Evaluation in Research Funding Agencies: Are Structurally Diverse Teams Biased Against?

Albert Banal-Estañol
Inés Macho-Stadler
David Pérez-Castrillo

This version: April 2019
March 2016

Barcelona GSE Working Paper Series

Working Paper nº 890

Evaluation in Research Funding Agencies: Are Structurally Diverse Teams Biased Against?

Albert Banal-Estañol,* Inés Macho-Stadler[†] and David Pérez-Castrillo[‡]

April 2019

Abstract

We analyze whether funding bodies are biased against diverse teams, which have often been linked to the production of transformative research. We develop a general framework that compares the drivers of success in the ex-ante grant decision process to the drivers of success in ex-post performance. We use our framework to systematically analyze the decisions of one of the major public funding organizations for scientific research worldwide, the UK's Engineering and Physical Sciences Research Council (EPSRC). We find that structurally diverse teams are not only penalized but are also biased against. Indeed, although teams that exhibit greater diversity in knowledge and skills, education, and/or scientific ability, are significantly less likely to obtain funding, they are generally more likely to be successful. Our mediating effects show that the evidence of a bias against diversity is weaker for teams led by prestigious researchers.

JEL Classification numbers: O32, I23

Keywords: Funding organization, diversity, bias, transformative research, research grant

*Corresponding author: Universitat Pompeu Fabra, Barcelona GSE and City, University of London. Ramon Trias Fargas, 23-25, 08003 Barcelona. Email: albert.banalestanol@upf.edu.

[†]Universitat Autonoma de Barcelona and Barcelona GSE. Email: ines.macho@uab.es.

[‡]Universitat Autonoma de Barcelona and Barcelona GSE. Email: david.perez@uab.es.

1 Introduction

Radical innovations are widely understood to be the main engines of technological progress and economic growth. Public agencies for research and innovation worldwide then aim to support people, firms and organizations that search for breakthroughs (Branscomb and Auerswald, 2002). Funding agencies for scientific research such as the National Science Foundation (NSF) or the UK Research Councils, for instance, strive to provide support to individual or teams of researchers that conduct "transformative" or "frontier" research—research that holds the potential to radically change our knowledge and understanding of current science and engineering concepts, thus having "an impact in an area of research much greater in magnitude than might normally be expected" (National Academies of Science, Engineering, and Medicine, 2015, pp. 9).[1]

Diversity is often claimed to be a crucial condition for radical innovation (Nelson and Winter 1982; Fleming, 2007). Indeed, it is often argued that research and innovation tends to evolve as a cumulative process (Nerkar, 2003; Carnabuci and Bruggeman, 2009). To break away from existing trajectories and prompt breakthroughs, it may then be necessary to incorporate and combine researchers of different fields of knowledge, skills, or abilities in the same team (Fleming, 2001; Arts and Veugelers, 2014). According to this argument, "structurally" diverse teams should be more likely to spark new insights and produce breakthroughs (Guimerà et al., 2005; Jones et al., 2008; Jones, 2009).

Despite its potential positive effects, many commentators claim that public research and innovation agencies are biased against (structural) diversity (Langfeldt, 2006; Laudel, 2006). Critics argue that diversity is penalized in the evaluation and award process because diverse teams are more difficult to evaluate (Lamont 2009), or because proposals from diverse teams are perceived as being less "safe" or less "doable" (Luukkonen 2012). Unfortunately, we have very little systematic evidence as to whether research and innovation agencies are actually biased: (i) do they indeed penalize diverse teams and (ii) are diverse teams more likely to produce high-impact research?

This paper analyzes whether funding agencies may be biased against diverse teams. Funding agencies are ideal for analyzing the general stance toward diversity because we (and evaluators) know so much about the attributes of the applicants. We develop a general framework that compares the drivers of success in the ex-ante evaluation and award process to the drivers of success in ex-post performance. We argue that, absent biases, funding agencies should be, ceteris paribus, more (less) likely to provide funding to teams with a certain attribute if such teams have a higher (lower) likelihood of generating high-impact research. If, for instance, diversity of knowledge increases the likelihood of generating high-impact research, as some of the literature suggests, then agencies should be more lenient toward teams that exhibit greater diversity of knowledge. Instead, the agency would be biased against diversity of knowledge if these teams have a higher likelihood

[1] The term "transformative research" has been used by the NSF. At the National Institutes of Health (NIH) the phrase is sometimes rendered as "translational research." Within the European Research Council (ERC), the term "frontier research" is often used.

of success but a lower likelihood of being funded.

We use our general framework to determine whether funding agencies are biased against structurally diverse teams. Following prior literature, we characterize structural diversity as differences in job-related attributes, such as knowledge and skills, educational background, and scientific ability.[2] The diversity literature argues that structural diversity has a positive effect in ex-post performance (Williams and O'Reilly, 1998). In contrast, in terms of evaluation, structural diversity may be systematically penalized by funding organizations. The literature suggests that applications from diverse teams may be more complex to evaluate and/or be (or perceived to be) less "safe" or less "doable" than applications from homogeneous teams. Decision-makers, for instance, may not identify fruitful combinations of fields of knowledge, research practices, or research cultures. They may also penalize projects of a team of diverse researchers working in different areas because they may view them as more likely to fail. This leads us to hypothesize that funding agencies may be biased against diverse teams.

We test our predictions on the award decisions of one of the major public funding organizations for scientific research worldwide, the UK's Engineering and Physical Sciences Research Council (EPSRC). We constructed a novel dataset that overlaps all the EPSRC applications, funded or not funded, with the calendar census of all the engineering departments of 40 major UK universities between 1991 and 2007 (Banal-Estañol et al., 2015).[3] We make use of publication and education data to construct variables that proxy for the diversity of the applicant teams.

We find that diverse teams are not only penalized but are also biased against. Indeed, although teams that exhibit greater diversity in knowledge and skills, education, and/or ability, are significantly less likely to obtain funding, they are generally more likely to be successful. This is true independently if we measure diversity at the "individual" level or at the "team" level nor if we measure the individual levels of diversity of the team leader (the principal investigator) or of the whole team (also including the coinvestigators). As discussed in previous literature, funding agencies may be biased against diversity because applications of diverse teams are more difficult to evaluate or because they are perceived as being less "safe" or less "doable." Consistent with both explanations, we show that the prestige of the Principal Investigator mitigates the negative impact of diversity in the award decision but it does not have a significant effect in mitigating or amplifying the positive effect of the team's diversity on ex-post performance.

Our approach in identifying funding agencies' biases against diversity differs, in terms of both method and objectives, from those that recently appeared in the literature.[4] Boudreau et al.

[2] This is in contrast to the concept of demographic diversity that refers to differences in readily detectable attributes such as gender, age, or race.

[3] Banal-Estañol et al. (2013; 2018) used part of this dataset to analyze endogeneous collaboration patterns between academics and firms and the consequences, in terms of publication activity, of these collaborations.

[4] More generally, other researchers have questioned funding organizations and the evaluation and award process (Viner et al., 2004). Some suggest that the outcome distribution is not wholly meritocratic (Wenneras and Wold, 1997; Hegde and Mowery, 2008). Grimpe (2012), for example, shows that obtaining a government grant is influenced not by scientist productivity but by other personal attributes, and by institutional and discipline characteristics.

(2016) use a randomized double-blind review process of a seed grant to identify biases against novel projects rather than against the diversity of individuals and/or teams.[5] Li (2017) uses applications on nearly completed research, which is published independently if it is funded, to identify reviewer biases in favor of applicants whose work is related to their own.[6] As our aim is to study the attitude of grantor agencies toward team diversity, we focus on grants that can also be submitted by teams (as opposed to Azoulay et al., 2011, Boudreau et al., 2016, and Li, 2017, who use applications submitted by individuals only).[7]

2 Framework

This section develops a framework to analyze whether funding agencies are biased against, or in favor of, any personal attribute of the pool of applicants. We first motivate and then provide a formal definition of what we think constitutes a "bias." Our definition distinguishes between penalizing and being biased against, and between rewarding and being biased in favor of, a certain attribute. We then explain the empirical approach we use to identify these biases.

2.1 Funding agencies' objectives

Funding agencies around the world profess to allocating academic research funding on the basis of scientific merit/excellence. The EPSRC, for instance, states in its evaluation criteria that "Research excellence will always be preeminent."[8] This is of course because funding agencies need to make the

[5] Project novelty is measured by comparing the pairs of keywords that could occur in an R&D proposal with the pairwise combinations that have appeared in previous projects. Criscuolo et al. (2017) use a similar procedure to show that organizations are more likely to select projects that are characterized by intermediate levels of novelty. Similarly, Mueller et al. (2011) suggest that even people who claim to be highly motivated to achieve novel outcomes are often biaised against ideas that depart from existing ways of doing things because of their inherent uncertainty. Despite its advantages, our database does not allow us to measure project novelty and separate project novelty from team diversity.

[6] To construct a measure of ex-post performance, we trace back the research output of each specific application team in each particular project, and not only the overall performance of the applicants. We make use of a recent development in the publication databases that include the funding sources of each article. Like Ferguson and Carnabuci (2017) in a study of patent evaluations, we take into account that we can construct this measure only for funded applications in the econometric method.

[7] As in Boudreau et al. (2016) and Li (2017), our focus is on the award decisions themselves rather than on the effects of the grant program. Azoulay et al. (2011) showed that researchers supported by funding bodies that tolerate early failure, reward long-term success, and do not limit freedom, such as the Howard Hughes Medical Institute, are more likely to produce breakthroughs than comparable grantees from the NIH, which has short review cycles, pre-defined deliverables, and renewal policies that are unforgiving of failure. Jacob and Lefgren (2011) find that receipt of an NIH grant only leads to about one additional publication over the following five years.

[8] The ERC's mission statement asserts that "The ERC's mission is to encourage the highest quality research in Europe through competitive funding and to support investigator-driven frontier research across all fields, on the basis of scientific excellence." In its objectives, Norway's research council states that "Grants are awarded on the basis of the scientific merit." See https://www.epsrc.ac.uk/newsevents/pubs/standard-calls-reviewer-helptext, https://erc.europa.eu/about-

best use of their scarce funds to generate high-quality, high-impact output (Tijssen et al., 2002). Many funding agencies even call for "transformative" research, which may have "an impact on an area of research much greater in magnitude than might normally be expected" (National Academies of Science, Engineering, and Medicine, 2015, pp. 9). The NSF, for example, has recently included an emphasis on "potentially transformative research" in its merit review criteria.

Some critics, though, argue that funding decisions are not based solely on scientific merit, but are also influenced by the personal attributes of the applicants. Bornmann and Daniel (2005) found, for instance, that conditional on prior scientific achievements, institutional affiliation and field of study affect the predicted probability of approval in fellowship applications. This may result, the critics claim, in little or no relationship between award decisions and the "true quality" of the applications being evaluated (Blackburn and Hakal, 2006; Laudel, 2006). In practice, this means that the applications funded are not those that would have had the highest ex-post performance, measured, for instance, in terms of quantity and/or quality of the ensuing publications (Li, 2017).

Methodologically, most prior research has analyzed whether a given attribute, such as institutional affiliation or research field, explains the award decision in regressions that control for other characteristics (Bornmann and Daniel, 2005). Yet, it is not always clear that all of these attributes do not have an influence on project output. Applicants from top institutions, for instance, may be, ceteris paribus, more likely to produce a higher quantity and quality of publications than those based at lower-ranked institutions.

We shall argue that if an attribute is related to ex-post performance, and a funding agency does anticipate it and takes it into account in the award process, this agency cannot be considered "biased." Indeed, we cannot call biased an agency that, ceteris paribus, is more likely to fund proposals of teams of applicants with greater past scientific performance, because past performance is likely to be correlated with future performance. More generally, we argue that agencies need to consider the effect of each of the attributes on ex-post performance. If a certain attribute increases (decreases) the likelihood of success then the agency should be more lenient (strict) toward proposals with this attribute. For instance, if applicants from top institutions are, ceteris paribus, likely to perform better than those based at lower-ranked institutions then the evaluation process should, ceteris paribus, positively discriminate in favor of researchers at top institutions.

2.2 Formal definition of bias

To formalize our discussion, let us denote the (unobserved) ex-post quality of the research results of project i by $g^*(Y_{i1}, ..., Y_{ik}, Z_i)$, where $Y_{i1}, ..., Y_{ik}$ represent characteristics of the team of applicants, such as their average past scientific performance, the quality of the institution they work for, or the diversity of some of their attributes, and Z_i is a vector of other factors. We call the project "successful" if and only if $g^*(\cdot)$ is above a certain threshold. The threshold could be the minimum quality required for the ensuing research results to be published in a scientific journal, or for the

erc/mission, and http://www.forskningsradet.no/en/Funding/FRINATEK/1254025689182.

publications emanating from the project to reach a certain number of citations or be among the top-cited papers. A higher emphasis on transformative research would naturally imply a higher threshold. For any given measure of success, a certain attribute j improves the quality of the research results of the project, and therefore its likelihood of success, if and only if $\partial g^*/\partial Y_{ij} > 0$.[9]

In the same vein, we denote the (unobserved) ex-ante value that the funding agency attaches to project i by $f^*(Y_{i1}, ..., Y_{ik}, Z_i')$, where Z_i' includes other factors, possibly different from those in Z_i. The value $f^*(\cdot)$ may take into account, but it is not necessarily identical to, the expected quality of the research results $g^*(\cdot)$. Indeed, the agency may discount or reward certain attributes independently of their effect on expected quality. As we shall argue below, agencies may for instance penalize diversity because of an aversion to risk. The agency shall fund proposal i if and only if $f^*(\cdot)$ is above a threshold. The threshold arises, for instance, because the agency maximizes the value of the projects it funds subject to a budget constraint. A certain attribute j enhances the value of the proposal for the funding agency, and therefore the likelihood of being funded, if and only if $\partial f^*/\partial Y_{ij} > 0$.

We call the agency "biased against" a certain attribute if proposals with this attribute have a higher likelihood of success but a lower likelihood of being funded, i.e., if $\partial g^*/\partial Y_{ij} > 0$ and $\partial f^*/\partial Y_{ij} < 0$. If team diversity, for instance, enhances the likelihood of success in the publication process, and therefore the quality of research results, then an unbiased agency that rewards scientific excellence should be more lenient toward diverse teams. Similarly, we consider an agency "biased in favor" of attribute j if proposals with this attribute have a lower likelihood of success but a higher likelihood of being funded, i.e., if $\partial g^*/\partial Y_{ij} < 0$ and $\partial f^*/\partial Y_{ij} > 0$. In our definition we allow for the possibility of the funding agency penalizing but not being biased against a certain attribute ($\partial g^*/\partial Y_{ij} < 0$ and $\partial f^*/\partial Y_{ij} \leq 0$) as well as for the possibility of rewarding but not being biased in favor of this attribute ($\partial g^*/\partial Y_{ij} > 0$ and $\partial f^*/\partial Y_{ij} \geq 0$).

Three comments are in order. First, our definition of bias is conservative in the sense that we are comparing the sign rather than the magnitude of the effects. Indeed, to have a bias against a certain characteristic, we require the effect in the ex-ante funding decision to pull in the opposite direction to that of the ex-post success. Econometrically, we will call it "strong" evidence of a bias if both opposing effects are significantly different from zero. If one of the two effects is significantly different from zero but the other is not, we call it "weak" evidence of a bias.

Second, funding decisions are of course taken under uncertainty about the quality of the research results. This will always generate mistakes ex-post. For example, a non-biased agency may be more lenient toward a proposal with a certain attribute because this attribute leads to a higher likelihood of ex-post success, on average. But it may of course be that this particular project is unsuccessful. Still, we expect random errors at an individual proposal level to be dissipated at the aggregate,

[9] Our empirical results are robust to different thresholds as measures of success. We use a dichotomous variable as a measure of success in ex-post performance to facilitate the comparison with the drivers of success in the ex-ante evaluation and award process.

provided the number of applications to be scrutinized is sizeable.

Third, our approach relies on the agency being able to observe the characteristics of the applicants to construct estimates of ex-post performance. In practice, the information available to the agencies worldwide varies. But most funding agencies, such as the EPSRC, use a single-blind peer review system of evaluation, where the applicants' identities and characteristics, in addition to their research plan, are observed by the reviewers. As shown in the peer evaluation literature, single-blind evaluation systems weigh the characteristics of the applicants much more heavily than the research plans expressed in their research proposals (Lee et al., 2000).[10] This is more the case for agencies making funding decisions than for academic journals reviewing manuscripts submitted for publication. Undoubtedly, evaluating untested ideas of research proposals is inherently more difficult than evaluating completed works submitted to a journal for publication (Porter and Rossini, 1985).

2.3 Empirical strategy

Econometrically, we shall view the value attached by the agency, as well as the quality of the research results, as two (unobserved) latent variables that linearly depend on all factors,

$$f^*(Y_{i1},...,Y_{ik}) = \alpha_1 Y_{i1} + ... + \alpha_k Y_{ik} + \gamma Z_i' + \varepsilon_i,$$

and

$$g^*(Y_{i1},...,Y_{ik}) = \beta_1 Y_{i1} + ... + \beta_k Y_{ik} + \phi Z_i + \nu_i,$$

where ε_i and ν_i are normally distributed, possibly correlated, error terms. The (observed) award variable, f_i, which takes a value of 1 if proposal i is awarded funding and a value of 0 if it is not, as well as the (observed) project success variable, g_i, which takes a value of 1 if project i is successful and a value of 0 if it is not, can be viewed as indicators for whether the latent variables are above a threshold $\underline{f}$ and $\underline{g}$, respectively, and can be estimated with a probit model (Van de Ven and Van Pragg, 1981),

$$f_i = \mathbf{1}(f^*(Y_{i1},...,Y_{ik}, Z_i') > \underline{f}),$$

and

$$g_i = \mathbf{1}(g^*(Y_{i1},...,Y_{ik}, Z_i) > \underline{g}).$$

According to our definition, in this linear probit specification, the agency is biased against a certain attribute j if $\alpha_j < 0$ and $\beta_j > 0$ and biased in favor of it if $\alpha_j > 0$ and $\beta_j < 0$. Notice we are implicitly assuming that included attributes are uncorrelated with unobserved ones. In particular, we assume that projects of applicants of a given ability, seniority, etc. have a distribution of project-potential which is independent of the observed degrees of team diversity.

[10] For example, the research track record of the applicants is typically a "critical component in evaluations of grant proposals" (Marsh et al., 2008, p. 167).

But, the main difficulty of using this approach is that we do not observe the non-funded proposals when estimating the probability of ex-post success. More importantly, there may be unobserved characteristics of the application (such as the quality of the research proposal) that influence the award decision that are correlated with the measure of success in terms of ex-post performance. As argued by Ferguson and Carnabuci (2017) in a study of patent evaluations, ignoring this type of differential selection may generate an overstatement of the effect of certain attributes on ex-post performance outcomes. Still, if we select all the right variables for our models, and leave few unobservable variables that affect our performance variable, then we may not have selection bias.

We follow a conservative approach and estimate success in ex-post performance using a two-stage econometric model that accounts for a potential differential selection. We make use of a Heckman probit selection model, which provides consistent, asymptotically efficient estimates for all the parameters. For the model to be well identified, we need at least one factor affecting the award but not the performance equation that can serve as an instrument for the exclusion restriction. We use the overall acceptance rate of all the applications in the same quarter, which should affect (and it does, empirically) the likelihood of funding of a given application but it should not affect its likelihood of success in ex-post performance directly. Here, we are implicitly assuming that the timing of submission is not strategic and, for instance, that (marginal) projects are not submitted when the probability of getting the application awarded is higher. Indeed, the applicants of the EPSRC, as well as of most funding agencies, do not know the budget nor the number of competing applications in a given quarter (although this information, at an aggregate level, may be available ex-post).[11]

3 Diversity

In this section, we analyze prior literature to develop hypotheses on the ex-post performance results of, as well as on ex-ante evaluation approaches to, our attribute of interest: diversity.

3.1 Types and measures of structural diversity

Diversity may in principle be applied to a wide number of attributes, ranging from religious to functional background. But, in practice, as stated by Van Knippenberg et al. (2004, pp. 1008), "diversity research has mainly focused on gender, age, race, tenure, educational background, and functional background." In an early review of the literature, Williams and O'Reilly (1998) suggest that the most important difference across types of diversity is between "social category"—differences in readily detectable attributes such as sex, age, and tenure—and "informational/functional diversity"—differences in less visible underlying attributes that are more

[11] An alternative to measuring the quality of unfunded grants in case applications are based on research that is already very advanced, is to use text-matching and link grant application titles with the titles and abstracts of semantically related publications (Li, 2017).

job-related, such as functional and educational background. Cummings (2004) also distinguishes between "demographic" and "structural diversity" (e.g., member differences in terms of sources of task information and know-how). Diversity of knowledge and skills has also been termed "cognitive" (or "deep-level") diversity to distinguish it from diversity in surface characteristics such as the demographic variables of age, gender, and race (Harrison et al., 1998; Shin et al., 2012; Taylor and Greve, 2006).

We focus on structural diversity, which is often viewed as a crucial condition for producing "Schumpeterian" novel combinations (Nelson and Winter, 1982; Fleming, 2007). Following the literature, we characterize (structural) diversity as (i) diversity in knowledge and skills, which in science is related to the concept of interdisciplinarity and in other settings would be related to the notion of different functional backgrounds, (ii) educational diversity, and (iii) diverse abilities.

The concept of diversity can be applied both to the individuals of the team or to the team as a whole. As argued by Bunderson and Sutcliffe (2002), diversity may be defined as a distribution across team members ("team" or "interpersonal" levels of diversity), or the extent to which the individuals who comprise the team are themselves diverse ("individual" or "intrapersonal" levels of diversity). For instance, a team may be diverse because it is composed of specialized researchers in different fields (team diversity) or because the researchers in the team are themselves interdisciplinary (individual diversity) (Wagner et al., 2011). An attribute that is underrepresented in a given group is likely to become salient and can thus also be considered a source of (individual) diversity (Kanter, 1977; Williams and O'Reilly, 1998).

When considering the extent to which the individuals of the team are themselves diverse (individual diversity), we distinguish between the Principal Investigator (PI) and the coinvestigators. Although both PIs and coinvestigators are understood to contribute to the project, the PI, as team leader, may be considered more important, both in terms of ex-ante evaluation as well as in terms of ex-post performance, than the other members of the team (Viner et al., 2004). But it is difficult to know by exactly how much. We therefore consider the two extremes and separately analyze the impact of the level of individual diversity of the PI and the average level of individual diversity of all the members of the team.

3.2 Ex-post performance

Structurally diverse teams, i.e., those that "possess a broader range of task-relevant knowledge, skills, and abilities" (Van Knippenberg et al., 2004, pp. 1009), have been argued to have a mainly positive effect on performance. Exposure to diverging views and perspectives may lead to more creative and innovative ideas and solutions (Bantel and Jackson, 1989; De Dreu and West, 2001). Diverse teams may also be more likely to have different opinions, thus raising communication costs and retarding problem-solving (Nooteboom, 1999). Still, increased relationship conflict may not necessarily lead to lower performance (Jehn et al., 1997). The need to reconcile conflicting viewpoints may force diverse teams to more thoroughly process task-relevant information and may prevent them from opting too easily for a seemingly good course of action.

We now discuss in more detail the effects of each of the three proxies of (structural) diversity.

Knowledge and skills Diversity in knowledge and skills in science is related to the notion of interdisciplinarity. In the definition of the National Academies of Science, "interdisciplinarity" is defined as "a mode of research by teams or individuals that integrates perspectives/concepts/theories and/or tools/techniques and/or information/data from two or more bodies of specialized knowledge or research practice" (Porter et al., 2007). As made clear by the definition, interdisciplinarity may be defined at the team level or at the individual level.

Several papers highlight the benefits of interdisciplinarity. Disis and Slattery (2010) argue that intellectually diverse teams, not dominated by a single view, are more likely to be successful. Hollingsworth (2007) surveys 291 major discoveries in biomedical sciences and finds that none of them occurred in a laboratory that was narrow in scope and oriented to a single discipline. The successful teams not only exhibited high levels of team diversity, but were all led by (individually) diverse directors, who had the capacity to integrate diversity and to address problems relevant to numerous fields of science.[12] Catalini (2018) finds that the labs from different fields that collaborated after a spatial reallocation were more likely to produce papers that would end up in the highest quartile of the citation distribution.[13]

Educational background The diversity research literature has stressed and documented the advantages of teams with different educational backgrounds. For instance, Smith et al. (1994) find that top management team educational diversity is positively associated with company financial performance. Dahlin et al. (2005) show that educational diversity enhances informational use. As explained by Williams and O'Reilly (1998), decisions made by groups with diverse information and perspective will be of higher quality than by groups of employees holding the same views. This diversity of information and perspective may steam from diverse educational backgrounds.

In academia, post-graduate research education is especially important because it is a key determinant of research culture. As shown by the higher education literature, "academic institutions possess distinctive cultures which are developed and sustained by identifiable actions of the community members" (Dill, 1982, pp. 304). Several papers have examined the causes and the consequences of academic culture (see e.g., Evans, 2007) and the transmission of the research cultures to the research students (Deem and Brehony, 2000). Barjak and Robinson (2008) show that academic research teams that draw on knowledge from different research cultures and nationalities are more successful (see also Bantel and Jackson, 1989). Research cultures vary, especially, across

[12] Another example of a successful interdisciplinary team is the famous institute Pasteur (Hage and Mote, 2010).

[13] Catalini (2018) also documents an increase in variance. Unfortunately, our data on ex-post performance, despite being very precise (as it refers to the results of a relatively small team that worked together on a particular research project), does not allow us to precisely estimate other parameters of the distribution of citations, such as the variance of citations, as Catalini (2018) does using a whole laboratory as a unit of observation.

institutional environments and by the degree of research intensity (Holligan et al., 2011).

Notice that educational diversity can be measured, not only at the team, but also at the individual level. Teams led, or composed of, researchers that have been educated outside a given institution, the "outsiders," may be considered salient.[14] Previous research has documented a tendency of institutions toward the exploitation of familiar knowledge (March, 1991). Recruiting outsiders instead of insiders has been shown to enhance team access to external ideas, enabling it to complement the exploitation of native ideas with the exploration of foreign ideas, thus improving performance (Singh and Agrawal, 2011).[15]

Scientific ability Teams of researchers who, other things equal, differ in ability may also have particular group dynamics that affect their productivity. As argued by Hamilton et al. (2003, 2012), diversity in ability may enhance team productivity if there is learning and collaboration within team members. Using individual productivity data from a garment plant, they show that holding average team ability constant, teams with more heterogeneous worker abilities are more productive. In academia, diversity in ability can also enhance output if there is a clear distribution of tasks and tasks are complementary. As a result, a team consisting of an above-average researcher and a below-average researcher may be more successful than a team of two average researchers.

3.3 Ex-ante evaluation

Despite the positive effects highlighted above, diversity may be systematically penalized by funding organizations for several reasons. First, the evaluation of a diverse team, independently if diversity is meant at the individual or at the team level, may be complex (Porter and Rossini, 1985; Nightingale, 1998). Peer review is better at evaluating applications (including curricula and research proposals) within defined fields of knowledge or levels of ability than across fields or levels. Decision-makers may not identify fruitful combinations of fields of knowledge, research practices, or research cultures nor value the usefulness of the potential results of these combinations for other areas. Evaluators often have expertise in (or preferences for) one topic or approach (Li, 2017), and therefore proposals from diverse teams may require experts from several disciplines or approaches. But then these applications may fail to reach the minimum standard of each of them. As stated by Lamont (2009, pp. 210), "combining traditional standards of disciplinary excellence with interdisciplinarity

[14] This is especially the case in countries or fields of research where academic "endogamy" is high. Endogamy is high in several European countries (greater than 50 percent in Belgium, France, Spain, and Sweden), somewhat lower in Germany and the UK, and dramatically lower in the US (Aghion et al., 2010). As it will be shown in the descriptive statistics, the percentage of academics in our dataset who are "endogamous" is around 37%.

[15] Outsiders can, of course, also come from abroad. Levin and Stephan (1999) find that exceptionally productive scientists and engineers working in the US have a higher probability of being foreign born and foreign educated than the underlying population of US scientists. Scellato et al. (2012) show that internationally mobile researchers, and in particular those that trained outside the destination country, contribute significantly to extending the quality of the research network in destination countries at no detriment to the quality of the research performed.

presents a greater challenge and creates the potential for double jeopardy for interdisciplinary scholars, because expert and generalist criteria have to be met at a same time."

Second, applications from diverse teams may be (or may be perceived to be) less "safe" or less "doable" than applications from homogeneous teams (Langfeldt, 2006; Laudel, 2006). If funding agencies are sufficiently risk averse (Stephan, 2013), applications with a good chance of generating high-impact research may not be funded if they involve a high risk of not generating any output at all.[16] Diversity in knowledge and skills, for instance, may be penalized as evaluators may view the projects of a team of specialized researchers working in the same area of research as less likely to fail. Diverse teams may be perceived as being more likely to fail because of a lack of focus or because of coordination and communication costs (Nooteboom, 1999). Diversity in educational backgrounds may also raise evaluators' concerns regarding feasibility because of potential conflicts or because outsiders are perceived to know their institution, culture, and mechanisms less than insiders. Finally, evaluators may be wary of teams of diverse abilities, as the below-average researcher may dominate the performance of the team. The evaluation process may end up placing more weight on the weakest link.[17]

3.4 Hypotheses

We now make use of the previous review of the literature to develop two hypotheses which, using the conceptual framework developed in the previous section, should allow us to predict whether funding agencies are biased against or in favor of diversity. To formalize the discussion, let us now relabel the attributes of the team of applicant Y_{ij}'s with D_i and X_i, representing, respectively, the levels of (one of the dimensions of) diversity, and a vector of the other characteristics of the team of applicants.

As previously argued, all else equal, diversity can be penalized in the evaluation process even if it might have a positive effect on performance.

- Hypothesis 1: Diversity is penalized in the award-decision process, i.e., the effect of D_i on $f^*(D_i, X_i, Z_i')$ is negative.
- Hypothesis 2: Diversity increases the quality of the project and thus the likelihood of success in ex-post performance, i.e., the effect of D_i on $g^*(D_i, X_i, Z_i)$ is positive.

[16] This behavior may be reinforced if agencies or evaluation panels follow a loss-averse behavior (Kahneman and Tversky, 1984), which consists of having asymmetric attitudes with respect to gains and losses. Their aversion to losses (the fear of financing projects that may not deliver any outcome) can be stronger than their liking of gains (even if these may represent breakthroughs). Reviewers may also be uncertainty averse. Uncertainty-averse individuals prefer a lottery with known probabilities to a similar lottery with unknown probabilities (Ellsberg, 1961).

[17] Committees of evaluators may also tend to have more disperse evaluations when they appraise diverse teams (Sah and Stiglitz, 1986). As before, disperse evaluations may make an application less likely to be funded, as it may be relatively more difficult to reach consensus among committee members (if evaluation is more subjective) or to reach minimum threshold levels (if evaluation is more quantitative).

According to our definition, the agency would be considered biased against diverse teams if hypotheses 1 and 2 were to hold. But it can be that the agency penalizes diversity (hypothesis 1 holds) but is not biased against it (hypothesis 2 does not hold).

4 Data and variables

We analyze the award decisions of the EPSRC, the main UK government agency ("Research Council") for funding research in engineering and the physical sciences. EPSRC funding is awarded on the basis of a single-blind peer review, competitive procedure. The mission of the EPSRC as set out in the 1993 Government White Paper on Science, Engineering and Technology 'Realising our Potential' is: "to promote and support high quality basic, strategic and applied research and related postgraduate training in engineering and the physical sciences (Chemistry, Physics and Mathematics), placing special emphasis on meeting the needs of the users of its research and training outputs, thereby enhancing the United Kingdom's industrial competitiveness and quality of life." The mission statement of the EPSRC has barely changed since its inception.[18]

Research Council funding has an enormous influence on the careers of academic scientists in the UK.[19] Some institutions even set individual research council income targets.[20] In the aggregate, more than half of the overall research funding of the engineering departments comes from the EPSRC. In addition, the research assessment exercises, which are used to allocate core research funding to UK universities, use research council income as an input measure in the assessment process.

4.1 Data sources, construction and sample

We start the construction of our database from all the EPSRC grant applications from 1991 to 2007. For each application, the EPSRC records contain the name of the PI and the coinvestigators (the other team members), the start and end dates, the holding organization, and the amount of funding requested.[21] PIs must be academic employees of an eligible UK organization. In almost all the applications, the PI and the coinvestigators are employees of the holding organization. We

[18] Nowadays, the mission statement is to "Promote and support, by any means, high quality basic, strategic and applied research and related postgraduate training in engineering and the physical sciences. Advance knowledge and technology (including the promotion and support of the exploitation of research outcomes), and provide trained scientists and engineers, which meet the needs of users and beneficiaries (including the chemical, communications, construction, electrical, electronic, energy, engineering, information technology, pharmaceutical, process and other industries), thereby contributing to the economic competitiveness of the United Kingdom and the quality of life."

[19] It is a key aspect in the academic promotion policies. City University's promotion policies, for instance, request evidence of contributions to research and include, next to "high quality publications," "a significant level of financial support/a number of grants from Research Councils." This is also the case in other countries. In the US, as stated by Stephan (2013): "External funding, which was once viewed as a luxury, has become a necessary condition for tenure and promotion."

[20] See "Grant income targets set at one in six universities," Times Higher Education, June 11, 2015.

[21] We observe no difference between funds requested and funds awarded.

also know whether the application has been funded or not. The unit of observation of our analysis will be a grant application of a team of one or more academics.

We match all the EPSRC grant applications with the academic calendar census data of all the engineering departments of 40 major universities in the UK (see Banal-Estañol et al., 2015, for details). We use the applications that include, as a PI or as a coinvestigator, at least one of the academic engineers of the calendar database. We discard the applications of teams of more than 10 academics, so that individual characteristics matter, but the results are very similar when we include all the proposals (only 1.5% of the applications involve more than 10 academics). Our initial sample has $17,835$ applications (teams) over 12 years (1996-2007) that include at least one researcher with full information (for $15,305$ projects we have all the information about the PI). In total, our dataset includes $3,786$ academics. We also use the applications of the period 1991-1995 for the construction of stock variables.

We use prior publication data to identify most of the job-related attributes of the team members in the application. We identify for each team member in each application all her publications in the Web of Science (WoS) in the five years prior to application date. For example, for a team member of an application of 2005, we take into account all her publications in the period 2000-2004. For each publication, we identify (i) the field assigned to the publishing journal,[22] (ii) the publishing journal's orientation category in the Patent Board classification (initiated by Narin et al., 1976, and updated by Hamilton, 2003), and (iii) the number of citations received by December 2007. These three pieces of information are going to allow us to proxy for a given researcher's fields of knowledge, skills, and ability, respectively.

We assemble other characteristics of each team member from other sources. We obtain information on the Ph.D. granting institution (for almost all researchers) from specialized websites (*ethos.bl.uk/Home.do* and *www.theses.com*) or from departmental or personal web pages. We also collect the time-varying information on academic rank from the academic calendar census. This information also allows us to construct proxies for educational background.

Finally, to construct a measure of success in ex-post performance, we again make use of the WoS database. The WoS database includes, since 2008, information on the funding sources of the publications, obtained from the publication acknowledgments. Because of the WoS coverage period, we searched for the publications that mention, as funding source, one of the $1,493$ EPSRC funded applications of the period 2005-2007. We identified 963 publications in the years 2008-2010 that acknowledge one of these EPSRC grants. This process allows us to univocally assign the

[22] WoS allocates the journals in the Science Citation Index (SCI) to 172 subject categories (such as "mechanics", "engineering, mechanical" and "mathematics, interdisciplinary applications"), which are further grouped by Leydesdorff and Rafols (2009) into 14 fields (such as "Materials Sciences", "Computer Sciences" and "Engineering") using factor analysis. Almost all of the WoS publications of our researchers are in journals of the SCI but a few of them are in journals of the Social Science Citation Index (SSCI). We group those in the latter index in a 15th field. The most common fields of the publications of our researchers are Computer Sciences (24%), Materials Sciences (22%), Engineering (20%), Chemistry (12%) and Biomedical Sciences, Environmental Sciences and Physics (4% each).

publications that came out of the project (rather than from other projects the same researchers may have). To address potential truncation biases, all our regressions control for the year in which the project started as well as its duration. As a final step, we identified the number of citations received by each of these publications by April 2016. Citation counts are generally an accepted criterion of scientific merit, since they measure the impact of the research results on other scientists (Bornmann and Daniel, 2005; Cole, 2000; Tijssen et al., 2002).

4.2 Measures

We now explain how we construct our dependent and control variables, as well as our proxies of diversity. Following prior literature, we include measures that reflect (i) the individual level of diversity of the PI, as the team leader, (ii) the average individual level of diversity of the whole team, and (iii) the team level of diversity of the whole team. Notice that we can only compute team measures of diversity for the teams of at least two members. As we shall see, we make use of several constructs used in the literature, including Blau's (1977) index of heterogeneity (which is based on the HHI, the Herfindahl-Hirschman Index) for categorical variables and Harrison and Klein's (2007) coefficient of variation for continuous measures.[23] Table 1 provides a summary of all the variables.

[Insert Table 1 here]

Dependent variables Following our empirical approach we construct two binary variables. The first, the *award* variable, refers to the ex-ante evaluation process. This variable takes a value of 1 if the application is awarded funding and a value of 0 if it is not. The second binary variable is our measure of success in ex-post performance, which we base on the sum of the "normalized" citations of all the publications coming out of the project. The normalized number of citations of a given publication is obtained by dividing the number of citations received by that publication by the average number of citations received by all the papers published in the same year and in the same field. We construct a dummy variable named *success* that assigns a value of 1 to the projects in the top 25% in terms of normalized citations among all the funded applications, and zero otherwise. As argued by Tijssen et al. (2002), gaining attention and recognition from colleagues is an important step in establishing a solid reputation of scientific excellence. Citations to a researcher's papers within other scientific publications written by fellow researchers can be used as measures of these external impacts on their scientific environments.

Control variables We include a significant number of control variables in all the regressions. We first control for ability, proxied by the variable *citations*, which adds the number of normalized citations (by year and field) of the researcher's publications in the five years prior to the

[23]The coefficient of variation, obtained by dividing the standard deviation by the mean, provides the most direct and scale-invariant measure of dispersion for a continuous variable.

application.[24,25] We also control for type of research. We use the four categories of the Patent Board classification of journals: (1) applied technology, (2) engineering and technological science, (3) applied and targeted basic research, and (4) basic scientific research. Part of the prior research considers the first two categories applied and the last two basic (Breschi et al., 2008) while other authors consider the first and the third categories applied and the second and the fourth basic (van Looy et al., 2006). We take both views into account and define the variable *research type* of a researcher in a given year as the fraction of her publications in the previous five years in the first category ("applied") relative to the publications in all four categories ("applied" and "basic"). This variable allows us to reflect the orientation of the academic on a continuous $[0, 1]$ interval scale.[26] We also control for the *application experience* of each researcher, defined as the number of applications in which she participated in the previous four years.

Additionally, we include the following demographic variables: *academic rank,* on a scale of 1 to 4 (corresponding to the UK categories of lecturer, senior lecturer, reader, and professor), the dummy variable *Russell group*, which indicates whether the researcher works in one of the universities of this prestigious set, and the *department size*, in terms of the number of (full-time-equivalent) research active members of the department, according to the RAE database. Academic rank controls for the prestige of the individual and Russell group controls for the quality and prestige of the hosting institution.[27]

Our regressions also control for the *duration* of the project and the per-capita amount of funding requested (*funds per cap*). In the award (but not in the performance) regressions, we include the fraction of grants awarded over all our sample in a given quarter, denoted as *fraction awarded* and constructed as the ratio between the total amount of money disbursed and the total amount requested.

Diversity in knowledge and skills We use multiple measures of diversity of knowledge and skills. Following the literature on interdisciplinarity (e.g., Porter and Rafols, 2009; Wagner et al., 2011; Wang et al., 2015; Yegros-Yegros et al., 2015), our measures of diversity of knowledge are based on the *number of fields* ("variety"), as well as on the degree of concentration ("balance"),

[24] We have also considered other measures of ability such as the normal count of published papers, the weighted-impact-factor sum of publications using the Journal Impact Factors (JIF) attributed to the publishing journal in the year of publication in the Science Citation Index (SCI), the average impact-factor per publication, or the number of top-cited papers. All of them give similar results.

[25] Some of the academic researchers are not in our sample for the whole period, for example, because they are junior or come from abroad. We take this into account by computing the five-year equivalent measures using the years available (out of the five years prior to the start of the project).

[26] We have replicated the exercises by using as a measure of research type the fraction of publications in the first and second groups, relative to the count of publications in all four groups. Qualitative results are very similar.

[27] We have considered an alternative measure for the quality of the holding organization based on the 2008 Research Assessment Exercise (RAE) evaluation of the quality of research undertaken by UK institutions. Using a variable that computes the fraction of papers that are at the top category of their discipline (the so-called four star papers, as opposed to the one, two or three star papers) gives similar results.

of the researchers' publications in the previous five years. As indices of concentration we use the *Blau's index of fields*, given by $1-\sum s_i^2$, where s_i is the share of the total publications in field i, and the *Shannon's index of fields*, given by $-\sum s_i \ln s_i$. As individual-level measures of diversity, we use the number, as well as the Blau and Shannon indexes, of the fields of the PI, as well as the averages of all the members of the team. As team-level measures of diversity, we use the contribution of the rest of the team to the scope of knowledge of the PI, defined by the number of fields in which the rest of the team has published but the PI has not (*num additional fields)*, as well as the number, Blau and Shannon indexes of the fields of the publications of the whole team.

To illustrate the difference between some of our individual and team-level measures of diversity of knowledge, consider two simple examples. Suppose a two-member team in which the PI has only published in field A whereas the other team member has only published in field B. This team should have no individual-level diversity of knowledge because it is composed of two specialists, whereas it exhibits team-level diversity as the researchers work in different fields. Accordingly, the *Blau's index of fields* of each researcher would be 0 whereas the *num additional fields* variable would be 1. Take instead, another team in which both team members have published the same number of papers in each of the two fields A and B. This team exhibits individual levels of diversity, as the researchers are interdisciplinary, but not team levels of diversity, as the researchers work in exactly the same fields. Accordingly, the *Blau index* of each of the team members would be 0.5 whereas the *num additional fields* would be 0.[28]

We consider an academic as having diversity of skills if she has the ability to publish in both basic and applied journals. We define the level of individual diversity of skills of the PI using the *Blau index of types* of her publications (basic and applied), which allows us to construct the HHI-based variable heterogeneity of the research types.[29] At the team level, we can make use of the *research type* of each individual which, as explained when describing the controls, is a continuous measure of the orientation of the academic. Indeed, we can compute the *coefficient of variation of the research type*, by dividing the standard deviation of research type across team members by their average level of research type.

Educational diversity We build several measures of individual educational diversity using the institution in which the researcher works and the institution in which she obtained her Ph.D. Building on the idea of salience, we construct a time-varying dummy variable for each researcher named *dum PhD outside*, which takes a value of 1 if, in that year, the institution in which she is currently working at is different from the one in which she obtained her Ph.D. We thus consider

[28] Notice that the Blau index of the publications of the whole team would not be able to distinguish between individual and team-level diversity, as it gives the same value (0.5) to the two teams in the two examples. This is why, although we use both, we favor the *num additional fiels* variable over the *Blau index* as team-level measure of diversity.

[29] As an alternative, we used a dummy variable which is equal to 1 if the value of the *research type* variable is intermediate, i.e., within the interval $(0.33, 0.66)$. Results remained unchanged.

"insiders" the researchers that did their Ph.D. in the same institution in which they currently work. We decompose the *dum PhD outside* variable further and define four other dummy variables: *dum PhD US*, *dum PhD foreign non-US, dum PhD outside RG*, and *dum PhD outside non-RG*, which take a value of 1 if the academic is an outsider and obtained her Ph.D. in, respectively, the US, a country outside the US and the UK, a UK university of the Russell group, and a UK university not in the Russell group. We use both the dummy variables for the PI, and a ratio variable for the whole team, computed as the average of the dummy variables across team members.

As a measure of team diversity, we define a variable named *num PhD outside origins,* which counts the number of institutions, other than the holding organization, in which the team members have obtained their Ph.D.

Diversity in abilities We proxy for the level of diversity of team abilities with the *coefficient of variation of citations,* computed again dividing the standard deviation of normalized citations across team members by their mean level of normalized citations (Harrison and Klein, 2007).

Team size A variable that measures the size of the team may capture, when used in conjunction with the other variables, residual levels of (team) diversity. Indeed, the team science literature has often associated diversity with the number of agents in the team (Wuyts et al., 2005; Uzzi et al., 2013). This literature argues that teams are more likely to integrate multiple and divergent perspectives, thus improving performance, than individual researchers (Singh and Fleming, 2010; Falk-Krzesinski et al., 2011). But, there is considerable evidence that although performance may initially rise as group size increases, this effect tails off or becomes negative above a certain group size threshold, i.e., either no increase or even a decrease in performance (for a review, see von Tunzelmann et al., 2003). It is argued that the larger the number of people in a group, the more effort has to be spent on unifying the broader set of inputs, and the more costly communication, coordination, and control tasks will become (Brooks, 1975). This discussion suggests that there is an inverted U-shaped relation between group size and performance. Thus, we create a variable named *num team members* (including the PI) and include it together with its square, to take second order effects into account.

4.3 Descriptive statistics

We now present descriptive statistics for some of our variables. As shown in the left panel of Table 2*a*, the percentage of applications that are awarded funding is 34%. The average number of normalized past citations is 3.5 and the average research type is 0.25. Applications have an average duration of 2.76 years and the amount requested per capita for the whole duration of the project is £129,000. Around 78% of the applications have a university from the Russell group as the holding institution, although these universities represent 60% of the pool of universities in our dataset (24 out of 40).

[Insert Table 2 here]

The right panel of Table 2*a* shows the diversity variables. The average number of members in a team is 2.5. The percentage of researchers with a Ph.D. from an outside institution in an application is, on average, 63% and the average number of the Ph.D. origins outside the hosting institution in a team is 1.2. For those projects for which we have information about at least two researchers, the average Blau index of heterogeneity of fields of the team is 0.55, the average coefficient of variation of research type is 0.94, and the coefficient of variation of past citations is 0.34 on average.

Table 2*b* provides a list of the aggregate number of proposals submitted by each university, as well as the fraction awarded funding. Although the universities of Oxford and Cambridge do not have the most applications, they do have the highest percentage of applications funded.

Table 2*c* replicates Table 2*a* for the subsample of awarded applications of the years 2005-2007. By definition, the percentage of awarded applications that are considered ex-post successful is 25%. Table 2*d* compares the means of the variables (i) award dummy, (ii) ability (citations), (iii) research type, (iv) application experience, (v) academic rank and (vi) Russell group dummy for two sets of teams: the "highly diverse" teams (respectively, "highly non-diverse" teams), defined as those that are above (resp., below) the median in every one of the three main measures of diversity: (i) diversity of knowledge proxied by *num additional fields*, (ii) diversity in educational diversity proxied by *ratio PhD outside*, and (iii) diversity in abilities proxied by *coef-var citations*. The table shows that the highly diverse teams are significantly less likely their applications awarded than highly non-diverse teams although their ability (proxied by citations to prior publications), application experience, academic rank and university prestige is similar.

5 Results

We estimate the effects of diversity on the likelihood that an application is funded, as well as on the likelihood that the project is successful ex-post. As explained earlier, diversity can be applied to the individuals of the team (individual levels of diversity) or to the team as a whole (team levels of diversity). We consider the effects of the individual levels of the PI and of the individual and team levels of diversity of the whole team, which we consider in our main results. In all the regressions, we control for year fixed effects and report robust standard errors. As mentioned in the theoretical framework, we focus on the sign of the effects rather than in their magnitudes.

5.1 Award decision

Before presenting our main results on the whole team, we report the results of the probit regressions on the effects of the individual measures of diversity of the team leader, the PI, in Table 3. As shown in Column 1, as well as in the rest of the regressions, the effects of the control variables are as expected. PI's ability (proxied by *citations* of prior publications) and *academic rank*, as well as holding university's eminence (proxied by being in the *Russell group* and by the *department size*), are important determinants of success in the EPSRC application process. More applied academics find it more difficult to obtain financing (*research type*). Experience in previous applications (*application experience)* does not seem to have a significant influence on the result. The *duration* has a negative effect whereas the per-capita amount of funding requested (*funds per cap*) does not have a significant effect. Finally, as expected, an application is more likely to be funded in a period in which the overall ratio between the money awarded and money requested (*fraction awarded*) is larger.

[Insert Table 3 here]

Columns 2, 3 and 4 show that our main measures of interdisciplinarity of the PI (*Blau index fields PI, Shannon index fields PI,* and *Blau index research types PI*) have a negative and significant influence on the likelihood of success in the grant application process. More diverse PIs in terms of knowledge and/or skills are more likely to find their application rejected. The *num fields PI* is not significant (although the correlation between this variable and *Blau index fields PI* is 0.74).[30] This suggests that an academic with a strong background in one field who has occasionally contributed to other fields (i.e., variety) is not penalized in the grant allocation process, whereas an academic with a balanced contribution in several fields is. Having been educated in another university also hinders the likelihood of success (the "*outsider*" dummies), as shown in columns 5 and 6. The (unreported) marginal impact (computed with all the other variables held at their means) shows that an outsider PI has a 3.1% lower probability of seeing her application funded than an insider PI (the base category). Separating by origin, having obtained a Ph.D. in the US does not have a significant impact. But in all the other cases, i.e., having a Ph.D. from a foreign country other than the US, or in another UK university, independently if it belongs to the Russell group or not, has a negative and significant effect. All previous results hold in the fully specified model, presented in column 7, except for that of diversity in skills, whose coefficient maintains the same sign but it is no longer significant.

Table 4 reports the results of the probit regressions on the effects of the team-level measures of diversity. All the regressions include the control variables as well as two individual-level (average) measures of diversity of the whole team, one for diversity in knowledge and skills and one for diversity in educational background, as in columns 2 and 6 of Table 3. First notice that, not only the effects of the controls, but also the effects of the measures of individual level of diversity of the whole team are similar to those of the team leader. Namely, greater individual-level diversity of knowledge and a larger share of outsiders from non-US universities have a negative influence in the likelihood of being awarded funding. All regressions also include the number of members of the team (*num team members*) and the square of this variable. Results suggest that there is a non-linear relationship between group size and success in the award process: the larger the size,

[30] The variable *num of fields PI* is not significant even when we replicate the same regression without the variable *Shannon index fields PI.*

the lower the likelihood that the team will be funded, but this effect diminishes with the size. The effects are significant in most columns although they lose some significance when we introduce some of the measures of team diversity.

[Insert Table 4 here]

Columns 2 to 8 analyze the effect of the team diversity variables on the teams with at least two members (except for column 3, which considers all the teams). Notice that the effect of these measures are computed in addition to the effect of the individual measures of diversity and of the measures of team size. Still, results are remarkably consistent. Columns 2, 3, 4 and 5 show that the level of team diversity in knowledge and skills decrease the likelihood of the team obtaining the grant. The coefficient of the number of the fields that the rest of the team work on but the PI does not (*num additional fields*), the heterogeneity of fields (*Blau index fields team*) and the coefficient of variation of research type (*coef-var research type*), are negative and significant (*num team fields* and the *Shannon index fields team* are negative but non-significant). Column 6 shows that educational diversity also reduces the likelihood of getting the application funded. The *num Ph.D. origins* is negative and significant. Column 7 reports that teams that exhibit more diversity in scientific ability, proxied by the coefficient of variation of the number of citations (*coef-var citations*), are less likely to be successful than more homogeneous teams. Column 8 confirms that all the previous effects hold in the fully specified model, except for the variable that proxies for individual diversity of knowledge (*Blau index fields ind*) which is no longer significant when added alongside the variable that proxies for team level of diversity of knowledge (*num additional fields*) and for the variable that proxies for the team level of educational diversity (*num PhD outside origins*) which is no longer significant when added alongside the variables that proxy for the individual level of educational diversity (the "*outsider*" dummies).

Taken together, the results of tables 3 and 4 suggest that diverse teams are less likely to see their proposals awarded funding than non-diverse teams. Therefore, these results provide support for Hypothesis 1.

5.2 Ex-post performance

Table 5 presents the results as to whether diverse teams are more likely to be successful in ex-post performance. We use exactly the same specifications as in Table 4. As explained before, we control for potential differential selection by using a two-step Heckman probit selection model. Below each regression we report the estimate of "rho," which measures the extent of the selection effect, together with its standard error.[31]

[Insert Table 5 here]

[31] Our estimates of rho suggest that we cannot exclude the possibility that the unobservables in the award regression are unrelated to those of the ex-post performance regression. In other words, selection into the sample of the second stage may end up being a random process.

While columns 3 and 4 show that heterogeneity of fields (measured through the *Blau index fields team* and *Shannon index fields team*) and *num fields* do not have a significant influence on ex-post performance, column 2 points at a significant positive effect of the team diversity in knowledge (*num additional fields*). On the other hand, team diversity in skills (*coef-var research type*) does not seem to have a significant impact on ex-post performance, as shown in column 5. Our results suggest that educational diversity has a positive impact on success. Three out of the four of our individual measures (the *outsider* variables) are significantly positive in many regressions. Moreover, our team measure (*num PhD outside origins)* is positive and significant in column 6. Diversity in ability (*coef-var citations*), on the other hand, does not seem to have a significant impact on the likelihood of success in column 7. Most of the regressions suggest that the linear and quadratic coefficients of the size of the team are positive and negative, respectively, as suggested by the team science literature. Although they often have the same sign, the estimates in the regression that includes all variables are not as significant as those in the other regressions. This is probably because of the low number of observations of that regression.

Taken together, the results of Table 5 suggest that diverse teams are generally more likely to be successful in terms of ex-post performance than non-diverse teams. Therefore, the results provide support for Hypothesis 2.

5.3 Mediating effects

Our empirical results so far stress that diverse teams are less likely to be successful in the award decision but, at the same time, they are generally more likely to be ex-post successful. We provided two possible explanations for this bias: the evaluation of a diverse team may be complex and the award process may perceive applications by diverse teams as being less "safe" or less "doable" than those of homogeneous teams.

In this section, we investigate which characteristics of the team leader mitigate or amplify the effects of diversity. We claim that the academic prestige of the team leader, for instance in terms of academic rank, may affect the two explanations. The agency may trust that a prestigious, well established, PI "knows what she is doing" when assembling a diverse team, whereas it may have more doubts about a PI with less prestige. In other words, safety may be less of an issue for teams led by prestigious PIs. Similarly, a diverse team led by a prestigious PI may be trusted even if her application is more complex to evaluate than that of a less diverse team.

We construct a dummy variable named *professor*, which is equal to 1 if the PI is a full professor and 0 otherwise (around 52% of the PIs are full professors). We include the interaction between this dummy variable and our diversity variables in the probit regressions on the likelihood that an application is awarded funding, as well as on the Heckman probit regressions on the likelihood that the project is ex-post successful. We use the specification that includes all the diversity and control variables (column 8 in tables 4 and 5). A positive (resp., negative) coefficient of the interaction term in the award decision indicates that the negative impact of diversity on the likelihood of obtaining a grant is mitigated (resp., amplified) if the PI is a prestigious researcher. Similarly, a positive (resp., negative) coefficient of the interaction term in the ex-post performance regression suggests that the positive impact of diversity in the likelihood of success in the project is amplified (resp., mitigated) if the PI is a prestigious researcher.

Columns 1 to 4 of Table 6 report the results for the award decision. They show that the prestige of the PI has a strong positive effect in mitigating the negative impact of diversity. The coefficients of the interaction terms are all positive and (except for the one of the four variables in column 3 for which we did not find a significant negative effect in Table 4) highly significant. Moreover, the sum of the coefficient of the interaction term and the main effect is often close to zero, indicating that the negative effect of diversity is not only mitigated but offset if the team is led by a prestigious PI.

[Insert Table 6 here]

The results for the ex-post success are reported in columns 5 to 8. None of the coefficients of the interaction terms are significant. This suggests that the prestige of the PI does not have a strong effect in mitigating or amplifying the positive effect of the team's diversity on the likelihood that the team will be successful.[32]

The results reported in Table 6 are robust in several dimensions. First, the interactions with the other diversity variables lead to qualitatively similar results. Second, we obtain the same results if we run the regressions of Table 4 for two split samples: those teams whose PI is a full professor, and those teams whose PI has a lower rank. The coefficients of the diversity variables for the subsample with prestigious PIs are less negative than those for the subsample with less prestigious PIs. Moreover, they are often not significant.

We have replicated the exercise by interacting the diversity variables with a dummy variable that is equal to 1 if the ability of the PI is above the median and 0 otherwise. The coefficients have the same sign and significance as in Table 6, although the effects for the award decision are less strong. Interestingly, all the interaction effects are insignificant if we interact the diversity variables with a dummy variable that is equal to 1 if the academic age of the PI is above the median and 0 otherwise. This suggests that it is the academic prestige or ability and not the academic age that mitigates the negative effect of diversity in the award decision process.

6 Discussion and conclusion

This paper analyzes whether funding bodies are biased against diverse teams, which have often been linked to the production of high-impact research. We develop a general framework that compares the drivers of success in the ex-ante award decision process against the drivers of success

[32] Similar results are obtained if we interact these variables with the regressions that include the measures of team diversity one by one, i.e., columns 1 to 5 of tables 4 and 5.

in ex-post performance. This approach allows us to distinguish between funding agencies that are penalizing and being biased against (as well as between rewarding and being biased in favor of) any given attribute of the team of applicants. We apply this framework to determine whether funding agencies are biased against diverse teams.

Our empirical results, based on EPSRC data, indicate that diverse teams are not only penalized but are also biased against. On the award decision, teams that exhibit greater diversity in knowledge and skills, education, and/or ability, are significantly less likely to obtain funding. This is true for all of our *numerous proxis* of (structural) diversity, independently if they measure individual or team levels of diversity, and independently if the individual levels of diversity are measured at the PI or at the whole team level. On ex-post performance, our results suggest that diverse teams are more likely to generate high-impact research. Teams that exhibit greater diversity in team levels of diversity of knowledge and individual or team levels of diversity in education account for a disproportionate share of the highly-cited projects. Even team size, which has been used as a proxy of diversity by the team science literature, tends to simultaneously decrease the likelihood of ex-ante approval and increase the likelihood of ex-post success.

In our conservative definition of a "strong" bias, we request the effect in the ex-ante funding decision to pull in the opposite direction as that of the ex-post success, and both opposing effects to be significantly different from zero. Whenever we do not find this strong evidence of a bias against diverse teams, we do find what we call "weak" evidence of a bias. Diversity in skills and scientific ability, for instance, significantly reduce the likelihood of funding despite not significantly affecting the likelihood of generating high-impact research.

We provide two possible explanations for the bias against diversity in the award decision process of funding organizations. First, the evaluation of a diverse team may be complex. Decision-makers may not identify or value fruitful combinations of fields of knowledge, research practices, or research cultures. Second, the award process may perceive applications by diverse teams as being less "safe" or less "doable" than those of homogeneous teams. Diversity in knowledge and skills, for instance, may be penalized as evaluators may view the proposals of a team of interdisciplinary researchers or a team of specialized researchers working in different areas as more likely to fail to generate any sort of outcome at all. We claim that both, safety and complexity, may be less of an issue for teams led by prestigious PIs. Consistent with this, our regressions show that the prestige of the PI mitigates the negative impact of diversity in the award regressions. In fact, the negative effect of diversity often disappears if the team is led by a prestigious PI. In contrast, the prestige of the PI does not have a significant effect in mitigating or amplifying the positive effect of the team's diversity on ex-post performance. These results are true independently if we measure prestige by academic rank or scientific ability.

Our results thus provide empirical support to the concerns voiced by academics on both sides of the Atlantic, about the important consequences of the proclivity for risk aversion and conservatism in the funding allocation process (Luukkonen, 2012; Stephan, 2013). Indeed, less diversity in funded research teams, and the resulting reduction of high-impact research, may make it unlikely that the economy will reap significant returns from its investments in R&D. One of the main reasons to place research in the university sector is that the society needs to try combinations of knowledge and skills of an unpredictable nature. Without government support, the society has a tendency to underinvest in this kind of research. Yet the system has evolved to do precisely the opposite of this, placing emphasis on safety and doability (Stephan, 2013). In this sense, our paper justifies the emergence of various mechanisms recently put in place to support different types of diversity. An example of such mechanisms include the Interdisciplinary Research program of the NIH to promote interdisciplinary collaborations.

Our framework can also be used to test for biases against other attributes of the team of applicants. In unreported regressions, we have tried (but failed) to find a bias in favor or against demographic diversity. Demographic diversity, in terms of gender, academic age, or academic rank, does not significantly affect the likelihood of grant approval nor the likelihood of ex-post success. This suggests that, without explicit mechanisms in place, evaluation processes may not end up positively discriminating, for instance, women. With respect to our control variables, our regressions show that team ability and seniority, as well as a university's eminence, are important determinants of success in the EPSRC application process. Ability and university eminence also improve the likelihood of ex-post success. But we find weak evidence of a bias in the case of seniority, because seniority, conditional on ability, increases the likelihood of grant approval but it does not affect the likelihood of ex-post success.

Our approach can also be used to test for biases against other project characteristics. Provided that data on project keywords or project descriptions are available, one could for instance test whether funding agencies not only penalize but are also biased against novel projects. Together with data on researcher characteristics, one could then distinguish between biases against diverse teams and biases against novel projects. Our two-step procedure can also be used, more generally, to test other objectives of funding agencies, beyond funding high-impact research, such as promoting original research, international collaboration, or collaboration with industry.

7

Acknowledgements

Previous versions of this paper have circulated under the title "Key success drivers in public research grants: Funding the seeds of radical innovation in academia?" We thank the editor, Keld Laursen, three reviewers, as well as Ashish Arora, Samuel Bentolila, Gianluca Carnabuci, Javier Gómez-Biscarri, Gabriel Montes-Rojas, Francesco Lissoni, Joan Llull, Carlos Serrano, Reinhilde Veugelers, and Jian Wang, for very useful comments, and Jonás Nieto for excellent research assistance. We also thank the very constructive comments received at the Jornadas de Economía Industrial 2014 in Barcelona, KID 2015 in Nice, the 15th REER Conference in Atlanta, the Ninth Annual Conference on Innovation Economics at Northwestern University, as well as at seminars at the CCP of the University of East Anglia, University College Dublin, Universidad de Salamanca, Pontificia Universidad Javeriana de Bogotá, Ben-Gurion University of the Negev, KU Leuven, and Waseda University. We gratefully acknowledge financial support from the Ministerio de Economía y Competitividad and Feder (ECO2015-63679-P, ECO2016-76998-P), Generalitat de Catalunya (2017SGR-711 and ICREA Academia), and Severo Ochoa Programme (SEV-2015-0563). The last two authors are fellows of CESifo and MOVE.

References

Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., Sapir, A., 2010. Governance and Peformance of Universities. Economic Policy 25(61): 7-59.

Arts, S., Veugelers, R., 2014. Technology Familiarity, Recombinant Novelty, and Breakthrough Invention. Industrial and Corporate Change 24 (6): 1215–1246.

Azoulay, P., Zivin, J.S., Manso, G., 2011. Incentives and Creativity: Evidence from the Academic Life Sciences. Rand Journal of Economics 42(3): 527-554.

Banal-Estañol, A., Jofre-Bonet, M., Meissner, C., 2015. The Double-edged Sword of Industry Collaboration: Evidence from Engineering Academics in the UK. Research Policy 44(6): 1160-1175.

Banal-Estañol, A., Macho-Stadler, I., Pérez-Castrillo, D., 2013. Research Output from University-Industry Collaborative Projects. Economic Development Quarterly 27(1): 71-81.

Banal-Estañol, A., Macho-Stadler, I., Pérez-Castrillo, D., 2018. Endogeneous Matching in University-Industry Collaboration: Theory and Empirical Evidence from the United Kingdom. Management Science 64(4): 1591–1608.

Bantel, K.A., Jackson, S.E., 1989. Top Management and Innovation in Banking: Does the composition of the top team make a difference? Strategic Management Journal 10: 107-124.

Barjak, F., Robinson S., 2008. International Collaboration, Mobility and Team Diversity in the Life Sciences: Impact on Research Performance. Social Geography 3(1): 23.

Blackburn, J.L., Hakal, M.D., 2006. An Examination of sources of Reer-Review Bias. Psychological Science 17(5): 378-382.

Blau, P.M., 1977. Inequality and Heterogeneity. New York: Free Press.

Bornmann, L., Daniel H.D., 2005. Selection of Research Fellowship Recipients by Committee Peer Review. Reliability, Fairness and Predictive validity of Board of Trustees' Decisions. Scientometrics 63(2): 297-320.

Boudreau, K.J., Guinan, E., Lakhani, K.R., Riedl, C., 2016. Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance and Resource Allocation in Science. Management Science 62(10): 2765-2783.

Branscomb, L.M., Auerswald, P.E., 2002. Between Invention and Innovation: An Analysis of Funding for Early-stage Technology Development. Working Paper.

Breschi, S., Lissoni, F., Montobbio, F., 2008. University Patenting and Scientific Productivity. A Quantitative Study of Italian Academic Inventors. European Management Review 5: 91-110.

Brooks, F.P., 1975. The Mythical Man-month. Reading, MA.: Addison-Wesley.

Bunderson, J.S., Sutcliffe K.M., 2002. Comparing Alternative Conceptualizations of Functional Diversity in Management Teams: Process and Performance effects. Academy of Management Journal 45(5): 875-893.

Carnabuci, G., Bruggeman, J., 2009. Knowledge Specialization, Knowledge Brokerage and the Uneven Growth of Technology Domains. Social Forces 88(2): 607-641.

Catalini, C., 2018. Microgeography and the Direction of Inventive Activity. Management Science 64(9): 4348-4364.

Cole, J.R., 2000. A short History of the use of Citations as a Measure of the Impact of Scientific and Scholarly Work. In: Cronin, B., Atkins, H. B. (Eds), The Web of Knowledge. A Festschrift in Honor of Eugene Garfield, Medford, New Jersey, USA: Information Today.

Criscuolo, P., Dahlander, L., Grohsjean, T. Salter A., 2017. Evaluating Novelty: The Role of Panels in the Selection of R&D Projects. Academy of Management Journal 60(2): 433-460.

Cummings, J.N., 2004. Work Groups, Structural Diversity, and Knowledge Sharing in a Global Organization. Management Science 50(3): 352-364.

Dahlin, K. B., Weingart, L. R., Hinds, P. J., 2005. Team Diversity and Information Use. Academy of Management Journal 48(6): 1107-1123.

De Dreu, C.K., West, M.A., 2001. Minority Dissent and Team Innovation: The Importance of Participation in Decision Making. Journal of Applied Psychology 86(6): 1191.

Deem, R., Brehony, K.J., 2000. Doctoral Students' Access to Research Cultures-are some more unequal than others? Studies in Higher Education 25-2: 149-165.

Dill, D.D., 1982. The Management of Academic Culture: Notes on the Management of Meaning and Social Integration Article in Higher Education. Higher Education 11: 30.

Disis, M.L., Slattery, J.T., 2010. The Road We Must Take: Multidisciplinary Team Science. Science Translational Medicine 2: 22.

Ellsberg, D., 1961. Risk, ambiguity, and the Savage axioms. Quarterly Journal of Economics 75: 643–669.

Evans, L., 2007. Developing Research Cultures and Researchers in Higher Education: the Role of Leadership. Paper presented at the Annual Conference of the Society for Research into Higher Education.

Falk-Krzesinski, H.J., et al., 2011. Mapping a Research Agenda for the Science of Team Science. Research Evaluation 20(2): 145-158.

Ferguson, J.P., Carnabuci, G., 2017. Risky Recombinations: Institutional Gatekeeping in the Innovation Process. Organization Science 28(1): 133-151.

Fleming, L., 2001. Recombinant Uncertainty in Technological Search. Management Science 47(1): 117-132.

Fleming, L., 2007. Breakthroughs and the 'long tail' of Innovation. MIT Sloan Management Review 49(1): 69.

Grimpe, C., 2012. Extramural Research Grants and Scientists' Funding Atrategies: Beggars Cannot be Choosers? Research Policy 41(8): 1448-1460.

Guimerà, R., Uzzi, B., Spiro, J., Nunes-Amaral, L.A., 2005. Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance. Science 308(5722): 697-702.

Hage, J., Mote, J., 2010. Transformational Organizations and a Burst of Scientific Breakthroughs: The Institut Pasteur and Biomedicine, 1889-1919. Social Science History 34(1): 13-46.

Hamilton, K.S., 2003. Subfield and Level Classification Journals. CHI Research.

Hamilton, B.H., Nickerson, J.A., Owan, H., 2003. Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation. Journal of Political Economy 111(3): 465-497.

Hamilton B.H., Nickerson, J.A., Owan, H., 2012. Diversity and Productivity in Production Teams. In Alex Bryson (ed.) Advances in the Economic Analysis of Participatory and Labor-Managed Firms (Volume 13) Emerald Group Publishing Limited: 99-138.

Harrison, D.A., Klein, K.J., 2007. What's the Difference? Diversity constructs as Separation, Variety, or Disparity in Organizations. Academy of Management Review 32(4): 1199-1228.

Harrison, D. A., Price, K. H., and Bell, M. P., 1998. Beyond Relational Demography: Time and the Effects of Surface- and Deep-level Diversity on Work Group Cohesion. Academy of Management Journal, 41: 96 -107.

Hegde, D., Mowery, D.C., 2008. Politics and Funding in the US Public Biomedical R&D System. Science 322(5909): 1797-1798.

Holligan, C., Wilson, M., Humes, W., 2011. Research Cultures in English and Scottish University Education Departments: an Exploratory Study of Academic Staff Perceptions. British Educational Research Journal, 37(4): 713-734.

Hollingsworth, R.J., 2007. High Cognitive Complexity and the Making of Major Scientific Discoveries. In: Sales A, Fournier M, editors. Knowledge, Communication, and Creativity, London, UK, Sage: 129-155.

Jacob, B.A., Lefgren, L., 2011. The Impact of Research Grant Funding on Scientific Productivity. Journal of Public Economics 95(9): 1168-1177.

Jehn, K.A., Chadwick, C., Thatcher, S.M., 1997. To agree or not to agree: The Effects of value congruence, Individual Demographic Dissimilarity, and Conflict on Workgroup Outcomes. International Journal of Conflict Management 8(4): 287-305.

Jones, B, Wuchty, S., Uzzi, B., 2008. Multi-University Research Teams: Shifting Impact, Geography and Social Stratification in Science. Science 322: 1259-1262.

Jones, B., 2009. The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder? Review of Economic Studies 76: 283-317.

Kahneman, D., Tversky, A., 1984. Choices, Values, and Frames. American Psychologist 39 (4): 341-350.

Kanter, R.M., 1977. Some Effects of Proportions on Group Life: Skewed sex ratios and Responses to token women. American Journal of Sociology 82(5): 965-990.

Lamont, M., 2009. How Professors Think: inside the curious world of Academic Judgment. Cambridge, MA: Harvard University Press.

Langfeldt, L., 2006. The Policy Challenges of Peer Review: Managing Bias, Conflict of Interests and Interdisciplinary Assessments. Research Evaluation 15(1): 31-41.

Laudel, G., 2006. The 'Quality Myth': Promoting and Hindering Conditions for Acquiring Research Funds. Higher Education 52: 375-403.

Lee, M., Om, K., Koh, J., 2000. The Bias of Sighted Reviewers in Research Proposal Evaluation: a Comparative Analysis of Blind and Open Review in Korea. Scientometrics 48(1): 99-116.

Leydesdorff, L., Rafols I., 2009. A Global Map of Science based on the ISI Subject Categories. Journal of the American Society for Information Science and Technology 60(2): 348-362.

Levin, S., Stephan, P., 1999. Are the Foreign Born a Source of Strength for U.S. Science? Science 285: 1213-14.

Li, D., 2017. Expertise vs Bias in Evaluation: Evidence from the NIH. American Economic Journal: Applied Economics 9(2): 60-92.

Luukkonen, T., 2012. Conservatism and Risk-taking in Peer Review: Emerging ERC Practices. Research Evaluation 21: 48-60.

March, J.G., 1991. Exploration and Exploitation in Organizational Learning. Organization Science 2(1): 71-87.

Marsh, H.W., Jayasinghe, U.W., Bond N.W., 2008. Improving the Peer-Review Process for Grant Applications: Reliability, Validity, Bias, and Generalizability. American Psychologist 63: 160.

Mueller, J.S., Melwani S., Goncalo, J.A., 2011. The Bias against Creativity: Why People Desire but Reject Creative Ideas. Psychological Science 21(1): 13-17.

Narin, F., Pinski, G., Gee, H., 1976. Structure of the Biomedical Literature. Journal of the American Society for Information Science 27(1): 25-45.

National Academies of Science, Engineering, and Medicine, 2015. Fostering Transformative Research in the Geographical Sciences. Washington DC. The National Acadmy Press.

Nelson, R.R., Winter, S.G., 1982. The Schumpeterian Tradeoff Revisited. American Economic

Review 72(1): 114-132.

Nerkar, A., 2003. Old is Gold? The Value of Temporal Exploration in the Creation of New Knowledge. Management Science 49(2): 211-229.

Nightingale, P., 1998. A Cognitive Model of Innovation. Research Policy 27(7): 689-709.

Nooteboom, B., 1999. Inter-firm Alliances: Analysis and design. Psychology Press.

Porter, A., Cohen, A.S., Roessner, D.J., Perreault, M., 2007. Measuring Researcher Interdisciplinarity. Scientometrics 72: 117-147.

Porter, A,. Rafols, I., 2009. Is Science Becoming more Interdisciplinary? Measuring and Mapping Six Research Fields over Time. Scientometrics, 81(3): 719-745.

Porter A., Rossini, J., 1985. Peer Review of Interdisciplinary Proposals. Science, Technology, & Human Values 10: 34.

Sah, R.K., Stiglitz, J.E., 1986. The Architecture of Economic Systems: Hierarchies and Polyarchies. American Economic Review 76: 716-727.

Scellato, G., Franzoni C., Stephan, P., 2012. Mobile Scientists and International Networks. National Bureau of Economic Research 18613.

Shin, S. J., Kim, T. Y., Lee, J. Y., Bian, L., 2012. Cognitive Team Diversity and Individual Team Member Creativity: A Cross-level Interaction. Academy of Management Journal 55(1): 197-212.

Singh, J., Agrawal, A., 2011. Recruiting for Ideas: How Firms Exploit the Prior Inventions of New Hires. Management Science 57(1): 129-150.

Singh, J., Fleming, L., 2010. Lone Inventors as Sources of Breakthroughs: Myth or Reality? Management Science 56(1): 41-56.

Smith, K.G. et al., 1994. Top Management Team Demography and Process: The role of Social Integration and Communication. Administrative Science Quarterly 39: 412-438.

Stephan, P., 2013. The Endless Frontier: Reaping What Bush Sowed? In: The Changing Frontier: Rethinking Science and Innovation Policy, NBER.

Taylor, A., Greve, H. R., 2006. Superman or the Fantastic Four? Knowledge Combination and Experience in Innovative Teams. Academy of Management Journal 49(4): 723-740.

Tijssen, R., Visser, M.S., van Leeuwen, T.N., 2002. Benchmarking International Scientific Excellence: Are highly cited Research Papers an appropriate Frame of Reference? Scientometrics 54(3): 381-397.

Uzzi, B., Mukherjee, S., Stringer, M., Jones, B., 2013. Atypical Combinations and Scientific Impact. Science 342: 468.

Van de Ven, W.P.M.M., Van Pragg B.M.S., 1981. The Demand for Deductibles in Private Health Insurance: A Probit Model with Sample Selection. Journal of Econometrics 17: 229-252.

Van Looy, B., Callaert, J., Debackere, K., 2006. Publication and Patent Behaviour of Academic Researchers: Conflicting, Reinforcing or Merely Co-existing? Research Policy 35(4): 596-608.

Van Knippenberg, D., De Dreu, C., Homan, A.C., 2004. Work Group Diversity and Group Performance: an Integrative Model and Research Agenda. Journal of Applied Psychology 89: 1008.

Viner, N., Powell, P., Green, R., 2004. Institutionalized Biases in the Award of Research Grants: A Preliminary Analysis Revisiting the Principle of Accumulative Advantage. Research Policy 33: 443-454.

Von Tunzelmann, N., Ranga, M., Martin, B., Geuna, A., 2003. The Effects of Size on Research Performance: A SPRU Review. Report for the Department of Trade and Industry.

Wagner, C.S., et al., 2011. Approaches to Understanding and Measuring Interdisciplinary Scientific Research (IDR): A Review of the Literature. Journal of Informetrics 5(1): 14-26.

Wang J, Thijs B, Glänzel W., 2015. Interdisciplinarity and Impact: Distinct Effects of Variety, Balance, and Disparity. PLoS ONE 10(5): 1-18.

Wenneras, C., Wold, A., 1997. Nepotism and Sexism in Peer Review. Nature 387: 341-343.

Williams, K.Y., O'Reilly, C.A., 1998. A Review of 40 Years of Research. Research in Organizational Behavior 20: 77-140.

Wuyts, S., Colombo, M.G., Dutta, S., Nooteboom, B., 2005. Empirical Tests of Optimal Cognitive Distance. Journal of Economic Behavior & Organization 58(2): 277-302.

Yegros-Yegros, A., Rafols, I., D'Este, P., 2015. Does Interdisciplinary Research Lead to Higher Citation Impact? The Different Effect of Proximal and Distal Interdisciplinarity. PLoS ONE 10(8): 1-21.

Table 1. List of variables

In this table we report the variables we use in the regressions and their definition. The last column indicates the category of each variable: D = dependent variable, C = control variables, DKS = diversity in knowledge and skills, DE = diversity in educational background, and DA = diversity in scientific ability. (ind) means individual levels of diversity and (team) means team levels of diversity.

Name of variable	**Definition of variable**	
Award	dummy equal to 1 if the application is awarded	D
Success	dummy equal to 1 if the project is in the top quartile in normalized citations	D
Citations	annual per-capita normalized citations of papers	C
Research type	ratio # of papers category 1 / # of papers all categories	C
Application experience	# of applications in previous 4 years per year	C
Academic rank	academic rank on a scale 1 to 4	C
Russell group	dummy variable equal to 1 if uni in the Russell group	C
Department size	# of members in the engineering faculty (in hundreds)	C
Duration	duration of the project (in years)	C
Funds per cap	ratio of requested funding / # of members of the team (in millions)	C
Fraction awarded	fraction of money awarded within a given quarter	C
Num fields	# of fields of the publications	DKS (ind/team)
Blau index fields	1 -HHI index of fields, i.e., 1 - ∑si^2 where si is the fraction of publications in field i	DKS (ind/team)
Shannon index fields	- ∑si ln(si) where si is the fraction of publications in field i	DKS (ind/team)
Num additional fields	# of fields where the team has published but the PI has not	DKS (team)
Blau index research types	1 -HHI index of the types of the publications	DKS (ind)
Coef-var research type	normalized std deviation of type of research of team members	DKS (team)
Dum/Ratio PhD outside	presence/fraction of PhD degrees from different than the current uni	DE (ind)
Dum/Ratio PhD US	presence/fraction of PhD degrees in the US	DE (ind)
Dum/Ratio PhD foreign non-US	presence/fraction of PhD degrees in a foreign country different from the US	DE (ind)
Dum/Ratio PhD outside RG	presence/fraction of PhD degrees in a UK Russell group uni different from holding uni	DE (ind)
Dum/Ratio PhD outside non-RG	presence/fraction of PhD degrees in a UK non-Russell group uni different from holding uni	DE (ind)
Num PhD outside origins	# of institutions in which members have PhD from other than the holding uni	DE (team)
Coef-var citations	normalized std deviation of citations of the team members	DA (team)
Num team members	sum of the # of coinvestigators and the PI	

Table 2a. Descriptive statistics on the sample for the award decision. In this table, we report the descriptive statistics for the dependent, control, and diversity variables that we use in the regressions for the award decision. All variables are defined in Table 1.

Dependent and control Variables	Observations	Mean	St dev	Median
Awarded	17,835	.34	.474	0
Citations	17,835	3.551	4.985	1.93
Research type	17,835	.252	.321	.1
Application experience	17,835	1.27	1.171	1
Academic rank	17,835	2.795	1.012	3
Russell group	17,835	.782	.413	1
Department size (in hundreds)	17,835	1.391	.725	1.236
Fraction awarded	17,835	.32	.083	.306
Duration (in years)	17,835	2.76	1.019	3
Funds per cap (in Million £)	17,835	.129	.28	.082

Diversity Variables	Observations	Mean	St dev	Median
Num team members	17,835	2.52	1.635	2
Blau index fields ind	17,835	.473	.214	.52
Num fields PI	15,286	3.334	1.623	3
Num additional fields	8,095	1.258	1.508	1
Blau index fields team	9,888	.553	.196	.61
Num fields team	17,835	4.007	1.898	4
Shannon index fields team	9,888	.381	.152	.4
Coef-var research type	5,807	.936	.589	1.01
Ratio PhD outside	17,835	.629	.406	.66
Ratio PhD US	17,835	.03	.146	0
Ratio PhD foreign non-US	17,835	.07	.222	0
Ratio PhD outside RG	17,835	.399	.413	.33
Ratio PhD outside non-RG	17,835	.129	.284	0
Num PhD outside origins	17,835	1.161	.877	1
Coef-var citations	9,000	.337	.211	.3

Table 2b. List of universities. This table presents the total number of applications and the fraction awarded for each university. * The University of Manchester was formed in 2004 by the merger of the University of Manchester Institute of Science and Technology (UMIST) and Victoria University. We assign the applications of the merging partners to the University of Manchester.

Russell Group	**Number of applications**	**Fraction awarded**
University of Birmingham	798	0.35
University of Bristol	396	0.35
University of Cambridge	1,062	0.44
Cardiff University	340	0.22
Durham University	187	0.30
University of Edinburgh	447	0.31
University of Exeter	151	0.28
University of Glasgow	459	0.34
Imperial College London	1,529	0.38
King's College London	192	0.28
University of Leeds	946	0.35
University of Liverpool	510	0.34
The University of Manchester*	1,496	0.33
Newcastle University	591	0.31
University of Nottingham	822	0.34
University of Oxford	663	0.39
Queen Mary	366	0.36
Queen's University of Belfast	403	0.30
University of Sheffield	1,073	0.36
University of Southampton	646	0.34
University College London	517	0.35
University of Warwick	253	0.34
University of York	96	0.35
Total	13,943	0.35

Non-Russell Group	**Number of applications**	**Fraction awarded**
University of Aberdeen	123	0.28
Aston University	183	0.32
Brunel University	260	0.17
City University	182	0.31
University of Dundee	182	0.31
University of Essex	140	0.33
University of Hull	132	0.28
Heriot-Watt University	318	0.28
Lancaster University	40	0.37
University of Leicester	199	0.28
Loughborough University	884	0.31
University of Reading	54	0.24
University of Salford	172	0.38
University of Strathclyde	536	0.30
Swansea University	368	0.39
University of Wales, Bangor	119	0.29
Total	3,892	0.30

Table 2c. Descriptive statistics on the sample for the ex-post perfomance. In this table, we report the descriptive statistics for the dependent, control, and diversity variables that we use in the regressions for the ex-post performance. All variables are defined in Table 1.

Dependent and control Variables	Observations	Mean	St dev	Median
Success	1,453	.248	.432	0
Citations	1,453	4.380	5.255	2.79
Research type	1,453	.231	.308	.07
Application experience	1,453	1.755	1.334	1.5
Academic rank	1,453	3.031	1.968	3
Russell group	1,453	.843	.364	1
Department size (in hundreds)	1,453	1.432	.730	1.288
Duration (in years)	1,453	2.729	1.178	3
Funds per cap (in Million £)	1,453	.142	.175	.1

Diversity Variables	Observations	Mean	St dev	Median
Num team members	1,453	2.745	1.923	2
Ratio PhD outside	1,453	.645	.403	.85
Ratio PhD US	1,453	.033	.154	0
Ratio PhD foreign non-US	1,453	.095	.254	0
Ratio PhD outside RG	1,453	.387	.417	.33
Ratio PhD outside non-RG	1,453	.130	.286	0
Num PhD outside origins	1,453	1.195	.936	1
Blau index fields ind	1,453	.481	.212	.53
IP num fields	1,231	3.600	1.758	3
Num additional fields	654	1.226	1.498	1
Blau index fields team	771	.563	.189	.61
Team num fields	1,453	4.286	2.022	4
Shannon index fields team	771	.396	.150	.41
Coef-var research type	481	.917	.599	.97
Coef-var citations	718	.302	.201	.265

Table 2d. Test of differences in descriptive statistics for very diverse and very non-diverse teams. This table presents the test of differences in the mean of two groups, the very diverse teams (defined as those above average diversity in terms of additional broad fields, scientific diversity, and outsiders) and the very non-diverse teams (defined as those below average in the three measures). ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Table 1.

	Very diverse team	Very non-diverse team	Difference in means *t* -stat
Awarded	.283	.372	-.089***
Citations	3.599	3.241	.359
Research type	.247	.294	.048***
Application experience	1.246	1.306	-.060
Academic rank	2.748	2.732	.016
Russell group	.774	.756	.018
Observations	980	713	

Table 3. Likelihood of award against the level of individual diversity of the PI. This table reports Probit regression results of the effect of several control and individual diversity variables of the PI of the projects on the likelihood that the project is awarded. Robust standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Table 1.

	Controls	Knowledge in skills			Educational background		All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Individual diversity							
Blau index fields PI		-0.082*					-0.126***
		[0.046]					[0.049]
Num fields PI			0.020				
			[0.015]				
Shannon index fields PI			-0.263**				
			[0.134]				
Blau index research types PI				-0.127**			-0.091
				[0.060]			[0.063]
Dum PhD outside					-0.077***		
					[0.024]		
Dum PhD US						-0.022	-0.024
						[0.065]	[0.065]
Dum PhD foreign non-US						-0.149***	-0.149***
						[0.047]	[0.047]
Dum PhD outside RG						-0.072***	-0.075***
						[0.026]	[0.026]
Dum PhD outside non-RG						-0.081**	-0.085**
						[0.035]	[0.035]
Controls							
Citations	0.010***	0.011***	0.010***	0.010***	0.014***	0.014***	0.014***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.003]	[0.003]	[0.003]
Research type	-0.063**	-0.064**	-0.060*	-0.042	-0.091***	-0.092***	-0.078**
	[0.032]	[0.032]	[0.032]	[0.033]	[0.034]	[0.034]	[0.035]
Application experience	-0.010	-0.009	-0.010	-0.009	-0.013	-0.013	-0.011
	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.011]
Academic rank	0.073***	0.075***	0.072***	0.076***	0.069***	0.068***	0.073***
	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Russel Group	0.116***	0.116***	0.115***	0.116***	0.106***	0.107***	0.107***
	[0.027]	[0.027]	[0.027]	[0.027]	[0.028]	[0.028]	[0.028]
Department size	0.062***	0.063***	0.062***	0.061***	0.064***	0.062***	0.062***
	[0.015]	[0.015]	[0.015]	[0.015]	[0.016]	[0.016]	[0.016]
Fraction awarded	1.523***	1.529***	1.531***	1.521***	1.526***	1.530***	1.536***
	[0.157]	[0.158]	[0.158]	[0.158]	[0.163]	[0.163]	[0.164]
Duration	-0.137***	-0.137***	-0.137***	-0.136***	-0.133***	-0.133***	-0.133***
	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
Funds per capita	-0.049	-0.049	-0.050	-0.050	-0.055	-0.055	-0.056
	[0.063]	[0.063]	[0.063]	[0.064]	[0.066]	[0.066]	[0.066]
Constant	-0.872***	-0.844***	-0.859***	-0.866***	-0.815***	-0.813***	-0.763***
	[0.090]	[0.091]	[0.091]	[0.090]	[0.095]	[0.095]	[0.097]
Observations	15,305	15,286	15,286	15,305	14,246	14,246	14,227

Table 4. Likelihood of award against the level of individual and team diversity of the team. This table reports Probit regression results of the effect of several control and individual and team diversity variables of the team members of the projects on the likelihood that the project is awarded. Robust standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Table 1.

	Basic	Knowledge and skills				Educational back	Scient ability	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Team diversity								
Num additional fields		-0.046***						-0.071***
		[0.012]						[0.015]
Blau index fields team			-0.256**					
			[0.122]					
Num team fields				-0.006				
				[0.014]				
Shannon index fields team				-0.194				
				[0.233]				
Coef-var research type					-0.092**			-0.079*
					[0.040]			[0.043]
Num PhD outside origines						-0.039***		0.045
						[0.015]		[0.032]
Coef-var citations							-0.122*	-0.196*
							[0.069]	[0.106]
Num team members	-0.102***	-0.020	-0.092***	-0.077**	-0.059	-0.066*	-0.080**	-0.011
	[0.019]	[0.041]	[0.020]	[0.034]	[0.043]	[0.034]	[0.036]	[0.050]
Num team members squared	0.011***	0.003	0.011***	0.009***	0.008*	0.008**	0.008**	0.002
	[0.002]	[0.004]	[0.002]	[0.003]	[0.004]	[0.003]	[0.004]	[0.005]
Individual diversity								
Blau index fields ind	-0.084*	-0.069	0.156	0.159	-0.068	0.009	-0.008	-0.010
	[0.047]	[0.104]	[0.124]	[0.128]	[0.097]	[0.071]	[0.075]	[0.136]
Ratio PhD US	-0.082	0.080	-0.079	0.013	0.114		-0.007	0.138
	[0.068]	[0.124]	[0.068]	[0.112]	[0.151]		[0.117]	[0.173]
Ratio PhD foreign non-US	-0.224***	-0.330***	-0.225***	-0.340***	-0.393***		-0.362***	-0.444***
	[0.049]	[0.091]	[0.049]	[0.082]	[0.109]		[0.088]	[0.130]
Ratio outside RG	-0.089***	-0.096**	-0.089***	-0.116***	-0.159***		-0.141***	-0.227**
	[0.027]	[0.049]	[0.027]	[0.044]	[0.059]		[0.047]	[0.092]
Ratio outside nonRG	-0.065*	-0.218***	-0.066*	-0.196***	-0.201**		-0.209***	-0.286***
	[0.038]	[0.070]	[0.038]	[0.064]	[0.084]		[0.067]	[0.110]
Controls								
Citations	0.009***	0.008**	0.008***	0.008**	0.008	0.008***	0.010***	0.006
	[0.002]	[0.004]	[0.002]	[0.003]	[0.005]	[0.003]	[0.004]	[0.005]
Research type	-0.103***	-0.056	-0.103***	-0.049	-0.277***	-0.038	-0.040	-0.327***
	[0.032]	[0.053]	[0.032]	[0.048]	[0.087]	[0.047]	[0.050]	[0.094]
Application experience	0.007	0.010	0.007	0.020	0.015	0.021	0.015	0.015
	[0.011]	[0.018]	[0.011]	[0.016]	[0.024]	[0.016]	[0.018]	[0.025]
Academic rank	0.073***	0.131***	0.072***	0.116***	0.114***	0.115***	0.123***	0.127***
	[0.011]	[0.021]	[0.011]	[0.018]	[0.026]	[0.018]	[0.020]	[0.028]
Russell Group	0.121***	0.092**	0.120***	0.099***	0.066	0.098***	0.100***	0.056
	[0.025]	[0.038]	[0.025]	[0.034]	[0.045]	[0.034]	[0.035]	[0.048]
Department size	0.048***	0.048**	0.049***	0.048**	0.034	0.055***	0.049**	0.056**
	[0.014]	[0.022]	[0.014]	[0.020]	[0.027]	[0.020]	[0.021]	[0.028]
Fraction awarded	1.624***	1.755***	1.624***	1.686***	1.503***	1.676***	1.721***	1.599***
	[0.146]	[0.218]	[0.146]	[0.197]	[0.258]	[0.197]	[0.208]	[0.276]
Duration	-0.100***	-0.005	-0.101***	-0.009	0.002	-0.010	0.006	-0.004
	[0.012]	[0.018]	[0.012]	[0.017]	[0.020]	[0.017]	[0.017]	[0.021]
Funds per capita	-0.136	-0.890***	-0.136	-1.025***	-1.033***	-1.013***	-1.044***	-0.880***
	[0.085]	[0.171]	[0.085]	[0.168]	[0.204]	[0.166]	[0.171]	[0.207]
Num fields PI		0.008						-0.014
		[0.013]						[0.017]
Constant	-0.729***	-1.171***	-0.724***	-1.050***	-0.913***	-1.124***	-1.081***	-0.929***
	[0.091]	[0.156]	[0.091]	[0.138]	[0.204]	[0.136]	[0.152]	[0.224]
Observations	17,835	8,037	17,835	9,811	5,777	9,811	8,922	5,116

Table 5. Likehood of expost success against the level of individual and team diversity of the team. This table reports the second stage of the Heckman Probit regression results of the effect of several control and individual and team diversity variables of the team members of the projects on the likelihood that the project is successful, that is, that it produces publications in the top 25% in terms of citations normalized by year and by field among all the projects. Robust standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. All variables are defined in Table 1.

	Basic	Knowledge and skills			Type	Educational Back	Scient. Ability	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Team diversity								
Num additional fields		0.070*						0.048
		[0.039]						[0.082]
Blau index fields team			0.090					
			[0.528]					
Team num fields				0.034				
				[0.051]				
Shannon index fields team				-0.461				
				[0.943]				
Coef-var research fields					-0.190			0.019
					[0.159]			[0.158]
Num PhD outside origins						0.146***		-0.147
						[0.051]		[0.092]
Coef-var citations							-0.358	-0.358
							[0.280]	[0.445]
Num team members	0.132*	0.030	0.129*	0.111	0.197	0.050	0.054	0.117
	[0.079]	[0.103]	[0.077]	[0.121]	[0.167]	[0.120]	[0.126]	[0.139]
Num team members squared	-0.009	0.002	-0.009	-0.008	-0.015	-0.006	-0.003	-0.001
	[0.009]	[0.011]	[0.008]	[0.012]	[0.016]	[0.012]	[0.012]	[0.015]
Individual diversity								
Blau index fields ind	-0.077	0.233	-0.163	0.221	0.041	0.113	-0.012	0.588
	[0.183]	[0.335]	[0.521]	[0.536]	[0.400]	[0.286]	[0.307]	[0.538]
Ratio PhD US	0.535**	0.776	0.528**	1.142***	1.295**		1.212***	1.262
	[0.248]	[0.587]	[0.247]	[0.422]	[0.539]		[0.436]	[0.961]
Ratio PhD foreign non-US	0.542***	0.382	0.539***	0.472*	0.479		0.627**	0.742**
	[0.170]	[0.256]	[0.170]	[0.285]	[0.393]		[0.298]	[0.377]
Ratio PhD outside RG	0.206*	0.393**	0.204*	0.370*	0.335		0.370*	0.717**
	[0.112]	[0.196]	[0.111]	[0.196]	[0.258]		[0.213]	[0.296]
Ratio PhD outside non-RG	0.050	0.324	0.049	0.408*	0.271		0.433*	0.667*
	[0.147]	[0.261]	[0.147]	[0.248]	[0.316]		[0.256]	[0.397]
Controls								
Citations	0.023***	0.020	0.023***	0.031**	0.029*	0.032***	0.036***	0.031
	[0.008]	[0.018]	[0.008]	[0.012]	[0.018]	[0.012]	[0.013]	[0.022]
Research type	-0.320**	-0.351	-0.321**	-0.419**	-0.885**	-0.439**	-0.494**	-0.326
	[0.136]	[0.265]	[0.136]	[0.211]	[0.387]	[0.206]	[0.217]	[0.632]
Application experience	-0.004	-0.065	-0.004	-0.037	0.007	-0.035	-0.065	-0.029
	[0.031]	[0.045]	[0.031]	[0.052]	[0.077]	[0.051]	[0.053]	[0.070]
Academic rank	-0.066	-0.175**	-0.066	-0.133	-0.160	-0.128	-0.126	-0.256***
	[0.051]	[0.075]	[0.050]	[0.104]	[0.146]	[0.099]	[0.111]	[0.096]
Russell Group	0.325***	0.150	0.325***	0.236	0.218	0.190	0.206	0.126
	[0.114]	[0.240]	[0.113]	[0.163]	[0.211]	[0.163]	[0.169]	[0.321]
Department size	0.044	-0.062	0.044	-0.008	-0.146	0.026	-0.022	-0.154*
	[0.054]	[0.073]	[0.054]	[0.083]	[0.097]	[0.079]	[0.084]	[0.092]
Duration	0.198***	0.192**	0.198***	0.221***	0.241***	0.220***	0.232***	0.232*
	[0.051]	[0.086]	[0.051]	[0.062]	[0.073]	[0.059]	[0.061]	[0.134]
Funds per capita	0.496	1.528**	0.492	1.429**	1.300	1.454**	1.312*	1.333*
	[0.314]	[0.654]	[0.313]	[0.677]	[0.821]	[0.652]	[0.672]	[0.748]
Num fields PI		0.006						-0.041
		[0.039]						[0.069]
Constant	-2.249***	-0.263	-2.253***	-1.786**	-1.566*	-1.452	-1.456	-0.252
	[0.337]	[1.501]	[0.329]	[0.875]	[0.937]	[0.959]	[0.937]	[1.612]
rho	0.346	-0.954	0.309	0.016	0.105	-0.076	0.009	-0.871
	[0.366]	[1.045]	[0.384]	[0.544]	[0.677]	[0.547]	[0.559]	[1.330]
Uncensored observations	1,453	654	1,453	771	481	771	718	428
Observations	5,284	2,421	5,284	2,874	1,821	2,874	2,693	1,610

Table 6. Mediating effects. Panel A of this table reports Probit regression results of the effect of several control, individual and team diversity variables of the team members of the projects, as well as some interaction terms on the likelihood that the project is awarded. Each column includes the interaction between the dummy variable Professor, which is equal to 1 if the PI is a full professor and 0 otherwise, with diversity variables. Column 1 includes the interaction of Professor with the variable Number of additional fields, column 2 with the Coefficient of variation of the research type, column 3 with the four variables used to identify Outside academics, and column 4 with the Coefficient of variation of the citations. Panel B reports the second stage of the Heckman Probit regression results of the effect of the same variables as Panel A on the likelihood that the project is successful, that is, that it produces publications in the top 25% in terms of citations normalized by year and by field among all the projects. Robust standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. The control and diversity variables are defined in Table 1.

	Panel A: Award				**Panel B: Expost success**			
	Knowledge and skills		Educational back	Scient Ability	Knowledge and skills		Educational back	Scient Ability
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Main effects (team diversity)								
Num additional fields	-0.092***	-0.064***	-0.067***	-0.065***	0.086	0.039	0.034	0.047
	[0.017]	[0.015]	[0.015]	[0.015]	[0.094]	[0.080]	[0.086]	[0.090]
Coef-var research type	-0.080*	-0.158***	-0.081*	-0.082*	0.024	0.027	0.059	0.027
	[0.043]	[0.048]	[0.043]	[0.043]	[0.155]	[0.224]	[0.155]	[0.161]
Num PhD outside origins	0.041	0.037	0.039	0.043	-0.148*	-0.142	-0.148	-0.152*
	[0.032]	[0.032]	[0.032]	[0.032]	[0.089]	[0.098]	[0.096]	[0.089]
Coef-var citations	-0.186*	-0.209**	-0.200*	-0.378***	-0.364	-0.399	-0.547	-0.116
	[0.107]	[0.107]	[0.107]	[0.123]	[0.441]	[0.450]	[0.520]	[0.518]
Num team members	-0.010	-0.016	-0.018	-0.021	0.119	0.116	0.114	0.130
	[0.051]	[0.051]	[0.051]	[0.051]	[0.136]	[0.149]	[0.151]	[0.137]
Num team members squared	0.001	0.002	0.002	0.002	-0.001	-0.001	-0.000	-0.003
	[0.005]	[0.005]	[0.005]	[0.005]	[0.015]	[0.015]	[0.016]	[0.016]
Main effects (individual diversity)								
Blau index fields ind	0.005	0.034	0.047	0.032	0.551	0.600	0.677	0.512
	[0.137]	[0.137]	[0.137]	[0.137]	[0.557]	[0.541]	[0.580]	[0.580]
Ratio PhD US	0.140	0.160	0.049	0.136	1.235	1.353	1.265	1.231
	[0.173]	[0.173]	[0.225]	[0.173]	[1.022]	[0.852]	[0.982]	[1.164]
Ratio PhD foreign non-US	-0.444***	-0.434***	-0.694***	-0.436***	0.768**	0.740*	0.997	0.765**
	[0.130]	[0.130]	[0.163]	[0.130]	[0.372]	[0.397]	[0.616]	[0.376]
Ratio PhD outside RG	-0.220**	-0.213**	-0.288***	-0.220**	0.720**	0.721**	0.908***	0.727**
	[0.092]	[0.092]	[0.099]	[0.092]	[0.303]	[0.294]	[0.333]	[0.320]
Ratio PhD outside non-RG	-0.278**	-0.267**	-0.418***	-0.272**	0.668	0.692*	0.672	0.648
	[0.111]	[0.111]	[0.133]	[0.111]	[0.413]	[0.377]	[0.517]	[0.439]
Interactions								
Num additional fields*Professor	0.055***				-0.074			
	[0.019]				[0.062]			
Coef-var research type*Professor		0.152***				-0.040		
		[0.039]				[0.178]		
Ratio PhD US*Professor			0.237				0.083	
			[0.305]				[1.242]	
Ratio PhD foreign non-US*Professor			0.536**				-0.395	
			[0.212]				[0.712]	
Ratio PhD outside RG*Professor			0.164*				-0.294	
			[0.085]				[0.288]	
Ratio PhD outside non-RG*Professor			0.313**				-0.061	
			[0.150]				[0.586]	
Coef-var citations*Professor				0.373***				-0.401
				[0.126]				[0.403]
Control Variables	YES	YES	YES	YES	YES	YES	YES	YES
rho					-0.943	-0.715	-0.697	-0.960
					[1.466]	[1.146]	[1.138]	[1.698]
Uncensored observations					428	428	428	428
Observations	5,116	5,116	5,116	5,116	1,610	1,610	1,610	1,610

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1