Key success drivers in public research grants: Funding the seeds of radical innovation in academia?*

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

September 2017

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. In contrast, demographic diversity, neither in terms of gender, age, or academic rank, does not significantly affect the likelihood of grant approval nor the likelihood of ex-post success. We provide evidence suggesting that funding agencies may be biased against structural diversity, not because applications of diverse teams are more difficult to evaluate, but because they are perceived to be less safe or less doable.

JEL Classification numbers: O32, I23

Keywords: Funding organizations, diversity, transformative research, radical innovation, research grants

*We thank Ashish Arora, Samuel Bentolila, Gianluca Carnabuci, Javier Gómez-Biscarri, Gabriel Montes-Rojas, Francesco Lissoni, Joan Llull, Carlos Serrano, Reinhilde Veugelers, Jian Wang, the senior editor Deepak Hegde, and three reviewers for very useful comments, as well as 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 and University College Dublin. We gratefully acknowledge financial support from the Ministerio de Economía y Competitividad and Feder (ECO2012-31962, ECO2015-63679-P, ECO2016-76998-P), Generalitat de Catalunya (2014SGR-142 and ICREA Academia), and Severo Ochoa Programme (SEV-2015-0563). The last two authors are fellows of CESifo and MOVE.

[†]Universitat Pompeu Fabra and Barcelona GSE. 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. Innovation agencies worldwide then aim to support firms that search for breakthrough technologies that can fundamentally change existing markets (Branscomb and Auerswald, 2002). Similarly, public funding agencies for scientific research, such as the National Science Foundation (NSF) or the UK Research Councils, 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.[1]

Diversity is often viewed as a crucial condition for radical innovation (Nelson and Winter 1982; Fleming, 2007). Indeed, knowledge 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 combine researchers of different fields of knowledge, skills, or abilities in the same team (Fleming, 2001). According to this argument, team diversity should be more likely to spark new insights (Guimerà et al., 2005; Jones et al., 2008; Jones, 2009).

Still, many commentators claim that funding agencies are biased against diverse teams (Langfeldt, 2006; Laudel, 2006). Critics argue that despite its positive effects diversity is penalized in the funding process, either 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 funding agencies penalize diverse teams, and if so, whether this is at the expense of transformative research.

This paper analyzes whether funding agencies are biased against teams that can produce transformative research. We develop a general framework that compares the drivers of success in the ex-ante award decision 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 transformative research. If, for instance, team diversity increases the likelihood of generating transformative 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 (or in favor of) diversity of knowledge if these teams have a higher (lower) likelihood of success but a lower (higher) likelihood of being funded.

We use our general framework to determine whether funding agencies are biased against diverse teams. We identify and characterize the two main types of diversity described in prior literature: "structural diversity" – differences in not so visible underlying attributes that are job-related, such as knowledge and skills, educational background and scientific ability – and "demographic diversity" – differences in readily detectable attributes such as sex, age, and tenure –. The diversity literature

[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.

argues that structural diversity has a positive effect in ex-post performance, whereas the effect of demographic diversity is mainly negative (Williams and O'Reilly, 1998). In contrast, in terms of evaluation, structural diversity may be systematically penalized by funding organizations whereas demographic diversity may be systematically rewarded. This leads us to hypothesize that funding agencies may be biased against structurally diverse teams and biased in favor of demographically 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).[2] We make use of publication and demographic data to construct variables that proxy for the structural and demographic diversity of the applicant teams. Funding agencies are ideal for analyzing the general stance toward diversity because we (and evaluators) know so much about the job-related attributes of the applicants.

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 ability, are significantly less likely to obtain funding, they are generally more likely to be successful. In contrast, demographic diversity, either in terms of gender, age or academic rank, does not significantly affect the likelihood of grant approval nor the likelihood of ex-post success. Therefore, our empirical results confirm our hypotheses on structural diversity but not those on demographic diversity. In terms of explanation, we provide evidence that funding agencies may be biased against structural diversity, not because applications of diverse teams are more difficult to evaluate, but because they are perceived as being less "safe" or less "doable."

Our approach in identifying the biases of funding agencies differs, in terms of both method and objectives, from those that recently appeared in the literature.[3] Boudreau et al. (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. 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.[4] As our aim is to study the attitude of

[2] Banal-Estañol et al. (2013; 2017) 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.

[3] More generally, other researchers have questioned funding organizations and the funding 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.

[4] To construct a measure of ex-post performance, we need to trace back the research output of each specific application team, 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 for funded applications in the econometric method.

grantor agencies toward team diversity, we focus on grants that can be submitted by individuals or teams (as opposed to Azoulay et al., 2011, Boudreau et al., 2016, and Li, 2017, who use applications submitted by individuals).[5]

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."[6] This is of course because funding agencies need to make the 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 gender 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,

[5] 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.

[6] 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-erc/mission, and http://www.forskningsradet.no/en/Funding/FRINATEK/1254025689182.

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 gender or affiliation, which should in principle be unrelated to ex-post performance, 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 personal 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) 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 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 $\partial g^*/\partial Y_{ij} > 0$.

In the same vein, we denote the (unobserved) 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 $g^*(\cdot)$. As we shall argue below, the agency may discount or reward certain attributes, such as team diversity, independently of their effect on expected quality. 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 very 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 ask both opposing effects to be significantly different from zero, in what can be seen as "strong" 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, 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).[7] 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).

[7] 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).

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 $\beta_j < 0$.

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 stringency of the evaluation process in a given quarter, which should affect (and it does, empirically) the likelihood of funding but there is no reason to think that it will affect the likelihood of success in ex-post performance directly.[8]

[8] 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).

3 Team diversity: Ex-post performance and ex-ante evaluation

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. Diversity is often viewed as a crucial condition for producing "Schumpeterian" novel combinations (Nelson and Winter, 1982; Fleming, 2007). Yet the evidence is somewhat mixed. While some studies have shown that diversity can be beneficial (Watson et al., 1993), others have suggested that it can negatively affect performance (Miller et al., 1998). In part, the lack of a clear picture stems from the different types of diversity being measured (Rodan and Galunic, 2004).

3.1 Types of diversity

Diversity may in principle be applied to a wide number of attributes, ranging from age to gender, from religious to functional background, and from task to relational skills. 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 job-related, such as functional and educational background. Cummings (2004) makes a similar classification, distinguishing between "demographic diversity" (e.g., member differences in sex, age, or tenure) and "structural diversity" (e.g., member differences in terms of sources of task information and know-how).

We follow this literature and distinguish between structural and demographic diversity. 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. Similarly, we contemplate demographic diversity in terms of (i) gender, (ii) age, and (iii) tenure.

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

3.2 Structural diversity

Structural diverse teams, 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). Structurally 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 (interpersonal) or at the individual level (intrapersonal).

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 interpersonal diversity, but were all led by (intrapersonal) diverse directors, who had the capacity to integrate diversity and to address problems relevant to numerous fields of science.[9] Catalini (2017) 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.[10]

Educational background The combination of knowledge and skills from different educational backgrounds or research cultures has also been found to improve performance, both in science and elsewhere. For instance, Smith et al. (1994) find that top management team educational diversity is positively associated with company financial performance. 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).

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

[10] Catalini (2017) 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 (2017) does using a whole laboratory as a unit of observation.

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.[11] Previous research in other settings 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).

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.

Despite the positive effects highlighted above, structural 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 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 structurally diverse teams, especially those led or composed of less prestigious researchers, may be (or may be perceived to be) less "safe" or less "doable" than applications from homogeneous teams (Langfeldt, 2006; Laudel, 2006). As funding agencies are risk averse (Stephan, 2013), applications with a good chance of generating transformative research may not be funded if they involve a high risk of not generating any output at all.[12] Diversity in

[11] 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).

[12] This behavior may be reinforced if agencies or evaluation panels follow a loss-averse behavior (Kahneman and

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 conflicts or because outsiders 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.[13]

3.3 Demographic diversity

Demographic diversity may have quite the opposite effects of structural diversity on ex-post performance (see, for instance, the review of the literature by Williams and O'Reilly, 1998). Diversity in age and tenure, for instance, has been shown to have negative effects. Hamilton et al. (2012) find than teams with more diversity in age are less productive. Stvilia et al. (2011) find a negative relationship between (academic) tenure diversity and team performance.

As an explanation, the social categorization perspective holds that observable similarities and differences are used as a basis for categorizing self and others into groups. People tend to like, trust, and favor in-group members more than out-group members (Turner et al., 1987). This means that team members are more positively inclined toward their team and the people within it if fellow team members are in-group members. Moreover, categorization processes may produce subteams within the team (i.e., "us" and "them"), and may give rise to problematic relations. As a result, diversity may lower team commitment (Tsui et al., 1992) and team cohesion (O'Reilly et al., 1989), and relational conflicts increase (Pelled et al., 1999).

In spite of the negative effects, demographic diversity may be systematically rewarded by funding organizations. Evaluation processes may positively discriminate in favor of minority groups, such as female researchers in engineering academia (explicitly or implicitly). For example, the guide of the evaluation process of the EU Horizon 2020 Programme explicitly states that "gender balance comes into play as a ranking factor to prioritize proposals above threshold with the same scores." Still, empirical research on gender bias has produced "data and interpretations which at times are contradictory" (Rees, 2011, p. 140). Although the initial assumption was that men are more favorably treated than women, recent studies have reversed such results (see, Lee et al., 2013,

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).

[13] Committees of evaluators may also tend to have more disperse evaluations when they appraise structurally 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).

and the references therein). For instance, Wennerås and Wold (1997) found that female fellowship applicants had to be more productive than a male applicant to receive the same competence score. However, replications of the study for the same institution in different periods found that gender-based differences can be reversed (Sandström and Hällsten, 2008).

Programs may also foster the exchange of best practices and experiences by promoting the pairing of senior faculty members with significant experience with more junior colleagues. Some grant competitions even explicitly state that "the successful application will support a team of junior and senior researchers" (Canadian National Transplant Research Program, 2016). Having a mix of junior and senior researchers has been thought of as a channel to generate long-term benefits for the society in terms of knowledge exchange and capacity building (Markham et al., 2013).

3.4 Hypotheses

We now make use of the previous review of the literature to develop four 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 structural or demographic diversity. To formalize the discussion, let us now relabel the attributes of the team of applicant Y_{ij}'s with S_i, D_i and X_i, representing, respectively, the levels of (one of the dimensions of) structural and demographic diversity, and a vector of the other characteristics of the team of applicants.

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

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

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

The discussion in the previous section suggests that, all else equal, demographic diversity can be rewarded in the evaluation process even if it might have a negative effect on ex-post performance.

- Hypothesis 3: Demographic diversity is rewarded in the award-decision process, i.e., the effect of D_i on $f^*(S_i, D_i, X_i, Z_i')$ is positive.
- Hypothesis 4: Demographic diversity decreases the quality of the project and thus the likelihood of success in ex-post performance, i.e., the effect of D_i on $g^*(S_i, D_i, X_i, Z_i)$ is negative.

Using the same arguments as before, the agency would be biased in favor of demographically diverse teams if hypotheses 3 and 4 were to hold.

4 Data and descriptive statistics

4.1 Data sources and sample

We analyze the award decisions of the EPSRC, the main UK government agency ("Research Council") for funding research in engineering and the physical sciences. Research Council funding has an enormous influence on the careers academic scientists in the UK.[14] Some institutions even set individual research council income targets.[15] 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.

The unit of observation of our analysis is a grant application of a team of one or more academics. For each application, the EPSRC records contain the name of the principal investigator (the PI), and the coinvestigators (the other team members), the start and end dates, the holding organization, and the amount of funding requested.[16] We also know whether the application has been funded or not. 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. Funding is awarded on the basis of a single-blind peer review, competitive procedure.

We match all the EPSRC grant applications from 1991 to 2007 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 $18,576$ applications (teams) over 12 years (1996-2007) that include at least one researcher with full information. 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 unique research field assigned by the WoS to the publishing journal,[17]

[14] 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."

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

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

[17] WoS assigns to each journal in the Science Citation Index (SCI) one (and only one) of its 14 broad fields (such as Environmental Sciences, Material Sciences, and Engineering). Almost all of the WoS publications of our researchers

(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.[18] 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 gender, Ph.D. year, and granting institution from specialized websites (*ethos.bl.uk/Home.do* and *www.theses.com*) or from departmental or personal web pages. We also collect time-varying information on academic rank and current institution from the academic calendar census. All this information allows us to construct demographic variables – gender, academic age, and academic rank – as well as proxies for educational background.

We also search for citation data to construct a measure of success in ex-post performance. 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). We again make use of publication data in the WoS database, which has been systematically collecting information on funding sources from the acknowledgments of the publications since 2008. As a result of this coverage period, we collect publication data for the $1,493$ funded applications in the period 2005-2007.[19] We identified 963 publications in the years 2008-2010 that acknowledge one of these EPSRC grants as a funding source. Finally, we identified the number of citations received by each of these publications by April 2016.

4.2 Variables

We now explain how we construct our dependent and control variables, as well as our proxies of structural and demographic diversity. Following prior literature, we include measures that reflect (i) the intrapersonal level of diversity of the PI, as the team leader, (ii) the intrapersonal level of diversity of the whole team, and (iii) the interpersonal level of diversity of the whole team. Notice that we can only compute interpersonal 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 across a number of categories and Harrison and Klein's (2007) coefficient of variation across continuous measures.[20] Table 1 provides a summary of all the variables.

[Insert Table 1 here]

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 broad field.

[18] We use the number of citations of the paper as a measure of past scientific performance. Similar results are obtained if we use the Journal Impact Factors (JIF) attributed to the publishing journal in the year of publication in the Science Citation Index (SCI).

[19] Our results on the award decisions are similar to those of the full sample although the significance of the coefficients is slightly lower.

[20] 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.

Dependent variables Following our empirical approach we construct two binary variables. The first is the *award* variable, which 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 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. 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 of the researcher's publications in the five years prior to the application.[21] 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.[22] 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 age*, defined as the difference between the application year and the Ph.D. year,[23] *academic rank,* on a scale of 1 to 4 (corresponding to the UK categories of lecturer, senior lecturer, reader, and professor), and the dummy variable *Russell group*, which indicates whether the researcher works in one of the

[21] 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. For details, see Banal-Estañol et al. (2016).

[22] 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.

[23] For some researchers, we have not been able to identify the year of the Ph.D. In those cases, we assign to the researcher as the Ph.D. year the year of the first publication of the researcher plus two. We use this convention because it is the best aproximation for those academics for whom we do have the Ph.D. year.

universities of this prestigious set. Academic rank controls for the prestige of the individual and Russell group controls for the quality and prestige of the hosting institution.[24]

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 overall EPSRC grants awarded 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 measure the level of intrapersonal diversity of knowledge of a given researcher using Blau's (1977) index of heterogeneity. That is, we define a time-varying variable named *heterogeneity of fields* as $1 - HHI$, where HHI is the Herfindahl-Hirschman index of the fields of the researcher's publications in the last five years.[25] We use both the *heterogeneity of fields* of the team leader, the PI, as well as that of the whole team, computed as the average of the diversity of fields of all the members of the team.

We proxy for the level of interpersonal diversity of knowledge of a team by measuring the contribution of the rest of the team to the scope of knowledge of the PI. That is, we define a variable named *num additional fields*, which counts the number of fields in which the rest of the team has published but the PI has not. We use this variable in conjunction with another one that counts the number of fields of the PI (*num fields of the PI).*

To illustrate the difference between our measures of intra and interpersonal 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 intrapersonal diversity of knowledge because it is composed of two specialists, whereas it exhibits positive levels of interpersonal diversity as the researchers work in different fields. Accordingly, the *heterogeneity of fields* variable of each researcher, as well as of the team, 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 is intrapersonally diverse, as the researchers are interdisciplinary, but is not interpersonally diverse, as the researchers work in exactly the same fields. Accordingly, the *heterogeneity of fields* of each of the team members, as well as that of the team, would be 0.5, whereas the *num additional fields* would be 0.[26]

[24] 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.

[25] The Herfindahl-Hirschman index is defined as the sum of the squares of the fraction of publications in each of the broad fields.

[26] Notice that an alternative measure of team diversity would be $1-HHI$, where HHI is the Herfindahl-Hirschman index of the fields of the whole team. This measure, however, would not distinguish between intra and interpersonal level of diversity. It gives the same value (0.5) to the teams in the two examples.

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 intrapersonal diversity of skills of the PI using the index of heterogeneity of the types of her publications (basic and applied), which allows us to construct the variable *heterogeneity of research types.*[27] As a measure of interpersonal diversity of skills, we 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 intrapersonal 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 decompose this 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 interpersonal 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).

Demographic diversity We construct the following proxies of demographic diversity. Building again on the idea of underrepresentation and salience, we identify whether the researchers in a team are women. As measures of intrapersonal demographic diversity, we first construct a dummy variable, named *dum female,* which takes a value of 1 if the PI is a woman. We also construct the continuous variable *ratio female*, defined as the fraction of women in the team.

We build two measures of interpersonal demographic diversity. We define a variable named *mixed academic rank*, which represents the balance between junior (academic rank levels 1 and 2) and senior researchers (levels 3 and 4) in the team. In particular, we compute the ratio of the number of junior researchers relative to all researchers, as well as the ratio of senior researchers relative to all researchers, and take the minimum of these two ratios. This variable takes a value of 0 if all have identical academic rank and reaches a value of 0.5 if the number of juniors and seniors

[27] 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.

are the same. We also construct the *coefficient of variation of academic age,* dividing the standard deviation of academic age across team members by the mean level of academic age.

Team size A variable that measures the size of the team may capture, when used in conjunction with the other variables, residual levels of (interpersonal) diversity (structural and/or demographic). 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 Table 2a, the percentage of applications that are awarded funding is 33.8%, out of which 25% are considered successful. The average number of normalized past citations is 0.37 and the average research type is 0.25. Applications have an average duration of 2.75 years and the amount requested per capita for the whole duration of the project is £128,000. The average number of members in a team is 2.5.

In terms of structural diversity, the average index of heterogeneity of fields is 0.51, and the average coefficient of variation of research type is 0.52. The percentage of researchers with a Ph.D. from an outside institution in an application is, on average, 70% and the average number of the Ph.D. origins outside the hosting institution in a team is 1.1. The coefficient of variation of past citations is 0.173 on average. In terms of demographic diversity, the coefficient of variation of academic age is, on average, 0.54. Only 6.5% of our academics are women.

[Insert Table 2 here]

Around 77% 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). Table 2b 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.

5 Empirical 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 may be measured individually (intrapersonal levels of diversity) or as a team (interpersonal levels of diversity). We thus consider the effects of the intrapersonal levels of the PI and of the intra and interpersonal 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 bootstrapped standard errors.

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 intrapersonal measures of diversity of the team leader, the PI, in Table 3. Columns 1 and 2 show that the level of interdisciplinarity of the PI has a negative and significant influence on the likelihood of success in the grant application process.[28] More diverse PIs in terms of knowledge and/or skills are more likely to find their application rejected. Having been educated in another university also hinders the likelihood of success, as shown in columns 3 and 4. The (unreported) marginal impact 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. Our sole measure of intrapersonal demographic diversity, which is whether the PI is a woman, is not significant, as is shown in column 5. All previous results still hold in the fully specified model, presented in column 6.

[Insert Table 3 here]

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), are important determinants of success in the EPSRC application process. More applied academics find it more difficult to obtain financing. Experience in previous applications 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 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 is larger.

Table 4 presents the results for the whole team. First notice that the effect of the measures of intrapersonal level of diversity of the whole team are similar to those of the team leader. Namely,

[28] Unreported regressions show that *num fields PI* is not significant (although the correlation between this variable and *heterogeneity of fields* is 0.75). This suggests that an academic with a strong background in one field who has occasionally contributed to other fields is not penalized in the grant allocation process, whereas an academic with a balanced contribution in several fields is.

greater diversity of knowledge and a larger share of outsiders from non-US universities have a negative influence in the likelihood of being awarded funding. Gender does not have a significant influence. All regressions also include the number of members of the team 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, 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 interpersonal measures of diversity in other columns.

[Insert Table 4 here]

Columns 2 to 8 analyze the effect of the interpersonal diversity variables on the teams with at least two members. Notice that the effect of these measures are computed in addition to the effect of the intrapersonal measures of diversity and of the measures of team size. Still, results are remarkably consistent. Columns 2 and 3 show that the level of interpersonal 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, and the coefficient of variation of type, are negative and significant. Column 4 shows that educational diversity also reduces the likelihood of getting the application funded. The number of Ph.D. origins is negative and significant. Column 5 reports that teams that exhibit more diversity in scientific ability, proxied by the coefficient of variation of the number of citations, are less likely to be successful than more homogeneous teams. As in the case of the intrapersonal measure of demographic diversity, columns 6 and 7 show that the proxies for interpersonal demographic diversity, in terms of academic age and tenure, are not significant. Column 8 confirms that all the previous effects on structural diversity hold in the fully specified model, except for the variable that proxies for the interpersonal level of educational diversity (*num PhD outside origins*) which is no longer significant when added alongside the variables that proxy for the intrapersonal level of educational diversity (the "*outsider*" dummies).

Taken together, the results of tables 3 and 4 suggest that structurally diverse teams are less likely to see their proposals awarded funding than non-structurally diverse teams. Demographic diversity, instead, does not affect the likelihood of getting the application awarded funding in a significant manner. Therefore, these results provide support for Hypothesis 1 but not for Hypothesis 3.

5.2 Ex-post performance

Table 5 presents the results as to whether diverse teams, both in terms of structural and demographic diversity, 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.[29]

[Insert Table 5 here]

While the first row shows that the level of intrapersonal diversity of knowledge does not have a significant influence on ex-post performance, column 2 points at a significant positive effect of the interpersonal diversity in knowledge (*num additional fields*). On the other hand, interpersonal diversity in skills does not seem to have a significant impact on ex-post performance, as shown in column 3. Our results suggest that educational diversity has a positive impact on success. Three out of the four of our intrapersonal measures (the *outsider* variables) are significantly positive in most regressions (rows 2 to 5). Moreover, our interpersonal measure (*num PhD outside origins)* is positive and significant in column 4. Diversity in ability, on the other hand, does not seem to have a significant impact on the likelihood of success. Finally, none of the variables that proxy for demographic diversity (row 6 and columns 6 and 7) is significant. All the regressions suggest that the size of the team has a non-linear effect on ex-post success, as suggested by the team science literature. The estimates, however, are not always significant as the interpersonal diversity variables seem to capture some of these effects.

Taken together, the results of Table 5 suggest that structurally diverse teams are generally more likely to be successful in terms of ex-post performance than non-structurally diverse teams. Demographic diversity, instead, does not significantly affect the likelihood of being successful. Therefore, the results provide support for Hypothesis 2 but not for Hypothesis 4.

5.3 Mediating effects

Our empirical results so far stress that structurally 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. In this section, we investigate which characteristics of the team leader mitigate or amplify these effects. We concentrate in particular on whether the team is led by a PI that is "prestigious" in terms of academic rank. We then discuss how these results change if we use ability or academic age instead of academic rank as mediating factors.

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 structural 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 structural diversity and control variables, including the academic rank (column 8 in tables 4 and 5). A

[29] 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. This may mean that we have managed to include the right variables for our models and left few unobservable variables that affect the outcome.

positive (resp., negative) coefficient of the interaction term in the award decision indicates that the negative impact of structural 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 structural diversity. The coefficients of the interaction terms are all positive and (except for one of the four variables in column 3) highly significant. Moreover, the sum of the coefficient of the interaction term and main effect is often close to zero, indicating that the negative effect of the structural 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 structural diversity on the likelihood that the team will be successful.[30]

The results reported in Table 6 are robust in several dimensions. First, the interactions with the other structural 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 structural 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 transformative research. We develop a general framework that

[30] Similar results are obtained if we interact these variables with the regressions that include the measures of interpersonal structural diversity one by one, i.e., columns $1, 2, 3$, and 4 of tables 4 and 5.

compares the drivers of success in the ex-ante award decision process against the drivers of success 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 structurally 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 intra or interpersonal levels of team diversity, and independently if the intrapersonal levels of diversity are measured at the PI or at the whole team level. On ex-post performance, our results suggest that structurally diverse teams are generally more likely to be successful. Teams that exhibit greater diversity in interpersonal levels of knowledge and intra or interpersonal levels of education are significantly more likely to generate transformative research. This indicates that the EPSRC is biased against structural diversity. 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 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 structural diverse teams, we do find what we can call “weak” evidence of a bias. Diversity in skills and scientific ability significantly reduce the likelihood of funding despite not significantly affecting the likelihood of generating transformative research.

We provide two possible explanations for the bias against structural 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 structurally 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 the academic prestige of the team leader may affect these two explanations differently. 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. Instead, a structurally diverse team led by a prestigious PI is not necessarily less complex to evaluate than a structurally diverse team led by a less prestigious PI. If safety is more of an issue than complexity, we expect the effects of structural diversity to be less strong for more prestigious than for less prestigious PIs.

Our results on the regressions with mediating effects indeed suggest that the prestige of the PI mitigates the negative impact of structural diversity in the award regressions. In fact, the negative effect of structural 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 structural diversity on ex-post performance. All these results are true independently if we measure prestige by academic rank or scientific ability. But, the mediating effects are insignificant if we replace prestige with academic age. This means that it is the job-relevant characteristics, not the demographic ones, that mediate in the effects of structural diversity.

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 structural diversity in funded research teams, and the resulting reduction of transformative 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).

Our empirical results confirm our hypotheses on structural diversity but not those on demographic diversity. Demographic diversity, either in terms of gender, age, or academic rank, does not significantly affect the likelihood of grant approval or the likelihood of ex-post success. We thus fail to find evidence, either strong or weak, of a bias in favor of demographic diversity. To our knowledge, the EPSRC does not have an explicit mechanism in place to achieve gender balance as the EU Horizon 2020 Programme does. Our results suggest that without explicit mechanisms in place, evaluation processes may not end up positively discriminating in favor of minorities.

There may, of course, be reasons (social, political, etc.) to support biases in favor of demographic diversity. Socially, we may want to favor female applicants to achieve gender balance, or to promote mixed teams of junior and senior researchers to promote knowledge exchange and capacity building. In terms of structural diversity, at least we may not want to be biased against it. Our empirical results suggest that we may be failing to achieve both of these objectives. In this sense, our paper justifies the emergence of various mechanisms recently put in place to support different types of structural and demographic diversity. Examples of such mechanisms include the Interdisciplinary Research program of the NIH to promote interdisciplinary collaborations and the EU Horizon 2020 Programme to promote positive gender discrimination.

Our approach can of course be used to analyze whether other funding agencies are biased against diverse teams. But our conceptual framework can also be used to test for bias against other attributes. With respect to our control variables, we find that team ability and seniority, as well as a university's eminence, are important determinants of success in the EPSRC application process. But ability and university eminence also improve the likelihood of ex-post success. We find weak evidence of a bias only 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.

7 References

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

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. (2016) "Key Success Drivers in Public Research Grants: Funding the Seeds of Radical Innovation in Academia?" Working Paper.

Banal-Estañol, A., Macho-Stadler, I., Pérez-Castrillo, D. (2017) "Endogeneous Matching in University-Industry Collaboration," forthcoming *Management Science*.

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 Jour-*

nal 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. (2017) "Microgeography and the Direction of Inventive Activity," forthcoming *Management Science*.

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.

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

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.

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.

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.

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

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.

Lee, J., Sugimoto, C. R., Zhang, G., Cronin, B. (2013) "Bias in Peer Review," *Journal of the American Society for Information Science and Technology* 64(1): 2-17.

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 (1991) "Exploration and Exploitation in Organizational Learning," *Organization Science* 2(1): 71-87.

Markham, C., et al. (2013) "Making Your Research Project Happen-Challenges, Facilitators and Tips for Planning," *Journal of Clinical and Preventative Cardiology* 2(3): 162-167.

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.

Miller, C.C., Burke, L.M., Glick, W.H. (1998) "Cognitive Diversity among upper-echelon Executives: Implications for Strategic Decision Processes," *Strategic Management Journal*: 39-58.

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.

O'Reilly III, C.A., Caldwell, D.F., Barnett, W.P. (1989) "Work Group Demography, Social Integration, and Turnover," *Administrative Science Quarterly*: 21-37.

Pelled, L.H., Eisenhardt, K.M., Xin, K.R. (1999) "Exploring the Black Box: An Analysis of Work Group Diversity, Conflict and Performance," *Administrative Science Quarterly* 44(1): 1-28.

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

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

Rees, T. (2011) "The Gendered construction of Scientific Excellence," *Interdisciplinary Science Reviews* 36(2): 133-145.

Rodan, S., Galunic, C. (2004) "More than Network Structure: How Knowledge Heterogeneity influences Managerial Performance and Innovativeness," *Strategic Management Journal* 25(6): 541-562.

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

Sandström, U., Hallsten, M. (2008) "Persistent Nepotism in Peer-Review," *Scientometrics* 74(2): 175-189

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.

Stvilia, B., et al. (2011) "Composition of Scientific Teams and Publication Productivity at a

National Science Lab," *Journal of the American Society for Information Science and Technology* 62(2): 270-283.

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.

Tsui, A. S., Egan, T. D., O'Reilly III, C. A. (1992) "Being different: Relational demography and organizational attachment," *Administrative Science Quarterly*: 549-579.

Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., Wetherell, M.S. (1987) *Rediscovering the Social Group: A self-categorization theory.* Basil Blackwell.

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.

Watson, W. E., Kumar, K., and Michaelsen, L. K. (1993) "Cultural Diversity's Impact on Interaction Process and Performance: Comparing Homogeneous and Diverse Task Groups," *Academy of Management Journal* 36: 590-602.

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.

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 education, DA = diversity in scientific ability, and DD = demographic diversity. (intra) means intrapersonal diversity and (inter) means interpersonal diversity.

Name of variable	**Definition of variable**	
Award	dummy equal to 1 if the applicaiton 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 age	difference between the year of the application and the date of the PhD	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
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 PI	# of fields of the publications of the PI	C
Heterogeneity of fields	1 -HHI index of the fields of the publications	DKS (intra)
Num additional fields	# of fields where the team has published but the PI has not	DKS (inter)
Heterogeneity of research types	1 -HHI index of the types of the publications	DKS (intra)
Coef-var research type	normalized std deviation of type of research of team members	DKS (inter)
Ratio PhD outside	fraction of PhD degrees from different than the current uni	DE (intra)
Ratio PhD US	fraction of PhD degrees in the US	DE (intra)
Ratio PhD foreign non-US	fraction of PhD degrees in a foreign country different from the US	DE (intra)
Ratio PhD outside RG	fraction of PhD degrees in a UK Russell group uni different from holding uni	DE (intra)
Ratio PhD outside non-RG	fraction of PhD degrees in a UK non-Russell group uni different from holding uni	DE (intra)
Num PhD outside origins	# of institutions in which members have PhD from other than the holding uni	DE (inter)
Coef-var citations	normalized std deviation of citations of the team members	DA (inter)
Ratio female	fraction of females in the team	DD (intra)
Mixed academic rank	minimum {ratio # of academic rank levels 1 and 2 / total, 1 minus this ratio}	DD (inter)
Coef-var academic age	normalized std deviation of academic age of the team members	DD (inter)
Num team members	sum of the # of coinvestigators and the PI	

Table 2a. Descriptive statistics

We report the descriptive statistics for the dependent, structural and demographic diversity, and control variables.

Variables	Observations	Mean	St dev	Median
Awarded	18,572	.338	.473	0
Citations	18,572	.366	5.919	1.72
Research type	18,572	.254	.324	.1
Application experience	18,572	1.257	1.168	1
Academic rank	18,572	2.788	1.023	3
Academic age	18,572	17.733	7.889	17
Russell group	18,572	.774	.418	1
Fraction awarded	18,572	.32	.083	.306
Duration (in years)	18,572	2.753	1.014	3
Funds per cap (in Million £)	18,572	.128	.275	.0825

Variables	Observations	Mean	St dev	Median
Num team members	18,572	2.5	1.631	2
Heterogeneity of fields	18,547	.511	.221	.58
Num additional fields	15,462	.610	1.219	0
Coef-var research type	10,451	.52	.64	.09
Ratio PhD outside	17,803	.698	.405	1
Num PhD outside origins	17,859	1.161	.877	1
Coef-var citations	17,674	.173	.227	.01
Ratio female	17,954	.065	.208	0
Coef-var academic age	9,348	.539	.503	.48
Mixed academic rank	18,572	.134	.205	0

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

Russell Group	**Number of applications**	**Fraction awarded**
University of Birmingham	809	0.35
University of Bristol	410	0.36
University of Cambridge	1,089	0.44
Cardiff University	346	0.22
Durham University	194	0.29
University of Edinburgh	463	0.32
University of Exeter	153	0.29
University of Glasgow	508	0.34
Imperial College London	1,560	0.38
King's College London	192	0.28
University of Leeds	965	0.35
University of Liverpool	511	0.34
The University of Manchester*	1,526	0.33
Newcastle University	654	0.32
University of Nottingham	862	0.33
University of Oxford	681	0.40
Queen Mary	381	0.35
Queen's University of Belfast	406	0.30
University of Sheffield	1,085	0.36
University of Southampton	670	0.34
University College London	519	0.35
University of Warwick	264	0.34
University of York	126	0.32
Total	14,374	0.35

Non-Russell Group	**Number of applications**	**Fraction awarded**
University of Aberdeen	132	0.28
Aston University	187	0.33
Brunel University	290	0.17
City University	189	0.31
University of Dundee	183	0.32
University of Essex	141	0.33
University of Hull	138	0.28
Heriot-Watt University	410	0.27
Lancaster University	59	0.36
University of Leicester	199	0.28
Loughborough University	994	0.31
University of Reading	55	0.24
University of Salford	172	0.38
University of Strathclyde	550	0.29
Swansea University	376	0.39
University of Wales, Bangor	123	0.29
Total	4,198	0.30

Table 3. Likelihood of award against intrapersonal level of diversity of the PI
This table reports the results of the probit regressions for the probability of an application being awarded against measures of intrapersonal diversity of the PI and control variables. Year fixed effects are also included in all regressions. Standard errors are bootstrapped.

	(1)	(2)	(3)	(4)	(5)	(6)
Intrapersonal diversity						
Heterogeneity of fields	-0.079**					-0.129***
	[0.039]					[0.046]
Heterogeneity of research types		-0.135***				-0.102**
		[0.045]				[0.050]
Dum PhD outside			-0.088***			
			[0.023]			
Dum PhD US				0.000		-0.013
				[0.066]		[0.075]
Dum PhD foreign non-US				-0.159***		-0.160***
				[0.044]		[0.051]
Dum PhD outside RG				-0.082***		-0.085***
				[0.022]		[0.026]
Dum PhD outside non-RG				-0.101***		-0.100***
				[0.036]		[0.029]
Dum female					0.032	0.018
					[0.058]	[0.041]
Controls						
Citations	0.008***	0.008***	0.011***	0.011***	0.008***	0.011***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Research type	-0.072**	-0.049	-0.100***	-0.100***	-0.073**	-0.085**
	[0.031]	[0.034]	[0.039]	[0.030]	[0.033]	[0.034]
Academic rank	0.076***	0.078***	0.072***	0.070***	0.075***	0.072***
	[0.009]	[0.008]	[0.011]	[0.011]	[0.010]	[0.010]
Application experience	-0.006	-0.006	-0.010	-0.010	-0.009	-0.010
	[0.010]	[0.012]	[0.013]	[0.010]	[0.010]	[0.011]
Russell group	0.146***	0.146***	0.139***	0.138***	0.130***	0.127***
	[0.027]	[0.027]	[0.022]	[0.024]	[0.028]	[0.025]
Fraction awarded	1.532***	1.525***	1.528***	1.531***	1.466***	1.471***
	[0.155]	[0.151]	[0.134]	[0.160]	[0.166]	[0.137]
Duration	-0.137***	-0.136***	-0.132***	-0.132***	-0.137***	-0.134***
	[0.011]	[0.012]	[0.013]	[0.013]	[0.043]	[0.014]
Funds per cap	-0.048	-0.049	-0.054	-0.054	-0.042	-0.047
	[0.064]	[0.076]	[0.079]	[0.079]	[0.083]	[0.096]
Constant	-0.779***	-0.801***	-0.742***	-0.742***	-0.763***	-0.643***
	[0.099]	[0.079]	[0.089]	[0.088]	[0.079]	[0.093]
Observations	15.286	15.305	14.246	14.246	14.407	13.554

*** p<0.01, ** p<0.05, * p<0.1

Table 4. Likelihood of award against intrapersonal and interpersonal level of diversity of the team

This table reports the coefficients of the probit model for the award decision against measures of intra and interpersonal diversity of the team and control variables. Year fixed effects are also included in all regressions. Standard errors are bootstrapped.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intrapersonal diversity								
Heterogeneity of fields	-0.132***		-0.168*	-0.048	-0.079	-0.085	-0.051	
	[0.040]		[0.086]	[0.064]	[0.071]	[0.076]	[0.066]	
Ratio PhD US	-0.053	0.108	0.143		0.024	0.008	0.038	0.187
	[0.064]	[0.130]	[0.179]		[0.136]	[0.126]	[0.116]	[0.130]
Ratio PhD foreign non-US	-0.232***	-0.340***	-0.401***		-0.369***	-0.348***	-0.351***	-0.453***
	[0.050]	[0.101]	[0.114]		[0.088]	[0.082]	[0.078]	[0.123]
Ratio PhD outside RG	-0.099***	-0.111**	-0.168***		-0.152***	-0.121***	-0.132***	-0.243***
	[0.026]	[0.051]	[0.053]		[0.041]	[0.042]	[0.039]	[0.099]
Ratio PhD outside non-RG	-0.071*	-0.241***	-0.247***		-0.234***	-0.228***	-0.221***	-0.325***
	[0.039]	[0.064]	[0.091]		[0.080]	[0.068]	[0.065]	[0.109]
Ratio female	-0.018	-0.012	-0.047	-0.052	-0.004	-0.096	-0.061	
	[0.048]	[0.083]	[0.116]	[0.078]	[0.091]	[0.083]	[0.070]	
Interpersonal diversity								
Num team members	-0.105***	-0.023	-0.059	-0.066*	-0.084**	-0.079*	-0.084***	-0.008
	[0.026]	[0.041]	[0.042]	[0.036]	[0.035]	[0.042]	[0.028]	[0.053]
Num team members squared	0.012***	0.003	0.008*	0.008***	0.009**	0.009**	0.009***	0.001
	[0.003]	[0.004]	[0.004]	[0.003]	[0.004]	[0.004]	[0.003]	[0.005]
Num additional fields		-0.060***						-0.086***
		[0.014]						[0.015]
Coef-var research type			-0.105***					-0.080*
			[0.040]					[0.040]
Num PhD outside origins				-0.041***				0.048
				[0.015]				[0.031]
Coef-var citations					-0.151**			-0.212**
					[0.066]			[0.097]
Coef-var academic age						-0.011		
						[0.036]		
Mixed academic rank							0.010	
							[0.068]	
Controls								
Citations	0.006***	0.006**	0.003	0.006**	0.007**	0.008**	0.006**	0.004
	[0.002]	[0.003]	[0.004]	[0.002]	[0.003]	[0.003]	[0.003]	[0.004]
Research type	-0.119***	-0.078*	-0.335***	-0.060	-0.064	-0.066	-0.065*	-0.352***
	[0.035]	[0.045]	[0.102]	[0.045]	[0.056]	[0.042]	[0.037]	[0.088]
Application experience	0.008	0.011	0.017	0.024	0.019	0.024**	0.022	0.016
	[0.010]	[0.018]	[0.025]	[0.015]	[0.019]	[0.012]	[0.017]	[0.029]
Academic rank	0.075***	0.134***	0.115***	0.117***	0.126***	0.127***	0.118***	0.130***
	[0.010]	[0.018]	[0.027]	[0.017]	[0.017]	[0.026]	[0.020]	[0.024]
Russell group	0.139***	0.107***	0.083*	0.119***	0.119***	0.116***	0.114***	0.086*
	[0.022]	[0.041]	[0.046]	[0.031]	[0.029]	[0.036]	[0.036]	[0.048]
Fraction awarded	1.585***	1.710***	1.442***	1.649***	1.692***	1.674***	1.652***	1.600***
	[0.150]	[0.249]	[0.227]	[0.219]	[0.231]	[0.207]	[0.221]	[0.293]
Duration	-0.098***	-0.002	0.005	-0.007	0.008	0.002	-0.007	-0.002
	[0.013]	[0.020]	[0.024]	[0.019]	[0.017]	[0.017]	[0.017]	[0.021]
Funds per cap	-0.132	-0.875***	-1.022***	-0.997***	-1.029***	-1.027***	-1.015***	-0.836***
	[0.109]	[0.201]	[0.234]	[0.146]	[0.169]	[0.209]	[0.175]	[0.207]
Num fields PI		0.002						-0.016
		[0.010]						[0.012]
Academic age						-0.004		
						[0.003]		
Constant	-0.609***	-1.100***	-0.744***	-1.026***	-0.955***	-0.941***	-0.955***	-0.858***
	[0.083]	[0.156]	[0.204]	[0.144]	[0.171]	[0.166]	[0.135]	[0.211]
Observations	17,375	7,945	5,702	9,694	8,815	9,215	9,694	5.116

*** p<0.01, ** p<0.05, * p<0.1

Table 5. Likelihood of ex-post success against intrapersonal and interpersonal level of diversity of the team								
This table reports the coefficients of the Heckman probit selection model for the ex-post success of the awarded projects against measures of intra and interpersonal diversity of the team and control variables. Year fixed effects are also included in all regressions. Standard errors are bootstrapped.								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intrapersonal diversity								
Heterogeneity of fields	-0.121		-0.138	0.007	0.030	-0.104	-0.042	
	[0.156]		[0.422]	[0.205]	[0.315]	[0.318]	[0.264]	
Ratio PhD US	0.586**	0.733	0.996*		1.117***	1.118***	1.052***	0.667*
	[0.197]	[0.542]	[0.555]		[0.413]	[0.479]	[0.339]	[0.369]
Ratio PhD foreign non-US	0.572***	0.494*	0.396		0.610**	0.558*	0.569**	0.755***
	[0.166]	[0.259]	[0.387]		[0.272]	[0.294]	[0.265]	[0.247]
Ratio PhD outside RG	0.223*	0.401**	0.292		0.320	0.357*	0.356**	0.395**
	[0.121]	[0.191]	[0.242]		[0.199]	[0.143]	[0.171]	[0.188]
Ratio PhD outside non-RG	0.008	0.214	0.068		0.275	0.239	0.246	0.168
	[0.133]	[0.206]	[0.368]		[0.225]	[0.255]	[0.241]	[0.293]
Ratio female	-0.120	-0.262	0.009	-0.167	-0.195	-0.220	-0.147	
	[0.166]	[0.306]	[0.431]	[0.269]	[0.267]	[0.272]	[0.306]	
Interpersonal diversity								
Num team members	0.160**	0.042	0.229	0.075	0.096	0.134	0.139	0.189*
	[0.079]	[0.102]	[0.157]	[0.112]	[0.122]	[0.133]	[0.109]	[0.111]
Num team members squared	-0.012	0.001	-0.017	-0.008	-0.007	-0.009	-0.010	-0.007
	[0.009]	[0.009]	[0.014]	[0.011]	[0.011]	[0.013]	[0.011]	[0.010]
Num additional fields		0.078*						0.026
		[0.045]						[0.061]
Coef-var research type			-0.140					-0.003
			[0.167]					[0.134]
Num PhD outside origins				0.145***				-0.092
				[0.051]				[0.091]
Coef-var citations					-0.232			-0.250
					[0.265]			[0.277]
Coef-var academic age						-0.196		
						[0.188]		
Mixed academic rank							-0.090	
							[0.262]	
Controls								
Citations	0.021***	0.023*	0.017	0.032***	0.031***	0.034**	0.031**	0.009
	[0.007]	[0.012]	[0.017]	[0.010]	[0.011]	[0.014]	[0.014]	[0.010]
Research type	-0.310**	-0.314	-0.716*	-0.403**	-0.404*	-0.546***	-0.371**	-0.400
	[0.143]	[0.232]	[0.434]	[0.204]	[0.242]	[0.201]	[0.188]	[0.315]
Application experience	-0.008	-0.075*	0.043	-0.039	-0.047	-0.031	-0.029	0.016
	[0.033]	[0.044]	[0.072]	[0.049]	[0.051]	[0.058]	[0.044]	[0.048]
Academic rank	-0.061	-0.181**	-0.132	-0.101	-0.099	-0.059	-0.105	-0.124
	[0.044]	[0.076]	[0.138]	[0.094]	[0.107]	[0.121]	[0.093]	[0.079]
Russell group	0.273**	0.014	0.094	0.115	0.152	0.141	0.133	0.216
	[0.130]	[0.201]	[0.189]	[0.169]	[0.172]	[0.172]	[0.155]	[0.241]
Duration	0.204***	0.188***	0.223***	0.218***	0.231***	0.230***	0.215***	0.286***
	[0.050]	[0.073]	[0.071]	[0.071]	[0.057]	[0.062]	[0.077]	[0.066]
Funds per cap	0.482**	1.515*	1.171	1.355	1.181*	1.181	1.271	0.687
	[0.238]	[0.785]	[0.881]	[0.864]	[0.633]	[0.872]	[0.803]	[0.505]
Num fields PI		0.024						-0.003
		[0.027]						[0.036]
Academic age						-0.013		
						[0.009]		
Constant	-2.178***	-0.238	-1.793*	-1.550*	-1.709*	-1.713*	-1.839**	-1.906**
	[0.306]	[1.210]	[0.949]	[0.940]	[0.895]	[0.888]	[0.838]	[0.919]
rho	0.314	-0.865	0.205	0.056	0.087	0.239	0.164	-0.079
	[2.365]	[4.160]	[2.660]	[0.918]	[0.943]	[2.389]	[3.862]	[3.993]
Uncensored observations	1.438	653	480	769	717	738	769	715
Observations	5,228	2,414	1,813	2,865	2,685	2,747	2,865	2,694

*** p<0.01, ** p<0.05, * p<0.1

Table 6. Mediating effects

This table reports in panel A the result of the probit regressions for the probability of an application being awarded, and in panel B the coefficients of the Heckman probit selection model for the ex-post success of the awarded projects, against measures of intra and interpersonal diversity of the team, the interaction between the dummy variable Professor and the structural diversity variables, and the control variables. Year fixed effects are also included in all regressions. Standard errors are bootstrapped.

	Panel A: Award				Panel B: Ex-post success			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main effects								
Ratio PhD US	0.189	0.212	0.102	0.188	1.053*	1.073*	0.713	1.057*
	[0.176]	[0.182]	[0.222]	[0.166]	[0.561]	[0.615]	[5.573]	[0.636]
Ratio PhD foreign non-US	-0.455***	-0.444***	-0.699***	-0.445***	0.644	0.599	0.942	0.670
	[0.114]	[0.132]	[0.145]	[0.119]	[0.415]	[0.463]	[0.619]	[0.419]
Ratio PhD outside RG	-0.237**	-0.230**	-0.300***	-0.236**	0.624**	0.608*	0.754**	0.655**
	[0.094]	[0.099]	[0.088]	[0.095]	[0.298]	[0.351]	[0.358]	[0.323]
Ratio PhD outside non-RG	-0.318***	-0.308***	-0.447***	-0.313***	0.464	0.453	0.466	0.451
	[0.105]	[0.113]	[0.109]	[0.105]	[0.346]	[0.417]	[0.576]	[0.345]
Num team members	-0.008	-0.013	-0.016	-0.018	0.137	0.137	0.151	0.155
	[0.052]	[0.059]	[0.055]	[0.050]	[0.148]	[0.164]	[0.192]	[0.160]
Num team members squared	0.001	0.001	0.002	0.002	-0.002	-0.002	-0.003	-0.004
	[0.006]	[0.006]	[0.006]	[0.005]	[0.014]	[0.016]	[0.016]	[0.014]
Num additional fields	-0.106***	-0.079***	-0.081***	-0.080***	0.102	0.058	0.056	0.059
	[0.015]	[0.015]	[0.017]	[0.015]	[0.089]	[0.064]	[0.067]	[0.060]
Coef-var research type	-0.081**	-0.154***	-0.083**	-0.084**	0.002	0.014	0.012	0.016
	[0.035]	[0.051]	[0.042]	[0.043]	[0.172]	[0.219]	[0.153]	[0.133]
Num PhD outside origins	0.045	0.042	0.043	0.047	-0.112	-0.105	-0.108	-0.122
	[0.028]	[0.036]	[0.033]	[0.032]	[0.098]	[0.123]	[0.102]	[0.101]
Coef-var citations	-0.202*	-0.226**	-0.217**	-0.378***	-0.300	-0.306	-0.267	0.076
	[0.104]	[0.108]	[0.085]	[0.136]	[0.431]	[0.436]	[0.344]	[0.526]
Interactions								
Num additional fields*Professor	0.053***				-0.071			
	[0.019]				[0.087]			
Coef-var research type*Professor		0.141***				-0.039		
		[0.037]				[0.161]		
Ratio PhD US*Professor			0.238				0.353	
			[0.284]				[5.687]	
Ratio PhD foreign non-US*Professor			0.527***				-0.530	
			[0.181]				[0.640]	
Ratio PhD outside RG*Professor			0.153*				-0.267	
			[0.084]				[0.257]	
Ratio PhD outside non-RG*Professor			0.286**				-0.032	
			[0.137]				[0.590]	
Coef-var citations*Professor				0.340***				-0.636
				[0.127]				[0.399]
Control Variables	YES	YES	YES	YES	YES	YES	YES	YES
rho					-0.492	-0.492	-0.537	-0.627
					[2.402]	[2.402]	[3.726]	[1.269]
Uncensored observations					428	428	428	428
Observations	5,116	5,116	5,116	5,116	1,610	1,610	1,610	1,610

*** p<0.01, ** p<0.05, * p<0.1