035	-0.065	0.105	-0.006	0.051	1.000	
12	For-profit lead	0.225	0.418	0	1	-0.031	0.025	0.071	-0.256	-0.050	0.018	0.051	-0.012	-0.003	-0.029	0.093	1.000

N = 1121

Table 5
Regressions estimating project evaluation scores, patent quantity, and publication quantity

	Project evaluation scores		Patent quantity		Publication quantity	
	(1)	(2)	(3)	(4)	(5)	(6)
For-profit lead	0.101*	0.030	0.952**	1.034**	-1.475***	-1.337***
	[0.046]	[0.044]	[0.345]	[0.354]	[0.239]	[0.240]
Prior experience	0.046	0.025	0.140	0.133	0.993***	0.992***
	[0.040]	[0.037]	[0.299]	[0.292]	[0.192]	[0.188]
Resource munificence	0.003	0.002	-0.035	-0.033	-0.011	-0.007
	[0.001]	[0.001]	[0.019]	[0.020]	[0.009]	[0.009]
Project duration	0.000**	0.000	0.000	0.000	0.000*	0.000**
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Project funding	0.005	0.008*	0.053	0.048	0.091*	0.080*
	[0.005]	[0.004]	[0.039]	[0.036]	[0.042]	[0.037]
Cooperative agreement	0.120	0.065	0.586	0.617	0.129	0.155
	[0.061]	[0.056]	[0.448]	[0.436]	[0.273]	[0.266]
EERE-funded	0.017	0.065	0.354	0.268	0.177	0.070
	[0.072]	[0.069]	[0.507]	[0.496]	[0.350]	[0.349]
Project similarity	-0.003	-0.003	0.013	0.019	0.011	0.011
	[0.005]	[0.005]	[0.038]	[0.035]	[0.020]	[0.019]
Earmarked project		-0.336***		0.518		0.563*
		[0.060]		[0.371]		[0.234]
Fiscal year dummies	Y	Y	Y	Y	Y	Y
Division dummies	Y	Y	Y	Y	Y	Y
N	321	321	321	321	321	321
R^2	0.237	0.333				
Adjusted R^2	0.189	0.289				
Pseudo R^2			0.083	0.086	0.117	0.120
Log Likelihood			-242.444	-241.719	-592.071	-589.765

*p<5%, **p<1%, ***p<0.1%. We report robust standard errors between brackets.

Table 6
Patent-level analyses of forward citations, text dissimilarity, and new bigram

	Patent forward citations		Patent text dissimilarity		Patent new bigram	
	(1)	(2)	(3)	(4)	(5)	(6)
Patent grant lag	-0.000	-0.000	-0.000	-0.000	0.001*	0.001*
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Patent team size	-0.040	-0.038	0.000	0.000	-0.117	-0.114
	[0.049]	[0.049]	[0.000]	[0.000]	[0.145]	[0.144]
Patent independent claims	0.006	0.003	0.000	0.000	0.306	0.311
	[0.063]	[0.065]	[0.001]	[0.001]	[0.201]	[0.199]
Patent scope	0.032	0.025	0.000	0.000	0.001	0.007
	[0.049]	[0.047]	[0.000]	[0.000]	[0.101]	[0.100]
Patent backward citations	0.012*	0.015**	-0.000*	-0.000***	-0.012	-0.014
	[0.005]	[0.005]	[0.000]	[0.000]	[0.009]	[0.009]
Patent NPL citations	0.010	0.007	0.000	0.000	0.019*	0.021*
	[0.006]	[0.005]	[0.000]	[0.000]	[0.008]	[0.008]
For-profit lead	-0.315	-0.447	-0.003	-0.001	-0.576	-0.443
	[0.293]	[0.299]	[0.002]	[0.002]	[0.476]	[0.510]
Earmarked project		-0.426		0.004**		0.321
		[0.251]		[0.002]		[0.536]
Filing year dummies	Y	Y	Y	Y	Y	Y
N	154	154	154	154	146	146
R^2			0.189	0.228		
Adjusted R^2			0.081	0.119		
Pseudo R^2	0.074	0.077			0.241	0.243
Log Likelihood	-367.051	-365.874			-76.658	-76.475

*p<5%, **p<1%, ***p<0.1%. We report robust standard errors between brackets. In Models 5 and 6, eight observations are dropped because, in the years in which the patents were filled, all patents contain 0 new bigrams and thus there is no variance in the dependent variable.

scientific spillovers. For most other dependent variables, no statistically significant difference can be found between the outputs of earmarked and non-earmarked projects. We even observe that for some dependent variables – i.e. scientific productivity and research novelty – earmarked projects score higher than non-earmarked projects. Together, these findings point to a misalignment between (1) how earmarked projects are evaluated by peer reviewers and (2) the extent of productivity, spillovers, and novelty of research- and science-based outputs.

3.7. Investigating underlying mechanisms

We consider and test several potential mechanisms that can explain the observed misalignment between expert project evaluations and tangible outcomes. First, we consider the possibility that projects that went through a competitive selection process are inherently different from those that went through a political selection process. Second, we explore whether earmarked projects might receive lower evaluation

Table 7
Publication-level analyses of forward citations, keyword novelty, and keyword age

	Publication forward citations		Publication keyword novelty		Publication keyword age	
	(1)	(2)	(3)	(4)	(5)	(6)
High-impact journal	1.053***	0.971***	0.113	0.145	0.055	0.068
	[0.101]	[0.100]	[0.213]	[0.215]	[0.140]	[0.140]
Publication team size	0.101***	0.094***	-0.095	-0.092	-0.002	-0.001
	[0.017]	[0.017]	[0.054]	[0.054]	[0.029]	[0.029]
Publication interdisciplinarity	-0.037	-0.018	0.092	0.090	-0.195***	-0.197***
	[0.037]	[0.036]	[0.089]	[0.089]	[0.058]	[0.058]
Publication keyword count	0.020	0.018	0.217***	0.221***	-0.040	-0.038
	[0.014]	[0.014]	[0.051]	[0.051]	[0.031]	[0.031]
Publication team internationalization	-0.076	-0.031	-0.143	-0.167	-0.079	-0.089
	[0.091]	[0.088]	[0.276]	[0.276]	[0.175]	[0.174]
Publication references	0.012***	0.012***	0.009**	0.009**	-0.022***	-0.022***
	[0.002]	[0.002]	[0.003]	[0.003]	[0.003]	[0.002]
Publication title length	-0.002	-0.002	-0.008*	-0.008*	0.002	0.002
	[0.001]	[0.001]	[0.004]	[0.004]	[0.003]	[0.003]
For-profit lead	0.126	-0.014	0.268	0.333	-0.327	-0.300
	[0.098]	[0.104]	[0.238]	[0.245]	[0.182]	[0.183]
Earmarked project		-0.493***		0.255		0.103
		[0.081]		[0.238]		[0.147]
Publication year dummies	Y	Y	Y	Y	Y	Y
N	1121	1121	1121	1121	1121	1121
R^2					0.551	0.551
Adjusted R^2					0.543	0.542
Pseudo R^2	0.058	0.062	0.097	0.099		
Log Likelihood	-5043.166	-5020.795	-348.134	-347.600		

*p<5%, **p<1%, ***p<0.1%. We report robust standard errors between brackets.

scores mainly because peer reviewers consider them to be less relevant to the overall Hydrogen Program. Third, we consider the hypothesis that researchers within earmarked projects might put less effort into presenting the results of their research for evaluation by peer reviewers. Fourth, we investigate the possibility of a systematic bias of peer reviewers toward earmarked projects. Finally, we explore whether the functional background of peer reviewers can explain this misalignment.

3.7.1. *Inherent differences in characteristics of earmarked and non-earmarked projects*

We first consider the possibility that the relationship between earmarks and project quality is related to a project selection effect, where projects that went through a competitive selection process are inherently different from those that went through a political selection process.

Table 8
t-tests for differences between earmarked and non-earmarked projects

	Mean non-earmarked (N=247)	Mean earmarked (N=74)	Difference	SE
For-profit lead	0.591	0.338	0.253***	0.065
Prior experience	0.534	0.365	0.170*	0.066
Resource munificence	55.272	49.846	5.426**	1.718
Project duration	1721.441	1206.500	514.941***	89.028
Project funding	2.770	2.089	0.682	0.597
Cooperative agreement	0.417	0.297	0.120	0.064
EERE-funded	0.830	0.892	-0.062	0.048
Project similarity	71.771	71.607	0.164	0.632
Start fiscal year	2006.445	2005.932	0.513	0.300
Fuel cell division	0.304	0.378	-0.075	0.062
Hydrogen production & delivery division	0.247	0.230	0.017	0.057
Hydrogen storage division	0.251	0.149	0.102	0.055
Other division	0.198	0.243	-0.045	0.054

*p<5%, **p<1%, ***p<0.1%.

To evaluate this possibility, we perform t-tests on each control variable for earmarked and non-earmarked projects. The findings, displayed in Table 8, show that earmarked projects (1) are less likely to be led by for-profit firms, (2) are less likely to be led by an organization with prior experience with the Hydrogen Program, (3) are led by organizations that are located in less resource munificent regions, and (4) tend to have a shorter duration.

To further explore the potential existence of a project selection effect, we first run the regressions for project evaluation scores, patent quantity, and publication quantity, while excluding the control variables 'for-profit lead', 'prior experience', 'resource munificence', and 'project duration'. Then, we enter these four control variables in a sequential manner into the regressions (see Table 9). If we detect strong changes in the sign, magnitude, and statistical significance of the coefficient for the earmark variable, this is an indication that the relationship between earmarks and project-level outcomes is influenced by inherent differences in characteristics between earmarked and non-earmarked projects. In Models 1-6, we examine project evaluation scores and observe that the coefficient for the earmark variable does not change considerably when the four control variables are sequentially added to the regression. In a similar vein, the coefficient for the earmark variable does not change considerably when the four control variables are sequentially added to the regression estimating patent quantity (Models 7-12). We examine publication quantity in Models 13-18 and notice that the coefficient of the earmark variable changes considerably when the four control variables are sequentially added in the regressions. Above all, we see that the coefficient of the earmark variable becomes substantially smaller when we add the variable 'for-profit lead' into the regression. This suggests that the positive association between earmarking and publication output that we detect in Model 13 in Table 9 is at least partially driven by the fact that organizations that are not for-profit firms (such as universities) are more likely to lead earmarked projects and that such organizations generate more project-based publications.

In sum, our first set of additional analyses suggest that the positive effect of earmarks on scientific productivity can be at least partially explained by a project selection effect. We find that projects, which went

Table 9
Testing project selection effect for regressions estimating project evaluation scores, patent quantity, and publication quantity.

	Project evaluation scores						Patent quantity						Publication quantity					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
For-profit lead		0.027				0.030		0.630*				1.034**		-1.495***				-1.337***
		[0.044]				[0.044]		[0.298]				[0.354]		[0.228]				[0.240]
Prior experience			0.027			0.025			0.018			0.133			0.969***			0.992***
			[0.037]			[0.037]			[0.282]			[0.292]			[0.205]			[0.188]
Resource munificence				0.002		0.002				-0.013		-0.033				-0.024*		-0.007
				[0.001]		[0.001]				[0.016]		[0.020]				[0.010]		[0.009]
Project duration					0.000	0.000					0.000	0.000					0.001***	0.000**
					[0.000]	[0.000]					[0.000]	[0.000]					[0.000]	[0.000]
Project funding	0.011***	0.011**	0.011***	0.011***	0.009*	0.008*	0.098**	0.085**	0.098**	0.097**	0.083*	0.048	0.123*	0.140**	0.107*	0.127*	0.057	0.080*
	[0.003]	[0.003]	[0.003]	[0.003]	[0.004]	[0.004]	[0.036]	[0.033]	[0.036]	[0.037]	[0.041]	[0.036]	[0.049]	[0.052]	[0.043]	[0.052]	[0.040]	[0.037]
Cooperative agreement	0.058	0.060	0.057	0.063	0.059	0.065	0.740	0.741	0.735	0.716	0.711	0.617	0.162	0.366	0.039	0.128	0.041	0.155
	[0.055]	[0.055]	[0.055]	[0.056]	[0.055]	[0.056]	[0.457]	[0.461]	[0.445]	[0.455]	[0.439]	[0.436]	[0.299]	[0.297]	[0.291]	[0.300]	[0.275]	[0.266]
EERE-funded	0.069	0.064	0.072	0.078	0.060	0.065	0.583	0.570	0.581	0.496	0.529	0.268	-0.132	0.327	-0.172	-0.188	-0.363	0.070
	[0.070]	[0.070]	[0.070]	[0.071]	[0.069]	[0.069]	[0.456]	[0.443]	[0.458]	[0.466]	[0.464]	[0.496]	[0.379]	[0.380]	[0.370]	[0.378]	[0.368]	[0.349]
Project similarity	-0.004	-0.004	-0.004	-0.004	-0.003	-0.003	0.014	0.009	0.014	0.012	0.024	0.019	-0.003	0.001	-0.009	-0.008	0.023	0.011
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.036]	[0.038]	[0.036]	[0.035]	[0.035]	[0.035]	[0.020]	[0.019]	[0.019]	[0.019]	[0.019]	[0.019]
Earmarked project	-0.376***	-0.368***	-0.373***	-0.367***	-0.358***	-0.336***	0.312	0.424	0.311	0.271	0.397	0.518	0.915***	0.431	0.893***	0.797**	1.072***	0.563*
	[0.058]	[0.061]	[0.058]	[0.058]	[0.058]	[0.060]	[0.353]	[0.355]	[0.351]	[0.367]	[0.373]	[0.371]	[0.269]	[0.258]	[0.253]	[0.269]	[0.254]	[0.234]
Fiscal year dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Division dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	321	321	321	321	321	321	321	321	321	321	321	321	321	321	321	321	321	321
R^2	0.324	0.325	0.325	0.327	0.327	0.333												
Adjusted R^2	0.288	0.287	0.287	0.289	0.289	0.289												
Pseudo R^2							0.073	0.078	0.073	0.073	0.074	0.086	0.073	0.100	0.085	0.076	0.081	0.120
Log Likelihood							-245.219	-243.768	-245.217	-245.005	-244.895	-241.719	-621.698	-603.419	-613.208	-619.702	-616.143	-589.765

*p<5%, **p<1%, ***p<0.1%. We report robust standard errors between brackets.

Table 10
Regressions estimating five aspects of project evaluation scores separately

	Relevance		Approach		Tech. A&P		Tech. Transfer		Future Res.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
For-profit lead	0.129**	0.039	0.146**	0.074	0.080	0.009	0.016	-0.049	0.142**	0.085
	[0.046]	[0.042]	[0.048]	[0.047]	[0.056]	[0.055]	[0.064]	[0.064]	[0.047]	[0.046]
Prior experience	0.083	0.058	0.051	0.030	0.045	0.027	0.021	0.003	0.045	0.026
	[0.045]	[0.042]	[0.043]	[0.041]	[0.047]	[0.045]	[0.055]	[0.053]	[0.040]	[0.039]
Resource munificence	0.002	0.000	0.002	0.001	0.004*	0.003	0.004*	0.003	0.002	0.001
	[0.001]	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]	[0.002]	[0.002]	[0.001]	[0.001]
Project duration	0.000*	-0.000	0.000*	0.000	0.000*	0.000	0.000***	0.000*	0.000**	0.000
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Project funding	0.009	0.012**	0.005	0.008*	0.004	0.007	0.005	0.007	0.001	0.004
	[0.005]	[0.004]	[0.005]	[0.004]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.004]
Cooperative agreement	0.111	0.040	0.162*	0.106	0.092	0.038	0.212*	0.162	0.091	0.045
	[0.065]	[0.058]	[0.063]	[0.059]	[0.074]	[0.067]	[0.086]	[0.083]	[0.063]	[0.060]
EERE-funded	-0.019	0.043	0.058	0.106	0.010	0.057	0.116	0.159	0.005	0.047
	[0.080]	[0.077]	[0.074]	[0.074]	[0.092]	[0.089]	[0.097]	[0.096]	[0.073]	[0.071]
Project similarity	0.001	0.000	-0.009	-0.009	-0.001	-0.001	-0.002	-0.003	-0.006	-0.006
	[0.006]	[0.005]	[0.006]	[0.005]	[0.006]	[0.005]	[0.007]	[0.007]	[0.005]	[0.005]
Earmarked project		-0.428***		-0.340***		-0.334***		-0.307***		-0.273***
		[0.068]		[0.065]		[0.069]		[0.079]		[0.062]
Fiscal year dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Division dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	317	317	321	321	317	317	321	321	309	309
R^2	0.208	0.348	0.248	0.335	0.179	0.249	0.251	0.295	0.217	0.281
Adjusted R^2	0.158	0.304	0.200	0.290	0.126	0.199	0.203	0.248	0.166	0.231

*p<5%, **p<1%, ***p<0.1%. We report robust standard errors between brackets. The number of observations is not the same for each project aspect, as some projects were not evaluated on every aspect in every year.

through a competitive selection process, are inherently different from those that went through a political selection process on particular characteristics (i.e. type of lead organization, prior experience, resource munificence of lead organization's environment, project duration) and these differences (i.e. type of lead organization) influence science-based productivity. The observed differences in project characteristics, however, do not seem to influence the significant negative association between earmarks and project evaluation scores. This implies that we cannot find evidence for a project selection effect that can potentially explain why earmarked projects underperform relatively to non-earmarked ones in terms of peer review evaluation scores.

Table 11
Regressions estimating average number of words per presentation

	Average number of words per presentation	
	(1)	(2)
For-profit lead	-117.589	-118.901
	[82.308]	[78.341]
Prior experience	133.065*	132.698*
	[62.848]	[63.916]
Resource munificence	-0.861	-0.879
	[2.090]	[2.137]
Project duration	-0.257***	-0.258***
	[0.063]	[0.062]
Project funding	2.069	2.124
	[6.448]	[6.167]
Cooperative agreement	-29.962	-30.949
	[89.912]	[87.802]
EERE-funded	-91.583	-90.690
	[104.224]	[103.205]
Project similarity	25.004***	25.000***
	[6.801]	[6.818]
Total words used in presentations	0.149***	0.149***
	[0.013]	[0.013]
Earmarked project		-6.138
		[92.382]
Fiscal year dummies	Y	Y
Division dummies	Y	Y
N	321	321
R^2	0.689	0.689
Adjusted R^2	0.668	0.667

*p<5%, **p<1%, ***p<0.1%. We report robust standard errors between brackets.

3.7.2. *Differential scores on five project aspects*

In the main analyses, we use the overall project score to capture the evaluations provided by peer reviewers. The overall scores are a weighted average of the five aspects (relevance, approach, technological accomplishments, collaboration, future research) on which projects are evaluated. It could be possible that the negative association between earmarked projects and peer review evaluation scores is not present for all five project aspects. Above all, earmarked projects may receive particularly low scores in terms of relevance, as critics often argue that earmarked projects funded through R&D grants tend to focus on activities that are not well-aligned with program-wide goals. To examine whether earmarked projects mainly score worse in terms of relevance, we examine each of the five project aspects separately in Table 10. We find a negative and statistically significant relationship between earmarked projects and the five individual project aspects. Hence, the negative relationship between earmarks and peer review evaluation scores does not seem to be driven by one particular project aspect on which earmarked projects receive relatively low scores compared to non-earmarked projects.

3.7.3. *Differences in effort put into presenting research results*

We consider the possibility that researchers in earmarked projects put less effort into presenting their research results for evaluation by peer reviewers, potentially negatively impacting evaluation scores. By looking at the qualitative comments provided by peer reviewers to support their scores, we already notice that the quality of slides, especially in terms of clarity and amount of elaboration, is often mentioned. We argue that, for researchers within earmarked projects, there is less at stake when it comes to these presentations, which might affect their quality. Because funding decisions for earmarked projects are largely outside the authority of program managers, the evaluations provided by

Table 12
Regressions estimating average number of words per reviewer, percentage of differentiation words, percentage of discrepancy words

	Words per reviewer		Differentiation words		Discrepancy words	
	(1)	(2)	(3)	(4)	(5)	(6)
For-profit lead	-11.458*	-9.079	0.002*	0.001	0.001	0.001
	[4.589]	[4.685]	[0.001]	[0.001]	[0.001]	[0.001]
Prior experience	2.353	2.917	0.001	0.001	-0.000	-0.001
	[3.781]	[3.732]	[0.001]	[0.001]	[0.001]	[0.000]
Resource munificence	0.041	0.059	-0.000*	-0.000*	-0.000	-0.000
	[0.129]	[0.133]	[0.000]	[0.000]	[0.000]	[0.000]
Project duration	-0.005	-0.003	0.000	-0.000	-0.000	-0.000
	[0.004]	[0.003]	[0.000]	[0.000]	[0.000]	[0.000]
Project funding	0.478	0.309	-0.000	-0.000	-0.000**	-0.000**
	[0.307]	[0.307]	[0.000]	[0.000]	[0.000]	[0.000]
Cooperative agreement	-12.768*	-11.497	-0.001	-0.002	-0.002*	-0.002*
	[6.405]	[6.243]	[0.001]	[0.001]	[0.001]	[0.001]
EERE-funded	-13.795	-16.430*	0.001	0.001	0.000	0.000
	[7.264]	[7.138]	[0.001]	[0.001]	[0.001]	[0.001]
Project similarity	0.069	0.062	0.000	0.000	0.000*	0.000*
	[0.403]	[0.401]	[0.000]	[0.000]	[0.000]	[0.000]
Total words in evaluations	0.012***	0.012***	0.000	0.000	0.000*	0.000*
	[0.001]	[0.001]	[0.000]	[0.000]	[0.000]	[0.000]
Overall project score	-25.052***	-19.837***	-0.013***	-0.014***	-0.001	-0.002**
	[5.527]	[5.947]	[0.001]	[0.001]	[0.001]	[0.001]
Earmarked project		13.771*		-0.002		-0.002**
		[6.479]		[0.001]		[0.001]
Fiscal year dummies	Y	Y	Y	Y	Y	Y
Division dummies	Y	Y	Y	Y	Y	Y
N	316	316	321	321	321	321
R^2	0.625	0.633	0.410	0.417	0.136	0.162
Adjusted R^2	0.599	0.605	0.369	0.374	0.076	0.100

*p<5%, **p<1%, ***p<0.1%. We report robust standard errors between brackets. In Models 1 and 2, the variable 'Total words in evaluations' does not count the words used in the evaluation reports from FY2003 and FY2004 as the number of reviewers per evaluation was not listed in those reports.

peer reviewers might be less consequential for these projects. Researchers in earmarked projects might therefore be less motivated to put a lot of effort and time in creating high-quality slides for their evaluation presentations.

To explore this alternative explanation, we look at the slides that researchers used in their presentations during the AMR meetings. We compute the outcome variable as the total number of words on the slides divided by the number of times the researchers presented during the AMR. For example, if a project was reviewed in four different years, and the four presentations in those years contained a total of 7000 words, the dependent variable equals 1750. In Model 2 in Table 11, we find that there is no difference between earmarked and non-earmarked projects in terms of average number of words per presentation (Model 2: $\beta_{\text{Earmarked project}} = -6.138$, $p = 0.947$). Hence, we cannot find evidence that the negative relationship between earmarks and peer review evaluation scores is driven by systematic differences in the effort that researchers within earmarked and non-earmarked projects put in the presentation of their work to peer reviewers.

3.7.4. *Possibility of a bias toward earmarked projects in peer reviews*

It is possible that peer reviewers have a cognitive bias toward earmarked projects because they often face (in)direct negative consequences from earmarks. Earmarks redirect funds to projects that are outside the scope of conventional selection processes, which reduces the total amount of funds that can be allocated within them. In other words, earmarked projects have the potential to carve into the funding of other non-earmarked projects. Consequently, when a large portion of appropriated funds is earmarked, ongoing and prospective projects may receive less funding, or even no funding at all. As most peer reviewers are also researchers, who rely on external funding to conduct their research activities, it is possible that, when they need to evaluate an earmarked project, this earmark label will bias their evaluation and subsequent scoring.

To explore the potential existence of a bias toward earmarked projects, we perform text analyses on the qualitative comments provided by peer reviewers to support their evaluation scores. Some of the comments in the evaluation reports already provide indications that peer reviewers might have a form of bias toward earmarked projects, for example: "An earmarked (congressionally directed) project is always going to be a weakness, simply because it does not undergo the rigors of [an initial] peer-review. In this case, some of the work (specifically the storage work) seems to be well-aligned and of high quality, but for earmarks that is really the exception rather than the rule (U.S. Department of Energy, 2005, p. 221)."

To perform this analysis in a comprehensive way, we create an algorithm that (1) extracts the text from the PDF documents of the Hydrogen Program Annual Merit Review and Peer Evaluation reports between FY2003 and FY2017, (2) cleans the text by removing redundant information (such as page numbers, page headers, etc.), (3) extracts the qualitative comments used to support the scores for each project aspect, and (4) tokenizes and lower-cases the words used in those comments. In this way, we obtain an overview of all the words that reviewers used in their qualitative comments to support the scores they gave to the projects. In the analyses, we apply the commonly used bag-of-words approach to detect bias in reviewer comments (e.g Van den Besselaar et al., 2018).

We compute three different dependent variables. The first dependent variable measures the average number of words provided in the comments by each peer reviewer. For example, if a project was reviewed in three different years, by four peer reviewers each time, and the total number of words in those reviews was 1800, then the dependent variable equals 150.

Next to the absolute number of words used in reviews of earmarked and non-earmarked projects, we also look at the nature of those words. The second dependent variable is based on the "differentiation" dictionary that is part of the 2015 version of the text analysis program Linguistic Inquiry and Word Count (LIWC). This dictionary includes words such as "however", "except", and "versus" and it has been used by prior research to capture nuanced and complex thinking (e.g Crilly et al., 2016). This dependent variable is calculated as the number of differentiation words divided by the total number of words. Higher values on this dependent variable indicate that the evaluations contain more complex and nuanced opinions.

The third dependent variable uses the "discrepancy" dictionary from the 2015 version of LIWC. In the context of peer review, this dictionary is particularly relevant because it can indicate the extent to which reviewers made suggestions for improvement. For example, words such as "should", "ought" and "preferable" are part of this dictionary. Higher values on this dependent variable indicate that the evaluations contain more improvement-focused comments.

In the analyses, overall project scores are held constant to ensure that we can better identify to what extent the quantity and nature of words used by peer reviewers in the evaluations differ because of the project selection mechanism (i.e. earmarked versus non-earmarked). In Model 2 in Table 12, we find that earmarked projects receive evaluations that, on average, contain more words per peer reviewer than non-earmarked projects (Model 2: $\beta_{\text{Earmarked project}} = 13.771$, $p = 0.034$). One interpretation of this result is that, because reviewers might feel they need to additionally justify their comments when reviewing earmarked projects, they elaborate more on their opinions. Moreover, looking into the nature of the words used in the reviews, we find that evaluations of earmarked projects contain similar ratios of differentiation words (Model 4: $\beta_{\text{Earmarked project}} = -0.002$, $p = 0.120$) but lower ratios of discrepancy words (Model 6: $\beta_{\text{Earmarked project}} = -0.002$, $p = 0.004$) compared to those of non-earmarked projects. This suggests that peer reviewers' evaluations of earmarked projects are indeed of a different nature (i.e. more verbose and less improvement-focused) than those provided to non-earmarked projects. These observations provide first indications that reviewers are likely to be biased toward earmarked projects and help to explain the misalignment in performance between earmarked projects' peer review scores and research- and science-based outputs.

3.7.5. Identity of reviewers evaluating projects

In a final check, we explore the possibility that the negative association between earmarks and project evaluation scores might be explained by the reviewers' background. Scanning the media and literature on earmarks, we noticed that some of the most outspoken critics of this funding tool had a university background. As a result, the potential bias towards earmarked projects might be more outspoken for university researchers than for reviewers with a different background. To study the effect of the profile of peer reviewers on project evaluation scores, we would ideally need data on the identity of the specific set of peer reviewers associated to each project evaluation. However, since evaluations in the Hydrogen Program are anonymized, this data was not available. Therefore, we look at the full list of reviewers, who provided evaluations during a specific year, as reported in the Hydrogen Program Annual Merit Review and Peer Evaluation reports. Examining this data, we find that out of the 2569 reviewers who evaluated projects between FY2003 and FY2017 in the Hydrogen Program, 357 (13.90%) were affiliated to a university[10]. Given this relatively low proportion of university researchers, we can rule out the possibility that the negative association between earmarks and project evaluation scores is mainly driven by the fact that most peer reviewers have a university background.

[10] For this analysis, we did not deduplicate reviewer names. This means that, if a person is a peer reviewer in FY2005, FY2008 and FY2010, this person is recorded as three separate reviewers in the total count. In other words, the number of unique reviewers in the program between FY2003 and FY2017 is lower than 2569. For the FY2009 VTO report, we could not identify the list of reviewers that provided evaluations to projects.

4. Discussion

Both academics and policy makers tend to assume that, to generate high-quality projects from public funding, competitive, rather than political, selection processes should be used (Boyle & Matheson, 2009; Doyle, 2011). The results of our paper, however, challenge this assumption. While competitively selected projects receive higher peer review scores, we do not find that politically selected projects consistently underperform competitively selected ones in terms of research and scientific output. Below, we discuss the implications of our findings for (1) the discussion on the potential existence of biases in peer reviews of funded projects, and (2) the ongoing policy debate on the relevance of earmarking. Subsequently, we highlight the core limitations of our research and suggest interesting avenues for future research.

4.1. Biases in the evaluation of selected R&D projects

Given the importance of peer review in competitive selection processes, existing research on allocation of R&D funding has paid substantial attention to understanding the conditions that shape evaluations of peer reviewers (Marsh et al., 2008). This research stream highlights that experts can be biased in their evaluations of R&D projects (Bornmann, 2011; Lee et al., 2013). For instance, there is evidence that experts tend to prefer proposals that are closer to their own area of expertise (Boudreau et al., 2016; Braun, 1998; Li, 2017) and that are coming from already-successful (Bol et al., 2018; Langfeldt, 2001) and familiar (Marsh et al., 2007; Sandström & Hällsten, 2008) actors. In addition, recent evidence shows that, when evaluators have to score different proposals, the position of a particular proposal in the order of evaluations is likely to influence the final score (Criscuolo et al., 2021; Elhorst & Faems, 2021).

The research on potential biases in the evaluation of proposals mainly focuses on the initial selection of projects, where the decision needs to be made whether a project gets funding or not. In this study, however, we collected data on the evaluations of reviewers regarding ongoing projects. We found clear indications that the scoring of reviewers of ongoing projects is influenced by how a project was selected for funding (i.e. earmarked versus non-earmarked). Despite finding evidence that earmarked projects did not consistently perform better or worse relative to non-earmarked ones in terms of scientific and research outputs, we observed that reviewers gave significantly lower evaluation scores to earmarked projects. In additional analyses, we found indications that peer reviewers have a form of bias toward earmarked projects, providing them with feedback that is more verbose and less improvement focused. Our findings therefore provide some evidence that peer review evaluation biases are not restricted to the initial selection of R&D projects in funding settings but are also present is subsequent project evaluations after the initial selection is made. This finding also points to the importance of combining different outcome indicators when evaluating the quality and impact of projects funded through R&D grants. Solely relying on one outcome indicator, such as project evaluation scores, might give an incomplete perspective on the performance of projects and their selection approaches.

4.2. Should earmarking be banned?

Our findings also contribute to the ongoing debate on earmarking of public R&D funds. Earmarks have often been criticized for representing wasteful spending of scarce public funds (Dickerson, 2021; Doyle, 2011). Following mounting criticism of this practice, U.S. Congress decided on an earmark moratorium in 2011. One of the main arguments to support this moratorium was the assumed underperformance of earmarked projects. At the same time, it needs to be noted that critics of earmarks tend to exclusively rely on anecdotal evidence to support this underperformance assumption, highlighting examples of "bad" earmarks such as the widely publicized "bridge to nowhere", "cowgirl museum", and "bear DNA" projects (Crespin et al., 2009; Frisch & Kelly, 2011).

In this study, we leveraged a particular setting (i.e. the DOE's Hydrogen Program) and combined inputs from different databases to adequately identify which projects were earmarked. Moreover, by collecting data on different outcome indicators for each project, we were able to systematically compare the performance of earmarked and non-earmarked projects. Our findings demonstrate that, while earmarked R&D projects receive lower peer evaluation scores and are associated to fewer scientific spillovers than non-earmarked projects, there is some evidence that they outperform them in terms of scientific productivity and research novelty. These findings suggest that we cannot simply assume that earmarked projects systematically underperform compared to non-earmarked ones.

It is important to highlight that our findings do not automatically question the decision to ban earmarks. Although we could not find a consistent underperformance of earmarks in terms of research and scientific output, there might be other reasons why societies in general, and policy makers in particular, prefer to stay away from politically driven selection processes. For instance, an argument could be made that, despite the administrative costs and potential biases that are related to competitive selection processes, they still provide a more transparent selection process than political ones. In this way, competitive selection processes can be seen as an important part of the fiduciary responsibility of governments when redistributing public resources to particular actors. Moreover, there are strong indications to suggest that earmarks have been applied to "grease the wheels" in politics (Lazarus, 2009), implying that members of Congress would sometimes use them to obtain leverage in political negotiations. In addition, cases have been reported of organizations paying bribes to members of Congress in exchange for earmarked funds (Kunz & O'Leary, 2012). Even when such politically selected projects do not systematically underperform in terms of output, we need to consider whether a selection approach, which is sensitive to abuse and corruption, is acceptable from an ethical perspective. In sum, important procedural and ethical arguments can be brought forward to defend a ban on earmarking. However, we claim that, when actors want to defend such a ban on earmarking, they need to rely on the appropriate argumentation and cannot simply assume that politically selected projects always underperform competitively selected ones.

4.3. Future research and limitations

This study has limitations that can spur interesting future research endeavors. First, whereas we use highly standardized data from a relevant empirical setting, it would be interesting to examine the influence of earmarks on the performance of projects in settings that are different from the Hydrogen Program in terms of involved organizations, project goals, and research outputs. Moreover, while the type of earmark that we examine in the current study is common in different U.S. government agencies (like the Department of Transportation, Department of Defense, Department of Agriculture), it is worth noting that in some agencies like the National Institute of Health (NIH), other types of earmarks are also used Hegde and Sampat (2015, p. 2282), for instance, focus on so-called soft earmarks at the NIH which are "[specified in] language that "urges" and "encourages" the NIH to support research on particular diseases […] but do not have the formal force of law". Examining the performance outcomes of different types of earmarks represents an interesting avenue for future research.

Second, we cannot exclude the possibility that earmarked projects have lower quality in other dimensions that are more difficult to measure at the project level, such as job creation (Lanahan et al., 2021), interorganizational collaboration (Bianchi et al., 2019), and the commercialization of novel products and services (Choi & Lee, 2017). Additional research, comparing competitively and politically selected projects on other output dimensions, would therefore be very relevant.

Third, data on peer reviewers in our empirical setting were anonymized, making it impossible to pinpoint which reviewers were associated to which project evaluation. It would be valuable to explore in more detail why evaluations of earmarked and non-earmarked projects differ depending on the characteristics of peer reviewers. One avenue that could be further explored is whether peer reviewers with a university background evaluate earmarked projects differently than those with a different background. Moreover, it would be interesting to examine whether peer reviewers that had previously been involved in selecting non-earmarked projects provide different evaluations than those that have not.

Finally, it would be interesting to compare politically driven selection processes with other selection processes besides the competitive one. For example, scholars have discussed alternative approaches such as a lottery system where a subset of pre-filtered projects would be funded at random (e.g Roumbanis, 2019), or an egalitarian system where funding is allocated in equal amounts to a particular subset of organizations (e.g Ioannidis, 2011). Comparing such alternative approaches with politically driven selection processes has the potential to significantly increase our understanding of how to effectively allocate scarce public funding through R&D grants.

CRediT authorship contribution statement

Holmer Kok: Conceptualization, Writing – original draft, Writing – review & editing, Methodology, Formal analysis, Investigation. **Dries Faems:** Conceptualization, Writing – original draft, Writing – review & editing. **Pedro de Faria:** Conceptualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful to the editor Ammon Salter and three anonymous reviewers for their helpful comments and guidance. We thank Franziska Roehnelt and Loucmane Sawadogo for research assistance. Earlier versions of this paper were presented during seminars at the House of Innovation at the Stockholm School of Economics, University of Groningen, Research Institute of Industrial Economics (IFN), University of Liverpool, VU Amsterdam, Rotterdam School of Management, and KU Leuven.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.respol.2022.104514.

References

Abramo, G., D'Angelo, C.A., Di Costa, F., 2020. Does the geographic proximity effect on knowledge spillovers vary across research fields? Scientometrics 123, 1021–1036.
Arts, S., Hou, J., Gomez, J.C, 2021. Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. Res. Policy 50, 104144.
Aschhoff, B., Sofka, W., 2009. Innovation on demand—Can public procurement drive market success of innovations? Res. Policy 38, 1235–1247.
Ayoubi, C., Pezzoni, M., Visentin, F., 2019. The important thing is not to win, it is to take part: What if scientists benefit from participating in research grant competitions? Res. Policy 48, 84–97.
Azoulay, P., Graff Zivin, J.S., Manso, G., 2011. Incentives and creativity: Evidence from the academic life sciences. Rand J. Econ. 42, 527–554.
Becker, B., 2015. Public R&D policies and private R&D investment: A survey of the empirical evidence. J. Econ. Surv. 29, 917–942.
Bianchi, M., Murtinu, S., Scalera, V.G, 2019. R&D subsidies as dual signals in technological collaborations. Res. Policy 48, 103821.

Bloom, N., Griffith, R., Van Reenen, J., 2002. Do R&D tax credits work? Evidence from a panel of countries 1979–1997. J. Public Econ. 85, 1–31.
Bloom, N., Van Reenen, J., Williams, H., 2019. A toolkit of policies to promote innovation. J. Econ. Perspect. 33, 163–184.
Bol, T., de Vaan, M., van de Rijt, A., 2018. The Matthew effect in science funding. Proc. Natl. Acad. Sci. 115, 4887–4890.
Bonaccorsi, A., Vargas, J., 2010. Proliferation dynamics in new sciences. Res. Policy 39, 1034–1050.
Bornmann, L., 2011. Scientific peer review. Annu. Rev. Inf. Sci. Technol. 45, 197–245.
Bornmann, L., Mutz, R., Daniel, H.D, 2007. Gender differences in grant peer review: A meta-analysis. J. Informetrics 1, 226–238.
Bornmann, L., Wallon, G., Ledin, A., 2008. Does the committee peer review select the best applicants for funding? An investigation of the selection process for two European molecular biology organization programmes. PLoS One 3, e3480.
Boudreau, K.J., Guinan, E.C., Lakhani, K.R., Riedl, C., 2016. Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. Manage. Sci. 62, 2765–2783.
Boyle, M.A., Matheson, V.A, 2009. Determinants of the distribution of congressional earmarks across states. Econ. Lett. 104, 63–65.
Brach, A., Wachs, M., 2005. Earmarking in the US department of transportation research programs. Transp. Res. Part A: Policy Practice 39, 501–521.
Braun, D., 1998. The role of funding agencies in the cognitive development of science. Res. Policy 27, 807–821.
Brunner, J., 2021. Washington state members of Congress are all-in for the return of earmarks. The Seattle Times. https://www.seattletimes.com/seattle-news/politics/washington-members-of-congress-are-all-in-for-the-return-of-earmarks/.
Choi, J., Lee, J., 2017. Repairing the R&D market failure: Public R&D subsidy and the composition of private R&D. Res. Policy 46, 1465–1478.
Corredoira, R.A., Goldfarb, B.D., Shi, Y., 2018. Federal funding and the rate and direction of inventive activity. Res. Policy 47, 1777–1800.
Crespin, M.H., Finocchiaro, C.J., Wanless, E.O, 2009. Perception and reality in congressional earmarks. The Forum 7, 1–16.
Crilly, D., Hansen, M., Zollo, M., 2016. The grammar of decoupling: A cognitive-linguistic perspective on firms' sustainability claims and stakeholders' interpretation. Acad. Manag. J. 59, 705–729.
Criscuolo, P., Dahlander, L., Grohsjean, T., Salter, A., 2021. The sequence effect in panel decisions: Evidence from the evaluation of research and development projects. Org. Sci. 32, 987–1008.
Czarnitzki, D., Hanel, P., Rosa, J.M, 2011. Evaluating the impact of R&D tax credits on innovation: A microeconometric study on Canadian firms. Res. Policy 40, 217–229.
Czarnitzki, D., Hünermund, P., Moshgbar, N., 2018. Public procurement as policy instrument for innovation. ZEW-Centre for Eur. Econ. Res. Discussion Paper 18-001, 18-001.
De Figueiredo, J.M., Silverman, B.S, 2006. Academic earmarks and the returns to lobbying. J. Law Econ. 49, 597–625.
De Figueiredo, J.M, Silverman, B.S, 2007. How does the government (want to) fund science? Politics, lobbying and academic earmarks. NBER Working Paper, 13459.
De Rassenfosse, G., Jaffe, A., Raiteri, E., 2019. The procurement of innovation by the US government. PLoS One 14, e0218927.
Dickerson, M.D., 2021. Return of the Swamp: Earmarks Would be a Costly Mistake. The Heritage Foundation Backgrounder No. 3602. Accessed January 20, 2022. https://www.heritage.org/sites/default/files/2021-03/BG3602.pdf. (Accessed 20 January 2022).
Doyle, R.B., 2011. The rise and (relative) fall of earmarks: Congress and reform, 2006–2010. Public Budget. Finance 31, 1–22.
Du, J., Leten, B., Vanhaverbeke, W., 2014. Managing open innovation projects with science-based and market-based partners. Res. Policy 43, 828–840.
Ebersberger, B., 2005. The impact of public R&D funding. VTT technical research centre of Finland, 588. VTT Publications, Finland.
Elhorst, P., Faems, D., 2021. Evaluating proposals in innovation contests: Exploring negative scoring spillovers in the absence of a strict evaluation sequence. Res. Policy 50, 104198.
Engstrom, E.J., Vanberg, G., 2010. Assessing the allocation of pork: Evidence from congressional earmarks. Am. Polit. Res. 38, 959–985.
Fang, F.C., Casadevall, A., 2016. Research funding: The case for a modified lottery. MBio 7, e00422-16. -16.
Feldman, M.P., Kelley, M.R, 2006. The ex ante assessment of knowledge spillovers: Government R&D policy, economic incentives and private firm behavior. Res. Policy 35, 1509–1521.
Finnigan, T., 2007. All about pork: the abuse of earmarks and the needed reforms. Citizens Against Govt. Waste. Accessed January 20, 2022. https://www.cagw.org/sites/default/files/pdf/2007/all_about_pork_2007_final.pdf. (Accessed 20 January 2022).
Fleming, L., 2001. Recombinant uncertainty in technological search. Manage. Sci. 47, 117–132.
Fleming, L., Greene, H., Li, G., Marx, M., Yao, D., 2019. Government-funded research increasingly fuels innovation. Science 364, 1139–1141.
Frisch, S.A., Kelly, S.Q, 2011. Cheese factories on the moon: Why earmarks are good for American democracy, 1st ed. Paradigm Publishers, Boulder, CO.
Ganguli, I., 2017. Saving Soviet science: the impact of grants when government R&D funding disappears. Am. Econ. J.: Appl. Econ. 9, 165–201.
Ghisetti, C., 2017. Demand-pull and environmental innovations: Estimating the effects of innovative public procurement. Technol. Forecast. Social Change 125, 178–187.
Ginther, D.K., Heggeness, M.L, 2020. Administrative discretion in scientific funding: Evidence from a prestigious postdoctoral training program. Res. Policy 49, 103953.
Ginther, D.K., Schaffer, W.T., Schnell, J., Masimore, B., Liu, F., Haak, L.L., Kington, R., 2011. Race, ethnicity, and NIH research awards. Science 333, 1015–1019.
Gittelman, M., Kogut, B., 2003. Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. Manage. Sci. 49, 366–382.
Goldstein, A.P., Kearney, M., 2020. Know when to fold 'em: An empirical description of risk management in public research funding. Res. Policy 49, 103873.
Goldstein, A.P., Narayanamurti, V., 2018. Simultaneous pursuit of discovery and invention in the US Department of Energy. Res. Policy 47, 1505–1512.
Grodal, S., O'Mahony, S, 2017. How does a grand challenge become displaced? Explaining the duality of field mobilization Acad. Manag. J. 60, 1801–1827.
Guerzoni, M., Raiteri, E., 2015. Demand-side vs. supply-side technology policies: Hidden treatment and new empirical evidence on the policy mix. Res. Policy 44, 726–747.
Hayden, E., 2015. Racial bias continues to haunt NIH grants. Nature News 527, 286.
Hegde, D., Sampat, B., 2015. Can private money buy public science? Disease group lobbying and federal funding for biomedical research. Manage. Sci. 61, 2281–2298.
Heinze, T., 2008. How to sponsor ground-breaking research: A comparison of funding schemes. Sci. Public Policy 35, 302–318.
Howell, S.T., 2017. Financing innovation: Evidence from R&D grants. Am. Econ. Rev. 107, 1136–1164.
Hyytinen, A., Toivanen, O., 2005. Do financial constraints hold back innovation and growth? Evidence on the role of public policy. Res. Policy 34, 1385–1403.
Iansiti, M., 1995. Shooting the rapids: Managing product development in turbulent environments. Calif. Manage. Rev. 38, 37–58.
Ioannidis, J.P.A., 2011. Fund people not projects. Nature 477, 529–531.
Jacob, B.A., Lefgren, L., 2011. The impact of research grant funding on scientific productivity. J. Public Econ. 95, 1168–1177.
Jaffe, A.B., 1998. The importance of "spillovers" in the policy mission of the advanced technology program. J. Technol. Transf. 23, 11–19.
Jaffe, A.B., de Rassenfosse, G., 2017. Patent citation data in social science research: Overview and best practices. J. Assoc. Inf. Sci. Technol. 68, 1360–1374.
Knoll, B., Riedel, N., Schwab, T., Todtenhaupt, M., Voget, J., 2021. Cross-border effects of R&D tax incentives. Res. Policy 50, 104326.
Kunz, K., O'Leary, S, 2012. The importance of federal earmarks to state coffers: An examination of distribution trends over the decade. J. Public Budget. Account. Financ. Manage. 24, 579–608.
Lanahan, L., Joshi, A.M., Johnson, E., 2021. Do public R&D subsidies produce jobs? Evidence from the SBIR/STTR program. Res. Policy 50, 104286.
Langfeldt, L., 2001. The decision-making constraints and processes of grant peer review, and their effects on the review outcome. Social Stud. Sci. 31, 820–841.
Law, M.T., Tonon, J.M., Miller, G.J, 2008. Earmarked: The political economy of agricultural research appropriations. Appl. Econ. Perspect. Policy 30, 194–213.
Lawler, A., 2000. Earmarks, rising costs threaten NASA missions. Science 289, 2018-2018.
Lazarus, J., 2009. Party, electoral vulnerability, and earmarks in the US House of Representatives. J. Politics 71, 1050–1061.
Lazarus, J., 2010. Giving the people what they want? The distribution of earmarks in the US house of representatives. Am. J. Political Sci. 54, 338–353.
Lee, C.J., Sugimoto, C.R., Zhang, G., Cronin, B., 2013. Bias in peer review. J. Am. Soc. Inf. Sci. Technol. 64, 2–17.
Li, D., 2017. Expertise versus Bias in Evaluation: Evidence from the NIH. Am. Econ. J.: Appl. Econ. 9, 60–92.
Li, D., Agha, L., 2015. Big names or big ideas: Do peer-review panels select the best science proposals? Science 348, 434–438.
Ma, A., Mondragón, R.J., Latora, V., 2015. Anatomy of funded research in science. Proc. Natl. Acad. Sci. 112, 14760–14765.
Manso, G., 2011. Motivating innovation. J. Finance 66, 1823–1860.
Marsh, H.W., Bond, N.W., Jayasinghe, U.W., 2007. Peer review process: Assessments by applicant-nominated referees are biased, inflated, unreliable and invalid. Australian Psychol. 42, 33–38.
Marsh, H.W., Jayasinghe, U.W., Bond, N.W, 2008. Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability. Am. Psychol. 63, 160.
Nicholson, J.M., Ioannidis, J.P., 2012. Conform and be funded. Nature 492, 34–36.
Office of Energy Efficiency and Renewable Energy., 2004. Peer Review Guide Based on a Survey of Best Practices for In-Progress Peer Review. https://www1.eere.energy.gov/analysis/pdfs/2004peerreviewguide.pdf (Accessed January 20, 2022).
Payne, A.A., 2002. Do US Congressional earmarks increase research output at universities? Sci. Public Policy 29, 314–330.
Raiteri, E., 2018. Evaluating the impact of public procurement on technological generality through patent data. Res. Policy 47, 936–952.
Reinhart, M., 2009. Peer review of grant applications in biology and medicine. Reliability, fairness, and validity. Scientometrics 81, 789–809.
Roumbanis, L., 2019. Peer review or lottery? A critical analysis of two different forms of decision-making mechanisms for allocation of research grants. Science Sci. Technol. Hum. Values 44, 994–1019.
Sandström, U., Hällsten, M., 2008. Persistent nepotism in peer-review. Scientometrics 74, 175–189.
Sciara, G.C., 2012. Financing congressional earmarks: Implications for transport policy and planning. Transp. Res. Part A: Policy Practice 46, 1328–1342.
Sharaf, O.Z., Orhan, M.F, 2014. An overview of fuel cell technology: Fundamentals and applications. Renewable Sustainable Energy Rev. 32, 810–853.
Sissine, F. J. 2006. DOE Budget Earmarks: A Selective Look at Energy Efficiency and Renewable Energy R&D Programs. https://www.everycrsreport.com/reports/RL33294.html. (Accessed 20 January 2022).

Silber, J., Teich, A., Nelson, S., Lita, S., 2002. Earmarking: The expansion of excellence in scientific research. AAAS science and technology policy yearbook, edited by Albert H. Teich, Stephen D. Nelson, and Stephen J. Lita. AAAS, Washington, DC, pp. 105–113.

Stephan, P., Veugelers, R., Wang, J., 2017. Reviewers are blinkered by bibliometrics. Nature News 544, 411.

The New York Time, 2020. Republicans and Democrats Need to Work Together. Earmarks Can Help. The New York Times. https://www.nytimes.com/2020/11/29/opinion/republicans-democrats-earmarks.html.

U.S. Department of Energy, 2005. DOE Hydrogen Program: 2005 Annual Merit Review and Peer Evaluation Report. https://www.hydrogen.energy.gov/annual-review/annual_review05_report.html (Accessed January 20, 2022).

Van den Besselaar, P., Sandström, U., Schiffbaenker, H., 2018. Studying grant decision-making: A linguistic analysis of review reports. Scientometrics 117, 313–329.

Wang, J., Lee, Y.N., Walsh, J.P, 2018. Funding model and creativity in science: Competitive versus block funding and status contingency effects. Res. Policy 47, 1070–1083.

Wennerås, C., Wold, A., 1997. Nepotism and sexism in peer-review. Nature 387, 341–343.