Fontana, M., Iori, M., Sciabolazza, V. L., & Souza, D. (2022). The interdisciplinarity dilemma: Public versus private interests. *Research Policy*, *51*(7), 104553. Published online 27 May 2022. doi: 10.1016/j.respol.2022.104553

Published Journal Article available at: https://doi.org/10.1016/j.respol.2022.104553

The interdisciplinarity dilemma: public versus private interests

Magda Fontana[a,b], Martina Iori[c], Valerio Leone Sciabolazza[d], and Daniel Souza[b]

[a]Despina Big Data Lab, University of Turin, Lungo Dora Siena, 100A, 10153, Turin, Italy
[b]Department of Economics and Statistics, University of Turin, Lungo Dora Siena, 100A, 10153, Turin, Italy
[c]Institute of Economics & EMbeDS, Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà, 33, 56127, Pisa, Italy
[d]Department of Economics and Law, Sapienza University of Rome, Via del Castro Laurenziano 9, 00161, Rome, Italy

Abstract

Researchers often receive contrasting incentives when conducting their work. On the one hand, an interdisciplinary approach is required to produce scientific advances and access to funding. On the other, academic scholarships and evaluation mechanisms are still organized following the criteria of traditional disciplinary fields. If pursuing interdisciplinary research results in contrasting outcomes, science may face an interdisciplinarity dilemma: should researcher pursue their own private interest to build a reputation? Or should they endeavor towards public interest? How costly in terms of reputation is to choose interdisciplinarity research (IDR) over (more) specialized research? We answer these questions by exploiting data on 23,926 articles published by 6,105 researchers affiliated with the University of Florida in the period 2008-2013. Through individual fixed-effect, we compare articles of the same scholar to roll out the influence of individual characteristics on the scientific impact of their research. We find that the diverse dimensions of IDR (Variety, Balance, and Disparity) have a different effect on the reputation of a scholar and on her contribution to societal research. We confirm the existence of trade-off between private and public interest. We also point out that the increase of IDR aiming at connecting distant disciplines reduces the usefulness of the resulting knowledge. Results are robust to various specifications and apply to all scholars, regardless of their gender, collaboration behavior, discipline, and performance. These findings pose challenging questions to policymakers.

Keywords: Interdisciplinarity; Research Policy; Academic Career; Generality; Incentives; Citations

JEL codes: I23; H5; 039

1 Introduction

In recent years, diverse patterns have emerged in science. Scientists have narrowed their expertise in response to the burden of knowledge (Jones, 2009) and rely more and more on teamwork by joining different fields of specific knowledge – interdisciplinary research (IDR) – to produce wide-ranging scientific advances (Cedrini and Fontana, 2018; Larsen and Ins, 2010). The growing importance of interdisciplinarity also results from the push of private and public funding and research institutions that find the overcoming of disciplinary barriers (Rylance, 2015) as the optimal solution to scientific and social problems. These new patterns have produced profound changes in the organization of science: universities created interdisciplinary research centers (Biancani et al., 2018; Hackett et al., 2021; Wuchty et al., 2007), and science has a whole has experienced an increasing trend of citation flows across disciplines in several fields of study (Angrist et al., 2020; Battiston et al., 2019). Moreover, studies have found that IDR is associated with more grant and patent submissions and with stable cooperation networks (Arnold et al., 2021; Jha and Welch, 2010; Singh and Fleming, 2010).

Thus, it might seem that interdisciplinarity is the optimal response to the ongoing transformation of science. On the one hand, it counteracts the effects of specialization by allowing researchers to join expertise and to face more challenging societal and scientific issues and therefore fulfilling the public interest to face the complexity of societal problems that increasingly require expertise from different fields. On the other, the relative abundance of funding available to undertake IDR (Singh and Fleming, 2010) makes interdisciplinarity a sensible option for scholars. However, recent literature (Arnold et al., 2021; The National Academies, 2005) raised doubts about the presence of a potential conflict between the private interest of researcher (career and reputation) and public interest (solution of societal issue and circulation of knowledge beyond disciplinary boundaries). Actually, the reorganization of academe towards receiving interdisciplinary is far from being completed. Scholarships and their assessment mechanisms are still organized in separated disciplines or even in subfields. The specialization of journals (Stigler et al., 1995), together with the decreasing importance of generalist journals (Goel and Faria, 2007, p. 538), suggests that academic reputation tends to be built within niches. Moreover, the increasing relevance of rankings of field-specific journals renders the interdisciplinary effort rather risky since these rankings are used to evaluate research performances of universities, departments, and individual scholars and, then, to assign funds and make hiring decisions (Cedrini and Fontana, 2018; Ritzberger, 2008).

In this paper, we aim at: i) contributing to the literature on IDR by adding to the scant evidence on the effect of IDR on the researchers' career (Leahey et al., 2017; Sun et al., 2021); ii) filling a gap in the extant analyses of the topic: previous research shows mixed evidence on how interdisciplinarity affects scientific impact – number of citations –, productivity, and research funding (Leahey et al., 2017; Sun et al., 2021) but an analysis of trade-off between private and public interests is yet to be explored.

Namely, we explore the idea that researchers often receive contrasting incentives when conducting their work. On the one hand, an interdisciplinary approach is required to produce scientific advances and access to funding. On the other, academic scholarships and evaluation mechanisms are still organized following the criteria of traditional disciplinary fields. If pursuing interdisciplinary research results in contrasting outcomes, science may face an interdisciplinarity dilemma: should researchers pursue their own private interest to build a reputation? Or should they endeavor towards public interest? How costly in terms of reputation is to choose IDR over (more) specialized research?

To investigate the trade-off, we study, at the researcher level, the effect of adopting an interdisciplinary approach: i) on the number of citations received by researcher's papers, as a proxy for reputational achievement; ii) on the circulation of researcher's papers across diverse fields, as a proxy for the public interest to face societal issues through the circulation of expertise beyond disciplinary boundaries.

Toward this purpose, we analyze a novel and unique dataset of 6,105 researchers affiliated with the University of Florida (UF) along with their publication records (23,926 articles) and individual characteristics (such as gender and affiliation) over the period 2008-2013.[1] Albeit small in comparison with the samples used in other studies (Yegros-Yegros et al., 2015), our dataset has the unique feature of providing detailed bibliometric and non-bibliometric information about a panel of scholars operating in a wide range of scientific fields and affiliated to the same university. This feature allows sorting out a number of confounding factors often neglected by the literature, as the

[1]The University of Florida is a large research university in the United States that comprises more than 5,000 researchers and 50,000 students. UF consistently ranks among the top ten public universities in the United States and is the flagship university in the state of Florida.

role played by institutional and national heterogeneity.

Thanks to the panel nature of data, we observe the variation of the degree of interdisciplinarity across articles by the same scholar.[2] With respect to extant literature (see, for instance, Yegros-Yegros et al., 2015), we account for the investigators' individual characteristics that may play a crucial role in determining the scholar's reputation. At the same time, by performing our analysis at the article level and comparing papers of the same researcher (through individual fixed effect), we avoid aggregations of data at the researcher level (Leahey et al., 2017), and we test the individual incentives in pursuing IDR. We measure the scholars' reputation by looking at the number of citations accrued by articles and their contribution to research with societal impact through articles' degree of generality. Interdisciplinarity has been intended so far uniquely as a way of combining different sources of knowledge, but, it is our conviction, that it is the circulation of such knowledge that realizes the public interest associated with IDR. As societies become more interconnected and grow in complexity, science needs to combine knowledge from different domains but also shares new findings with them. Following Carley and Porter (2012) and Fontana et al. (2020), we measure generality by calculating the dispersion of citations across disciplines through the Hirschman-Herfindahl Index.[3]

We measure interdisciplinarity by highlighting its main dimensions (Porter and Rafols, 2009; Yegros-Yegros et al., 2015): the number of fields embedded in a paper (Variety); the evenness of their distribution (Balance), and the similarity between them (Disparity).[4] The use of multiple and distinct indicators allows capturing all the facets of a complex concept like interdisciplinarity.

Our identification strategy relies primarily on the use of individual, disciplinary-based citation patterns and year fixed effects, which allow registering the effect of a change in interdisciplinarity on the scientific impact of a researcher while sorting out potential confounding factors and the influence of a change in other dimensions. The additional information contained in our database, moreover, give us the chance to shed light on different sources of heterogeneity and assess whether the impact

[2] In principle, also other datasets, such as MAG, may allow creating a longitudinal dataset about scholars using an identification code. However, such identification codes are obtained through inferential methods, and they are not directly registered by scholars or their institutions. On the contrary, our information is more reliable since the association of articles to the same scholars is done by the UF, and there is no inference involved.

[3] The index is widely applied in the economics of innovation literature to measure the range of inventions that derive from a patent (Bresnahan and Trajtenberg, 1995; Squicciarini et al., 2013).

[4] The literature also uses the Rao-Stirling diversity (Stirling, 2007), an index that synthesizes the three dimensions. In addition to the loss of details, it has been shown (Fontana et al., 2020, Figure 10) that the Rao-Stirling diversity is highly correlated with Disparity. We, therefore, decided not to include it in our analysis.

of IDR differs across gender, collaboration types, research proficiency, and disciplinary affiliation.

Our findings confirm the existence of a trade-off between private and public interests in one of the three observed interdisciplinarity dimensions. An increase in the evenness of the distribution of disciplines in article references (Balance) results in a decrease of the number of accrued citations, but increases its generality. In addition, we find that the increase of the number of disciplines recombined in a paper (Variety) has a positive effect on the number of citations and generality. This seems to signal a private incentive to and public benefit from pursuing IDR, however, when the distance of the involved disciplines (Disparity) increases both citations and generality decline. Therefore, a trade-off emerges, independently of the involved interests, among the dimensions of IDR. Importantly, results are confirmed even when considering scholars with different characteristics or affiliations. In other words, all scholars face the similar incentives and constraints in engaging in interdisciplinary projects.

This evidence suggests that much effort should be put into coordinating private and public interests by tuning hiring and rewarding mechanisms with funding policy whenever interdisciplinarity is concerned. Secondly, the private and public benefits of IDR do not grow infinitely: in spite of its undeniable importance, interdisciplinarity is not the panacea for all scientific and societal issues.

This paper aims to make three contributions. Firstly, we provide evidence on the existence of a trade-off between private (researchers) and public (society) benefits in pursuing IDR. Secondly, we propose a novel approach to analyze researchers' scientific outcomes from a micro perspective without the aggregation of bibliometric data. Finally, we introduce the generality of knowledge as an additional measure of interdisciplinarity and, at the same time, as a relevant indicator of the achievement of the public goal to obtain interdisciplinary solutions to societal problems.

The paper proceeds as follows. Section 2 presents the theoretical background and motivating evidence, while Section 3 summarizes our research hypotheses. In Sections 4 and 5, we describe our empirical strategy and data, respectively. Section 6 discusses the results, and Section 7 concludes.

2 Interdisciplinary research and researchers' incentives

A vast and growing literature has stressed the existence of *multiple logics* within the academia (Llopis et al., 2022): researchers might engage in activities that pursue rather different goals. They can engage in quasi-market actions such as academic patenting or academic entrepreneurship (Sterzi et al., 2019), they can act to increase their reputation within the academia, and, finally, they can endeavor towards research with a higher societal impact (Mazzucato, 2018).

It has been convincingly argued by Llopis et al. (2022) that the multiplicity of objectives can make it difficult for researchers to respond to conflicting incentives and that policies that sustain different logics might aggravate the issue. Several studies have explored the trade-off between market and scientific activities (see, for instance, Tartari and Breschi, 2012), while the individual and institutional tension between reputation building and societal activities remains unexplored.

In this paper, we adhere to the definition of reputation proposed by Llopis et al. (2022, p. 2): the scientist's academic status within her peer community. Reputation gives scientists recognition and leverage in competitions and funding. We assume that such status is mainly built through the publication of articles (Subramanian et al., 2013) and their subsequent citations (Hamermesh and Pfann, 2012; Jamali et al., 2016; Jones, 2021). Instead, we define societal research as the activity that tackles issues that are "complex, systemic, interconnected, and urgent, requiring insights from many perspectives" (Mazzucato, 2018, p. 803). We assume that, given its nature, societal research requires, primarily, insights from different perspectives and the subsequent circulation of the derived knowledge beyond disciplinary fields (The National Academies, 2005). We then ask if scientists can simultaneously achieve reputation – recognition in their own field – and contribute to societal research. We use the interdisciplinarity of scholars to provide them with a degree of involvement in reputation-seeking behavior and societal research.

The existing literature on IDR primarily focuses on scholars' scientific outcomes, rather than researchers themselves (Leahey and Barringer, 2020; Hackett et al., 2021), and therefore is only partially relevant to this study. Several studies highlight the mixed effect of the various aspects of interdisciplinarity on the scientific impact, measured as the number of citations received by single articles (see, among others, Fontana et al., 2020; Yegros-Yegros et al., 2015).[5] Results vary across the dimensions of IDR and disciplines took into account. However, those studies commonly identified an inverted U-shaped relationship between the interdisciplinarity and impact of an article. Moving from articles to research projects and grants, Bromham et al. (2016) suggested the existence of a

[5]For a survey of the literature on interdisciplinarity see Wagner et al. (2011), for a review on the relationship between interdisciplinarity and impact see Zeng et al. (2017, section 6.1.1).

bias against interdisciplinarity in funding evaluations.

While the effect of interdisciplinary on knowledge production and scientific impact has been extensively studied in the literature, the impact of pursuing IDR on scholars' productivity, career, and funding performance is still underexplored. The existing evidence, however, seems to highlight that IDR comes with a cost. Leahey et al. (2017) provided one of the first studies on potential scholars' costs and benefits associated with interdisciplinarity research. They collected 32,000 articles published by 854 researchers from a wide range of fields and universities. The authors computed researcher-level bibliometric indicators by considering scholars' publications in the entire period of analysis. Overall, they found that an increase in the average interdisciplinarity of scholars' work improves their visibility in the scientific community, measured as the cumulative number of citations, and decreases their productivity, as indicated by the number of articles published. Sun et al. (2021) analyze 44,419 research grant awarded by the research councils in the UK and find that interdisciplinary research is less impactful than specialized research in the short run but, eventually, is more rewarding in terms of volume and value of funding.

We are supported in our research questions by a preliminary evidence on the effect of interdisciplinarity on yearly wages and research funding (number of grants) in a subsample of scholars at the University of Florida in the period 2008–2013 (more details about data are in Section 5). Once controlled for scholars' academic age, we use their wages to represent a signal of academic reputation and the number of awarded grants to indicate their potential contribution to societal research. Then, we create an interdisciplinary profile of researchers, i.e. the extent to which they are prone to conduct interdisciplinary research, by using the maximum number of unique disciplines in the references of an article written by an investigator in one year.

Figure 1 shows that there exists a negative and statistically significant correlation between interdisciplinary and researchers' wage, while we observe a positive and statistically significant correlation between interdisciplinary and the number of grants received by researchers. In other words, scholars that conduct research in delimited fields of study receive higher wages, while more interdisciplinary researchers are awarded with more grants. This evidence thus corroborates our hypothesis that researchers receive contrasting incentives when engaging in interdisciplinary work. By increasing the interdisciplinary content of their research, scholars also increase their societal relevance and receive more grants. At the same time, this reduces scholars' reputation within their

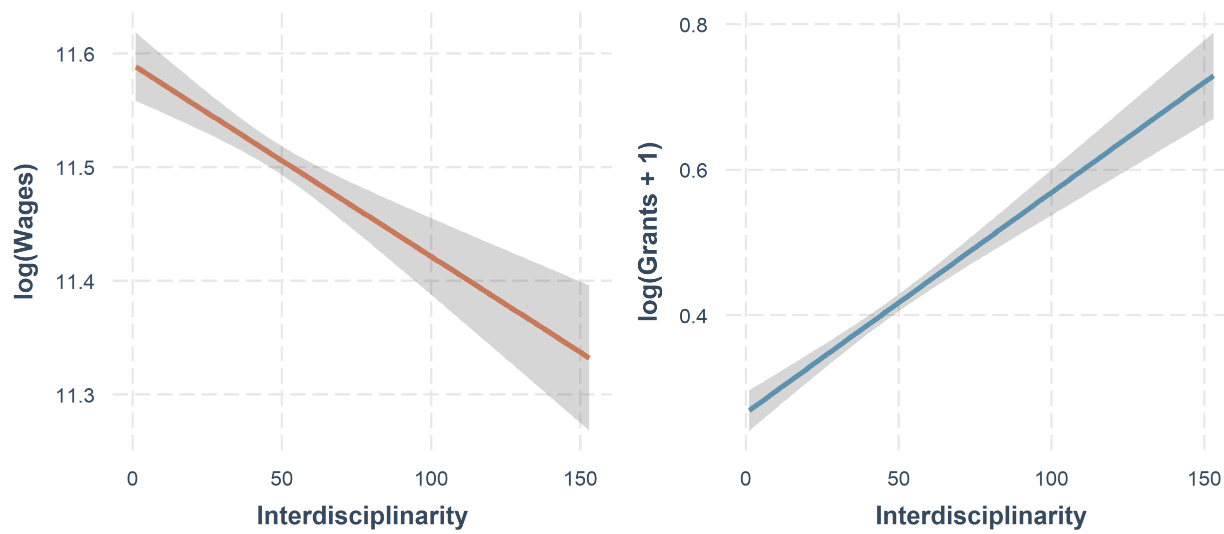

Figure 1: Correlation between the interdisciplinarity profile of a sample of researchers at UF and their academic achivements: the yearly wage (left-side panel) and the number of grants obtained in a year (right-side panel). Full results in A6.

academic circle resulting in lower wages.

To confirm and better understand the mechanisms that lead to these contrasting outcomes, in the following sections, we will investigate the reasons behind the observed difference by looking at the main drivers of reputation building and societal impact. Namely, keeping all other variables constant, we will focus on the number of citations accrued by a scholar as one important evaluation criterion in career progressions and therefore in the wage level. We then look at the diffusion of the knowledge across disciplines as the fulfillment of the interdisciplinarity required by funding agencies. It is worth anticipating that, while retaining the scholar's perspective, we will perform our analysis starting from papers. This allows us to characterize scholars' research at a more fine-grained level than what is allowed by variables that concern scholars. Moreover, by considering articles and not aggregating data at the researcher level, we are able to distinguish among the different dimensions of IDR.

3 IDR and researchers' trade-offs: research hypotheses

To capture the different facets of IDR that might influence research' scientific outcomes, we measure interdisciplinarity as the Diversity of the combined knowledge, i.e. "the apportioning of elements or options in any system" (Stirling, 2007). In fact, several mechanisms exist through which IDR might affect scientific impact, and the existence and extent of the supposed trade-off between private and public benefits might also vary considerably across the dimensions of IDR. We rely on the literature that decomposes Diversity in three independent components (Fontana et al., 2020; Hackett et al., 2021; Porter and Rafols, 2009; Stirling, 2007; Yegros-Yegros et al., 2015), defined at the article level: Variety, Balance, and Disparity.[6] The three dimensions of Diversity have specific meanings and autonomy, and refer respectively to the number of different disciplines involved in the making of the paper, their relative frequency, and their distance.

Variety is the basic form of interdisciplinarity: it returns the number of different disciplines that are referenced in the paper. It provides *prima facie* evidence on the intensity of interdisciplinarity of an article, but gives no information on the relative importance of the involved disciplines.

Balance overcomes this drawback by building on Variety in order to quantify the distribution of disciplines in the article references. Namely, it measures the evenness of the distribution of disciplines in references. Low values of Balance indicate that the paper references articles from a prevailing discipline, while high values of Balance correspond to an even distribution of disciplines in references.

Disparity measures a further dimension of Diversity: the proximity of the referenced disciplines in the knowledge space. The underlying idea is that disciplines that frequently co-occur in references are closer than those that co-occur rarely with respect to all other occurrences. High values of Disparity signal that a paper references fields that are very distant – have a low proximity – in the knowledge space. This indicator is rather different from Variety and Balance in that it does not heavily depend on the system of data classification as they do: proximity is calculated over the entire sample of articles and, therefore, provides the effective relative distance between pairs of disciplines. We will provide further details on the operalization of these indicators in Section 4.

The channels through which the IDR dimensions can affect the reputation and the societal contribution of a scholar are diverse. Firstly, there might exist a trade-off between the different dimensions of IDR. Increasing Variety implies that the pool of possible citing scholars increases. As a result, this component of IDR might positively impact both the number of citations and the diffusion of knowledge across fields. However, this might not hold when the referenced disciplines are

[6] Diversity also includes a compound indicator, the Rao-Stirling diversity, that is more suitably computed when the distinct role of the IDR components is not relevant to the object of analysis.

very distant to one another or when the focal paper is hardly identifiable with a field of study. This results in a trade-off for the researcher that pursues IDR, since increasing Variety will eventually end up in increasing Disparity.

Moreover, while the increase in some components of IDR is likely to positively affect the circulation of knowledge (public benefit), it might penalize the scholar prestige in a highly specialized academic environment (private benefit). This aspect might be particularly relevant for Balance: an even distribution of references to different disciplines may encourage the diffusion of the paper across a wide range of fields, but, at the same time, the paper will not have a target scientific community and will hardly be highly cited.

Combining these insights, we developed the first two hypotheses that we will test in our empirical analysis:

Hypothesis 1a (HP1a): *If IDR has an effect on the scholars' reputation, this impact differs across the various dimensions of IDR: while high Variety increases the potential to be cited by a larger set of scholars, a growth in Balance and Disparity might reduce the number of citations received by an article, since it will hardly fit within a defined field of study. Therefore, a trade-off in scholars' private benefits exists.*

Hypothesis 1b (HP1b): *If IDR has an effect on the circulation of knowledge, this impact differs across the various dimensions of IDR: while high Variety and Balance increases the potential diffusion of knowledge across fields, a growth in Disparity might reduce the circulation of an article across disciplines, since it will be more difficult to integrate in the existing literature. Therefore, a trade-off in public benefits exists.*

The different impacts of IDR components on scientific outcomes also result in a trade-off between private and public benefits. In this respect, we will test the following hypothesis:

Hypothesis 2 (HP2): *If IDR has an impact on the scholars' reputation and circulation of knowledge, the effect differs across these two indicators of scientific outcome: the increase in Balance in IDR hampers receiving a high number of citations, but favors knowledge diffusion across disciplines. Therefore, a trade-off between public and private benefits in pursuing IDR exists.*

4 Empirical strategy

The aim of our empirical analysis is to compare articles with different interdisciplinary content and assess whether they have a different scientific impact.

In order to make sure that articles are fairly compared, we elaborate an empirical design which allows us to compare only articles with similar characteristics, but with a different interdisciplinary content, published by the same author within the same field of study during the same year. Of course, interdisciplinarity is only one of the many factors determining the scientific impact of an article. If these factors are not considered, we would have a problem of omitted variables biasing our analysis. For this reason, we make sure that comparison is conducted sorting out specific features of the article and time-varying characteristics of the author which may concur to explain the scientific impact of an article.

In practice, our analysis is conducted using the following model:[7]

$$Y_{ijft} = IDR_{ijft}\beta + X_{it}\gamma + K_{jf}\delta + \alpha_i + \phi_f + \theta_t + \epsilon_{ijft}. \tag{1}$$

Here, the dependent variable (Y_{ij}) is a measure of the scientific impact of a paper j written by an investigator i at time t in the field of study f, measured alternatively as the number of citations received by j or its generality index (see Section 4.1 for their definitions), and the regressor of interest is IDR_{ijft}, which measures the various interdisciplinarity dimensions of paper j as defined in Section 4.1 (i.e. Variety, Balance, and Disparity).

The variables ϕ_f, θ_t, and α_i denote fields of study, year, and investigator fixed effects, respectively. These allow to compare only articles with similar characteristics, considering different sources of unobserved heterogeneity which may interfere with the effect that interdisciplinarity has on the scientific impact of an article: i.e., time-invariant characteristics of the article's field of study, publication year, and author. The variables K_{jf} and X_{it} are a proxy of the characteristics of the article and the author, respectively. They sort out potential problems of omitted variables in the model specification by controlling for specific features of the article (i.e., the number of authors,

[7] Estimates are obtained using an ordinary least squares regression. For the model specification where the dependent variable is the number of citations, we test the robustness of our results to the choice of the estimator. Specifically, we estimate our model using both Poisson and Negative Binomial regressions. Results are qualitatively unchanged. They are presented in Table B1 in Appendix B.

the presence of collaborators affiliated to an institution outside the United States, the adoption of a monodisciplinary approach)[8] and time-varying characteristics of the author (i.e., the H-index of the investigator i at time t, that is an author-level metric that measures cumulative productivity and citation impact of the researcher) which may concur to explain the scientific impact of an article. In order to avoid over-weighting extreme values in our estimates, and correctly deal with the highly skewed nature of our continuous variables, these are all log-transformed. The descriptive statistics for these variables in our data are presented in Table 1.

In the model, the parameter of interest is β, i.e., the estimated coefficient associated to IDR_{ijft}. This has to be interpreted as the average effect of an increase in the interdisciplinary content of an article on its scientific impact, *all else being equal*: i.e., when comparing articles with similar characteristics, but with a different interdisciplinary content, published by the same author within the same field of study during the same year. The robustness of our estimates relies on the fact that we are able to sort out from the model any identification threat arising from the presence of omitted variables (i.e., specific features of the article, K_{jf}, and time-varying characteristics of the author, X_{it}), and from performing unfair comparisons between articles due to potential unobserved heterogeneity in our data (i.e. time-invariant characteristics of the author, α_i, the field of the article, ϕ_f, and the year in which the article was published, θ_t). Importantly, the estimated value of β can be considered as representative of a large population, since our data covers a large number of authors working in many different fields across several years.

It is important to stress that we can rely on this sound empirical design because of our rich and innovative source of data which keeps track of the career of researchers working in several disciplines over different years, and allow us to use individual, field of study, and time fixed effects. To the best of our knowledge, we are the first to use this model specification in this strand of research.

4.1 Interdisciplinarity and scientific outcome indicators

As anticipated in Section 3, following Stirling (2007), we define three different dimensions of interdisciplinarity: Variety, Balance, and Disparity. We compute these indicators by using the disciplines of the papers listed in the references of the focal articles.

Variety measures the number of different disciplines referenced by the paper. Thus, we define

[8] Please observe that in these cases Balance and Disparity are not defined.

Variety (V_j) as:

$$V_j \equiv \sum_{s \in F} 1, \tag{2}$$

where F is the set of disciplines s in references of a paper j.

Balance, instead, refers to the evenness of the distribution of disciplines. We operationalize Balance (B_j) as a normalized Shannon Entropy, defined as:

$$B_j \equiv \frac{1}{\log V_j} \sum_{s \in F} f_s \log f_s, \tag{3}$$

where V_j is Variety measured as above and f_s is the frequency of discipline s in references of paper j. After normalization, this index assumes values between 0 and 1.

Finally, Disparity (D_j), which concerns the distance among referenced disciplines, is defined as the normalized sum of proximity among fields:

$$D_j \equiv \frac{1}{V_j(V_j - 1)} \sum_{\substack{r,s \in F \\ r \neq s}} (1 - p_{rs}), \tag{4}$$

where p_{rs} is the proximity between disciplines r and s. The computation of proximity is usually based on the co-occurrence of disciplines in articles, normalized by the size of fields. A common indicator is cosine similarity, which measures the cosine between fields' vectors of co-occurrences in references. Disparity is bounded between 0 and 1 and is independent of Variety and Balance. It is worth noting that Balance and Disparity are not defined for articles that cite only one discipline (i.e. when Variety is equal to one).

Figure 2 exemplifies the three measures of interdisciplinarity in the case of a paper that cites three unevenly-distributed disciplines, with different proximity to each other.

For what concerns the scientific impact, we operationalize researchers' reputation in academia as the total number of citations received by a paper in a five-year period after the publication date (Hamermesh and Pfann, 2012). It is described as:

$$C_j \equiv \sum_{t=y_{pub}}^{y_{pub}+5} c_{jt}, \tag{5}$$

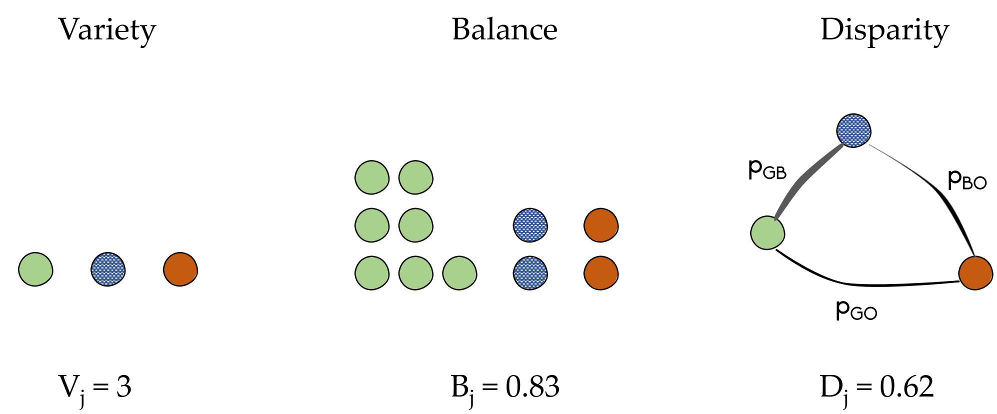

Figure 2: Example to illustrate the IDR measures. The example article cites three different disciplines (Green, Blue, Orange), with a prevalence of Green (7) over Blue (2) and Orange (2). In Disparity, the strength of links between fields of study is proportional to their mutual proximity. In this example, Green and Blue are similar to each other (they are often cited together, i.e. they frequently co-occur in references), while Orange is more distant.

where y_{pub} is the article's publication year and c_{jt} represents the citations received by a paper j in year t. We count citations over a five-year time window to have an indicator that is consistent between papers published in different years.

To measure knowledge diffusion across disciplines, instead, we rely on an index of generality of knowledge. A bit of knowledge that influences many, possibly distant, disciplines can be thought of as more impactful than one that is received only by few disciplines (Carley and Porter, 2012). This index captures the degree of applicability and influence on different fields of study of the knowledge contained in a paper. It is computed using the Hirschman-Herfindahl concentration index of citations across disciplines (Hall et al., 2001; Trajtenberg et al., 1997) and is defined as:

$$G_j \equiv 1 - \sum_{f=1}^{|F|} \frac{N_{jf}^2}{N_j^2}, \tag{6}$$

where N_{jf} is the number of forward citations received by a paper j from papers in the field of study f, while N_j is instead the total number of forward citations received by the paper. By definition, Generality is bounded between 0 and 1. Articles having their citations spread among many disciplines will have a high value of this indicator.[9]

5 Data

We construct a novel and unique dataset that includes detailed information about researchers and their publications: we study all the researchers affiliated to the University of Florida in the period 2008-2013. UF is the flagship university in the state of Florida, it is a large research university comprising more than 50,000 students and 5,000 full-time faculty. Over the past ten years, research awards to the university have increased by 45%: from $619 million in 2011 to 900.7 million in 2020.[10] UF is a member of the Association of American Universities, an organization of sixty-two academically prominent public and private research universities in the United States and Canada, and it consistently ranks among the top ten public universities in the United States. UF therefore represents an excellent example of a prominent and large research-oriented institution, and for this reason it has been already used as a case study to investigate how scientific collaborations are formed (Leone Sciabolazza et al., 2017), the mechanisms of scientific team assembly (Smith et al., 2021), and the design of new research policies (Leone Sciabolazza et al., 2020).

From the UF's registry office, we obtained information on researchers' gender, department affiliation, and publication record.[11] The individual-level information is anonymous, thus researchers' names are substituted by a unique identifier. The investigators' publication records provided by the UF's registry office include articles' title, journal in which the article was published, and the publication year. We exploit the publication title to retrieve the Digital Object Identifier (DOI) assigned to each article, i.e. the unique identifier of the publication in all bibliometric databases.[12] Through DOIs, we then collect articles' citations and references from the Lens database, while papers' fields of studies and authors' institutional affiliations were collected from the Microsoft Aca-

[9] One shortcoming of this measure is that it is not defined in articles that did not receive any citations in the five-year windows. This may lead to selection bias concerns that are discussed in the following sections.

[10] From: *University of Florida hits record $900 million in research awards*, University of Florida News (2020). Available at: https://news.ufl.edu/2020/08/record-research-awards/.

[11] We focus on articles published in peer-reviewed journals, excluding books and other types of academic production from our analysis.

[12] This process exploits Crossref and Scopus APIs. The search procedure is described in Appendix A.1.

demic Graph (MAG) database.[13] We use information about citations received by papers to compute both scientific impact indicators and researchers' H-index, which will be our proxy for the quality of scholars.[14]

To determine disciplines associated to articles and compute interdisciplinarity indicators, we rely on the classification scheme implemented by MAG to retrieve the field of studies associated to each paper. This scheme is a hierarchical classification that identifies 19 disciplines (first level) and 292 sub-disciplines (second level). The taxonomy uses state-of-the-art artificial intelligence methodologies to extract semantic content from documents, exploring natural language processing techniques and networks semantic reasoning to delineate disciplines (Sinha et al., 2015; Wang et al., 2019). There are several advantages in using this classification: it is based on concepts and language used at the paper-level, thus it avoids any bias that may arise from arbitrariness in the details of classifications that rely on human experts (Wang and Schneider, 2020);[15] it uses a heterogeneous network semantics analysis that exploits the context in which the publication's text is embedded, linking it to authors, affiliations, and locations (Wang et al., 2019); and it also mitigates the assignment errors that results from the loss of granularity when we adopt journal-based categorizations. Moreover, journal-based taxonomies have difficulties in dealing with generalists journals like *Nature*, *Science*, and *PLoS ONE*.

In our final database, we observe 6,105 researchers at UF, of which 34% are women, with at least one article in a peer-reviewed journal in the period 2008-2013. On average, the period of activity of each scholar in our sample (i.e. the number of years in which she publishes at least one journal article) is three years. At UF, researchers belong to different colleges, which, in turns, are aggregated in four academic units: Liberal Arts and Sciences, Engineering, Health Sciences, and Food and Agricultural Sciences. Scholars in Health Sciences, especially in the college of Medicine, prevail in our sample (more details in Table A1). In addition to these pieces of information, the

[13]The Lens database and Microsoft Academic Graph database used to complement information on articles by UF's researchers are becoming widespread for bibliometric analysis in recent years. Given that the fields of study information is crucial for our IDR measures, we decide to rely on these sources to maintain consistency and uniformity between our databases. Both sources can be freely accessed for research purposes and available at the following links: Microsoft Academic Graph and Lens.

[14]The H-index is an author-level metric that measures cumulative productivity and citation impact of each researcher. It takes into account the scholar's best cited papers and their number of citations. A researcher with n papers with at least n citations will have a H-index of n.

[15]For example, the total number of categories of the two most frequently used systems of classification, Web of Science (WoS) journal subject categories (SC) and the All Science Journal Classification (ASJC) from Scopus, varies drastically: there are 252 SCs and 330 ASJCs.

UF's registry office also reports yearly wages and the number of awarded grants for a limited number of researchers. These data have been exploited in the motivating evidence (see Section 2) and are described in Table A5.

The full publication record of UF's researchers consists of 23,926 articles published in peer-reviewed journals. As reported in Figure A1, the number of publications by year is quite stable over time, with about 4,000 articles per year. Overall, these papers made 646,280 references and received 366,024 citations in five years from the publication date. Considering only the years of activity, each UF's researcher published an average of 2.22 papers per year. 23% of these papers involves international collaborations, and 46% of them has more than one UF's researcher as an author. More details about researchers' and articles' characteristics are in Table 1.

Each article in the sample belongs to one or more disciplines, as measured by the MAG field of study classification. We exploit the most fine-grained level of this hierarchical classification (second level) to define articles' degree of interdisciplinarity and generality (see Section 4.1 for the definition of indicators). This level consists of 292 categories, that can be aggregated in 19 more general fields of study (first-level classification). While the second level of classification is used to compute all article-level indicators, we consider the first level of classification to define discipline fixed effects and, thus, control for different citation patterns across disciplines. In this section, we refer to these 19 fields of study at the first level of classification also for descriptive purposes, in order to describe articles' characteristics. The distribution of papers over these 19 categories is reported in Table 2. As expected, the average number of references and the average number of citations is heterogeneous across fields of study. More details about the number of references and citations by discipline are available in Table A2.

The information about MAG fields of study is also used to compute the knowledge space in which the scholars perform their research. The knowledge space, which summarizes the proximity between disciplines, is the core of the Diversity indicator, one of the dimension of IDR considered in this paper. As we are interested in fine-grained definitions of IDR indicators, we consider the knowledge space at the second level of discipline classification (292 fields of study). To avoid biases due to the small number of papers in our sample and obtain a more reliable measure of similarity between disciplines (as a proxy of the easiness in combining different topics and techniques in a single research), we use an index of proximity among fields of study computed over the universe

Table 1: Summary statistics.

Variables	*Mean*	*SD*	*Min*	*50%*	*Max*	*Obs*
Panel A: Researcher-level Data						
Nb. Papers/Year	2.22	2.08	1	1.6	40.33	6,105
Nb. Citations/Year	17.65	42.37	0	5.0	955.50	6,105
H-index	2.37	2.88	0	1.5	35.83	6,105
Gender (Woman=1)	0.34	0.48	0	0	1.00	6,105
Panel B: Paper-level Data						
Nb. Citations	20.30	46.34	0	10	2,530	23,926
Generality	0.72	0.18	0	0.77	0.98	22,658
Variety	37.06	19.54	1	36	153	23,926
Balance	0.84	0.09	0	0.85	1	23,926
Disparity	0.68	0.07	0	0.70	0.94	23,926
Nb. References	40.21	33.01	1	34.00	926	23,926
Nb. of Authors	5.64	9.90	1	4	1,269	23,926
International Collab.	0.23	0.42	0	0	1	23,926

Notes: Panel A shows selected measures of productivity of 6,105 researchers affiliated to the University of Florida from year 2008 to 2013. Gender is a dummy variable that assumes the value 1 when the researcher is a woman. Panel B shows descriptive statistics of the 23,926 articles published by these researchers in the time window 2008-2013. Nb. Citations is the total number of citations received in a 5 years period after the publication. Generality captures the degree of applicability of the knowledge codified in a paper on different fields of study. It is worth noting that generality is not defined for papers with zero citations. International collaboration is a dummy variable that assumes the value 1 when at least one co-author in the paper is affiliated to an institution outside the United States.

of articles in MAG. This proximity measure is based on the Network Similarity Package, a series of processing functionalities for MAG that allow us to compare two fields of study and obtain a similarity score that represents how close these fields are, based on the frequency they appear together in a same paper.[16] Based on this measure of similarity, we represent the network of fields of studies, i.e. the knowledge space, in Figure 3. The graph connects disciplines whose co-occurrence is frequent in the universe of MAG articles. Nodes represent fields of study at the second level of MAG classification, but, to ease the interpretation of the knowledge space, their shapes and colors correspond to disciplines at the upper level of classification (conversion table is available in Appendix

[16]For details on the Network Similarity package, see Microsoft Research (2020).

Table 2: Distribution of focal papers by field of study (first level of classification).

Field of Study	**Total**	**Average Nb. References**	**Average Nb. Citations**
Biology	7781	46.07	22.26
Medicine	6305	35.64	22.14
Chemistry	2628	41.60	19.86
Psychology	1703	46.51	17.31
Physics	1686	39.27	22.71
Mathematics	996	26.48	9.87
Materials science	785	34.34	21.68
Computer science	508	31.73	12.65
Geology	506	48.77	17.96
Economics	503	37.41	13.82
Engineering	404	27.78	13.71
Sociology	199	34.58	8.27
Environmental science	85	39.41	52.58
Geography	59	44.47	29.29
History	39	29.44	3.36
Political science	29	26.90	11.10
Business	26	57.19	24.00
Philosophy	20	31.75	2.15
Art	7	11.29	1.57

Notes: This table shows the distribution of focal articles per fields of study at the first level (19 categories). The average number of references relates to papers cited by our articles of interest and the average number of citations takes into account total number of citations within 5 years from the publication.

C). In the graph, sub-disciplines belonging to environmental science, medicine, and biology are on the left. At their right, we can observe the interconnection between economics and business. The bottom part of the network, instead, shows the connection between fields in mathematics (starting from the left), engineering, computer science, chemistry, physics, and material science. At the top of the figure, the interpenetration between art (included literature), psychology, sociology, history, and geography is evident.

Beyond the information about the relative distance between disciplines, the field of study classification and the knowledge space allow us to define the three different dimensions of IDR in our sample, as explained in Section 4.1. Figure 4 shows average values of Variety, Balance, and Disparity by field of study (at the first level of classification). While the average values of these indicators do not differ considerably across disciplines, some fields of study have unique characteristics in terms

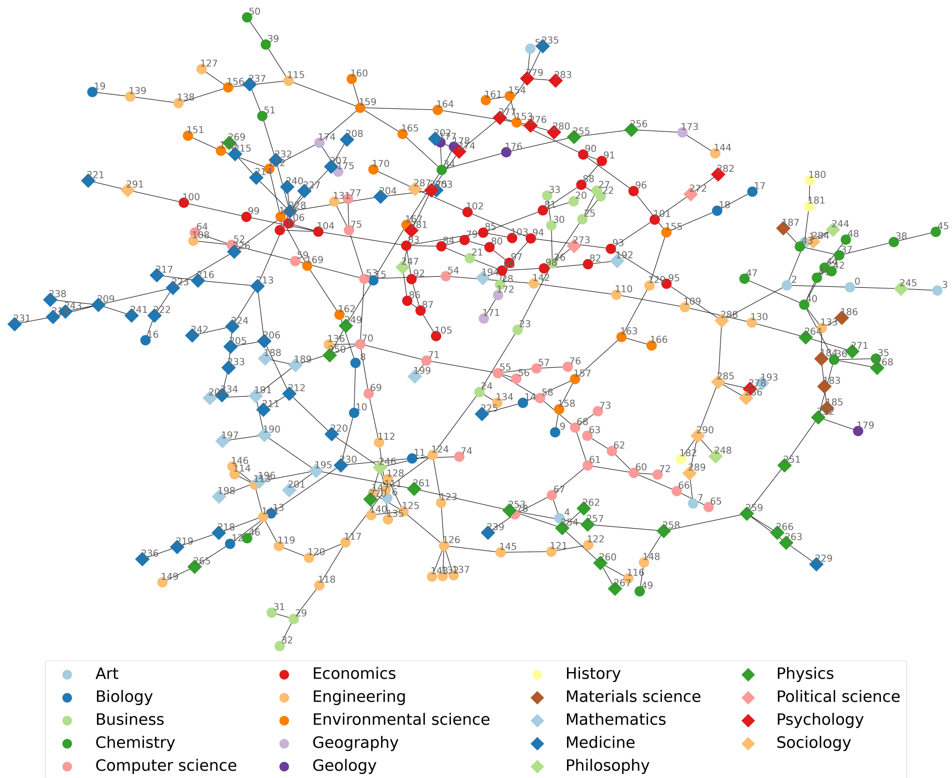

Figure 3: Knowledge space among fields of studies. The network shows the proximity between fields of study at the second level of the MAG classification (292 fields of studies). To ease the graph' interpretation, authors grouped fields of studies by discipline (the first level of MAG classification), which are represented by different colors and shapes, as reported in the plot legend. The conversion between the two levels as well as the field of studies corresponding to node IDs are reported in Table C1.

of interdisciplinarity. The most evident one is art, as it has the lowest average Variety and Disparity and the highest Balance in the sample. Those values characterize art as a poorly diversified discipline, in which, however, different fields are evenly combined in article references. The opposite occurs in business. In this field of study, the articles show, on average, a high Variety and Disparity – meaning that they are highly diversified – but a relatively low Balance – signaling the presence of a core field in article references. The importance of a core field of study (low Balance) is especially relevant in philosophy, biology, and physics. History, instead, results as a highly specialized field since it has a relatively low value in all three indicators.

In the following section, we explore in more details the different dimensions of interdisciplinarity by using the publication records of UF's researchers described in this section. The design of

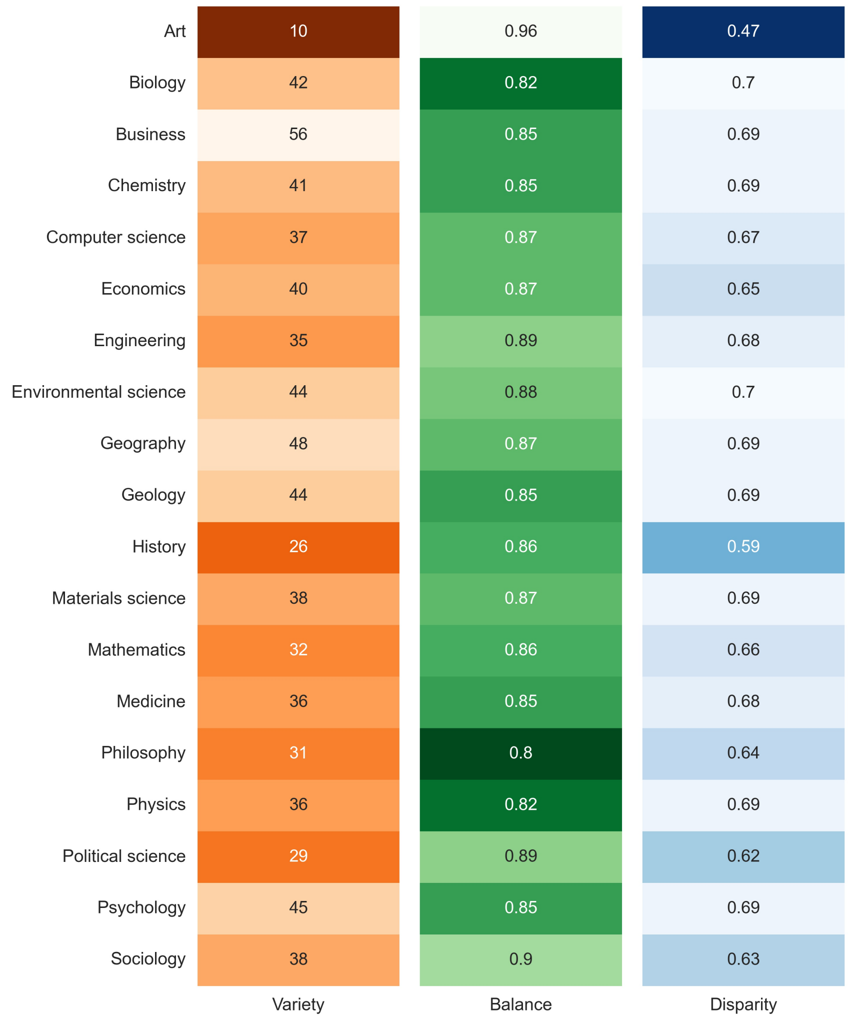

Figure 4: Average Variety, Balance, and Disparity per field of study (first level of classification).

our empirical strategy requires both article-level and researcher-level information. By matching article-level and individual-level information for each UF's author of our papers, we obtain 46,156 observations at the paper-researcher level as the co-authorship between UF's researchers is frequent in our sample. Descriptive statistics at the paper-researcher level are available in Table A3.

6 Results

6.1 Main results

In this section, we present the results from the estimation of equation (1) to assess the average effect of an increase in the interdisciplinary content of an article on its scientific impact, *all else* *being equal*: i.e., when comparing the scientific impact of articles with similar characteristics, but with a different interdisciplinary content, published by the same author within the same field of study during the same year. Findings from this exercise will be used to investigate the evidence in favor of hypotheses HP1a, HP1b, and HP2.

HP1a and HP1b posit that the impact of IDR on a given measure of scientific outcome differs according to the interdisciplinarity dimension considered. In order to test these hypotheses, we will assess whether the different dimensions of interdisciplinarity have a similar effect in determining the scientific impact of a paper (i.e. either citations or generality), or some of them are considered desirable and are rewarded by the academia while others are less desirable and thus penalized. Specifically, H1a states that, while Variety has a positive influence on the number of citations, the opposite occurs with Balance and Disparity. If the latter is verified, we would find evidence of the existence of a trade-off in researchers' public benefits. H1b, instead, conjectures a positive effect of Variety and Balance on the diffusion of knowledge and a negative impact of Disparity on the same scientific outcome. In this case, we expect to find evidence on the presence of a trade-off in the public interest and societal benefit.

HP2 states that the impact of a given dimension of IDR differs according to the measure of scientific impact considered: while we expect to observe a negative effect of Balance on the number of citations, a positive impact of the same IDR dimension is supposed to have a positive influence on the generality of knowledge. In order to test this hypothesis, we will investigate whether the same dimension of interdisciplinarity (Balance) has the same impact when considering different measures of scientific impact, i.e. citations and generality, or this is rewarded in some cases and penalized in other cases. If evidence supports the latter scenario, then results would confirm our hypothesis and the existence of a trade-off between private and public benefits in pursing IDR.

We begin our investigation by testing HP1a. To this purpose, we assess the effect of an increase in the interdisciplinary content of an article on its scientific impact when this is measured in terms of number of citations (reputation).[17] Results are reported in Table 3. In column (1), we jointly estimate the effects of the three dimensions of interdisciplinarity, so to assess the effect of an increase in the interdisciplinary content of a paper in one dimension (e.g., Variety), while accounting for

[17]It is worth noting that we compute the interdisciplinary indicators by considering the second level of discipline classification (292 fields of study), while we use the first level of classification (19 fields of study) to define discipline fixed effects that controls for the presence of different citation patterns across disciplines.

changes in other interdisciplinary dimensions (e.g. Balance and Disparity). In order to consider different potential sources of unobserved heterogeneity which may interfere with the effect that interdisciplinarity has on the scientific impact of an article, we control for monodisciplinarity and include individual, field of study, and year fixed effects into our model specification. We find that only an increase in Variety of a paper has a positive and statistically significant effect on its number of citations, whereas the other two indicators (Balance and Disparity) have a negative and statistically significant effect. Our results are thus in favor of HP1a, and we observe a trade-off in scholars' private interest.

In column (2), we augment our model specification by including a control for the number of authors in the paper. The estimated coefficient of this variable indicates that increasing the number of authors has a positive and statistically significant impact on the number of citations of a paper. This is consistent with the idea that the narrower expertise of researchers requires having larger teams to producing widely-cited research. Most importantly, all our previous findings in favor of HP1a are confirmed: i.e., the sign and the statistical significance of the three dimensions of interdisciplinarity are unchanged with respect to column (1).

In column (3), we add to our model specification a dummy variable registering whether one of the co-authors of the paper is affiliated to an institution outside the US. We find that having an international collaborator in the team has a positive and statistically significant effect on the number of citations, hinting that working in an international team may expand the visibility of one's work. Again, all our results supporting HP1a are left qualitatively unchanged.

In column (4), we add a control for the H-index of the investigator, thus estimating equation (1) with its entire set of controls. We find that having a higher H-index has a positive and statistically significant effect on the number of citations. Most relevant to us, the evidence in favor of HP1a is still confirmed.[18] Even after including the entire set of controls in our model specification, interdisciplinarity has a large and significant impact on citations, and the direction of this effect depends on the dimension of interdisciplinarity considered. Specifically, we find that a 10% increase in Variety increases by 5.38% the number of citations received in a 5 years time period by a researcher with an article. This result is in line with Leahey et al. (2017), who finds the same positive effects on the total number of citations. At the same time, we find the opposite effect for the other measures

[18] The results are robust to controlling also for disciplines' average H-index. Results are available upon request.

of interdisciplinarity. A 10% increase in Balance decreases by 35.20% the citations accumulated with a paper within 5 years. This supports the idea that an even distribution of the references among fields of study negatively impacts citations, i.e. articles built on a core field of study are more easily recognized as relevant by specialized readers. Finally, a 10% increase in Disparity diminishes by 11.38% the number of citations received with an article within 5 years, suggesting that academic audiences might find it difficult to receive articles that integrate more distant knowledge, in accordance with Yegros-Yegros et al. (2015). Overall, these results suggest that IDR has a positive effect on citations if papers integrate knowledge from various, but not too distant, fields of studies while referring mainly to a specific discipline (and audience).

Table 3: The effects of interdisciplinarity on citations.

	Dependent variable:			
	log(Citations + 1)			
	(1)	(2)	(3)	(4)
log(Variety)	0.647^{***}	0.551^{***}	0.552^{***}	0.550^{***}
	(0.016)	(0.015)	(0.015)	(0.015)
log(Balance + 1)	-4.454^{***}	-4.580^{***}	-4.548^{***}	-4.554^{***}
	(0.201)	(0.191)	(0.191)	(0.191)
log(Disparity + 1)	-0.992^{***}	-1.295^{***}	-1.282^{***}	-1.268^{***}
	(0.246)	(0.229)	(0.229)	(0.229)
log(Number of Authors)		0.453^{***}	0.447^{***}	0.445^{***}
		(0.013)	(0.013)	(0.013)
International Collaboration			0.037^{*}	0.037^{*}
			(0.015)	(0.015)
log(H-index + 1)				0.127^{***}
				(0.020)
Variety = 1	YES	YES	YES	YES
Fields of Study Fixed Effects	YES	YES	YES	YES
Year Fixed Effects	YES	YES	YES	YES
Researcher Fixed Effects	YES	YES	YES	YES
Number of Researchers	6,105	6,105	6,105	6,105
Observations	46,156	46,156	46,156	46,156
R^2	0.442	0.479	0.479	0.480
Adjusted R^2	0.356	0.399	0.399	0.400

Notes: This table presents OLS estimates of the effects of interdisciplinarity on citations, following equation 1. Observations are at the paper-researcher level. The dependent variable is the logarithm of total citations accrued in five years. All regressions include individual, year and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

We continue our investigation by moving to HP1b and considering the effect of an increase in the interdisciplinary content of an article on its scientific impact when this is measured in terms of knowledge diffusion across disciplines. Estimates are conducted with the same model specifications adopted in the previous exercise, and the results are presented in Table 4.[19]

We find evidence in favor of HP1b across all model specifications: the impact of IDR on the generality of an article differs according to the interdisciplinarity dimension considered. Specifically, we observe that an increase of Variety and Balance has a positive and statistically significant impact on the diffusion of knowledge, while the effect of Disparity is negative and statistically significant. In other words, while some dimensions of interdisciplinarity are essential for spreading ideas and concepts across multiple fields (Variety and Balance), the combinations of very distant knowledge is not well received by the scientific community.[20] All in all, we find evidence that a trade-off in public benefit exists: the diffusion of knowledge beyond disciplinary fields is favored by the increase in the degree of interdisciplinarity in scientific research, but this advantage is limited when the recombined fields of study are distant from each other.

Notably, estimates from column (4) show that even after controlling for our entire set of controls, the effect of interdisciplinarity on the diffusion of knowledge across disciplines is statistically significant, regardless of the dimension considered.[21] Moreover, the magnitude of this impact is considerable for almost all the dimensions observed. In fact, our results show that only Variety has a modest effect on generality, with a 10% increase in the number of unique fields of study in the paper's references leading to an increase of the generality index by 0.36 percentage points. On the contrary, Balance and Disparity have a sizeable positive effect on the diffusion of knowledge. In particular, a 10% increase in Balance raise the generality index by 2.29 percentage points. On the contrary, a 10% increase in Disparity decreases the paper generality by 1.90 percentage points.[22]

[19] We report in this table the second stage of a two-step Heckman correction model to control for potential selections in our sample (i.e. the fact that some papers have zero citations). This exercise does not rely on the use of a specific exclusion restriction, and it only makes use of the variables included in the second stage of the model (i.e. our covariates). It is worth noting that, even when an exclusion restriction is not used, identification is formally achieved, though results may be less precise in terms of statistical significance. This should be not of any practical concern, however. Our aim is to test whether our results remain qualitatively unchanged even when controlling for the potential presence of selection issues. Reassuringly, the evidence produced by our exercise confirms all our model predictions. Results of the first stage are available in Table B2 of the Appendix B.

[20] With respect to our previous exercise, we observe that a larger number of co-authors has a positive and statistically significant effect on the generality of the paper, while the presence of international collaborators in the team and the H-index of the researcher have no statistically significant effect.

[21] These results are robust to controlling also for disciplines' average H-index. Results are available upon request.

[22] As a robustness check, we have estimated the model in column (4) replacing our indicator of knowledge diffusion,

Table 4: The effects of intedisciplinarity on the diffusion of knowledge across fields.

	Dependent variable: log(Generality + 1)			
	(1)	(2)	(3)	(4)
log(Variety)	0.034^{***}	0.038^{***}	0.038^{***}	0.038^{***}
	(0.002)	(0.002)	(0.002)	(0.002)
log(Balance + 1)	0.287^{***}	0.237^{***}	0.238^{***}	0.238^{***}
	(0.024)	(0.025)	(0.025)	(0.025)
log(Disparity + 1)	-0.195^{***}	-0.201^{***}	-0.201^{***}	-0.201^{***}
	(0.038)	(0.037)	(0.037)	(0.037)
log(Number of Authors)		0.011^{***}	0.011^{***}	0.011^{***}
		(0.001)	(0.001)	(0.001)
International Collaboration			0.001	0.001
			(0.001)	(0.001)
log(H-index + 1)				-0.001
				(0.002)
IMR	-0.117^{***}	-0.069^{***}	-0.069^{***}	-0.069^{***}
	(0.013)	(0.015)	(0.015)	(0.015)
Variety = 1	YES	YES	YES	YES
Fields of Study Fixed Effects	YES	YES	YES	YES
Year Fixed Effects	YES	YES	YES	YES
Researcher Fixed Effects	YES	YES	YES	YES
Number of Researchers	5,938	5,938	5,938	5,938
Observations	44,084	44,084	44,084	44,084
R^2	0.341	0.342	0.342	0.342
Adjusted R^2	0.238	0.239	0.239	0.239

Notes: This table presents second stage results from Heckman's two-steps estimation of the effects of interdisciplinarity on the diffusion of knowledge across fields, following equation 1. Observations are at the paper-researcher level. The dependent variable is the logarithm of the generality index, defined in equation 5. All regressions include the Inverse Mills Ratio (IMR) to control for sample selection bias and individual, year and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

We now investigate HP2 by confronting the results from Table 3 and Table 4: i.e. comparing the direction of the effect of a given dimension of interdisciplinarity across different measures of scientific impact (number of citations vs knowledge diffusion beyond disciplinary boundaries).

We find evidence that the effect of Balance differs with the measures of scientific impact considered. In fact, papers with lower Balance have more citations but reach a less diverse audience of academics. This finding is consistent with HP2: i.e., it exists a trade-off between the number of citations that one can accrue, and the generality that one can achieve. The effect on generality corroborates what observed for the effect of Balance on the number of citations: papers that refer more evenly to the discipline pool have no target field and, thus, for them, it is more difficult to accrue citations from a specialized literature. At the same time, these articles have a broader appeal because they bridge audiences that were previously separated, boosting the societal impact of the work.

Two relevant implications follow from our results. First, since interdisciplinarity has a statistically significant effect on citations and generality of knowledge, but the direction of the effect depends on the IDR dimension considered, then researchers face a dilemma in how to approach IDR. In fact, despite the three dimensions are distinct, they are not completely independent. For instance, by increasing Variety (which has a positive effect on one's research impact), one will eventually increase Disparity (which has a negative effect instead).[23]

Secondly, Balance has strong but opposite effects on citations and generality. This indicates that researchers face a trade-off between increasing their reputation and reaching out to other disciplines. The costs of IDR in terms of citations are important enough to negatively impact researchers academic careers, but the public benefits regarding the diffusion of knowledge are substantive and cannot be dismissed. This disconnection between private and public returns, i.e. the interdisciplinarity dilemma, sets a challenge to the design of research policies.

i.e. the generality index, with the related and adapted Herfindahl index as developed by Gruber et al. (2013) to measure the breadth of technological recombinations in patents. We find that our results are qualitatively unchanged. Results are available upon request.

[23]We attempt to approach this question by including a polynomial term for Variety in our main specification and estimating their effects for both of our outcome variables. Our preliminary results show that the linear and the quadratic terms associated to Variety are positive and statistically significant for both citations and generality, meaning that we do not find evidence of an optimal level of Variety. Future research should be dedicated to understanding how to consider all dimensions in order to assess the optimal level of interdisciplinarity. Results are available upon request.

6.2 Heterogeneous effects

In this section, we explore whether the effects of IDR vary according to the characteristics of the investigators, and provide different incentives to engage in interdisciplinary work. To this purpose, we estimate equation (1) by considering only researchers with specific features, and test whether our hypotheses (HP1a, HP1b, and HP2) are confirmed regardless of the population of scientists considered. Because of our empirical design, estimates have to be interpreted as a measure of the additional effect of an increase in the interdisciplinary content of an article on its scientific impact, given the overall effect of producing an interdisciplinary content being an author with specific features: i.e., results indicate the marginal (rather than the total) effect of an increase in the interdisciplinary content of an article, given the characteristics of the author.

We begin by testing HP1a, and results are presented in Table 5. In column (1) and (2), we estimate our model by considering alternatively articles written by male and female scientists. Although the effects of the IDR dimensions on the number of citations seems more pronounced for women, they are qualitatively the same. In other words, the effect of an increase in the interdisciplinary content of an article on the number of citations accrued by a female author is similar to the effect estimated for a male author: HP1a is confirmed for both of them. Of course, the fact that we find no striking differences in the effect of interdisciplinarity when separately estimating our model for women or for men, does not indicate that men and women receive the same number of citations to their articles. The total number of citations obtained by the two categories of authors for an article may still be very different. Engaging in interdisciplinary work, however, seems not to play a significant role in explaining potential differences in the scientific impact between the two categories, because all authors are subject to the same dilemma regardless of their gender.

In column (3) and (4), we estimate our model by considering alternatively articles written with or without international collaborators. This is because international collaborations may influence the heterogeneity of the team and the knowledge-integration process, which, in turn, may affect the interdisciplinarity of the article. If this is the case, it could be that these two groups are not subject in the same way to the dilemmas associated to IDR. This is not what our evidence suggests, however. In fact, we do not find any qualitative difference on the effects that the different dimensions of interdisciplinarity have on the number of citations, when separately considering these two groups.

Table 5: Heterogeneity analysis of interdisciplinarity effects on citations.

	Dependent variable: log(Citations + 1)									
Samples	Men	Women	International Collaboration	Only US	Superstar	Non-Superstar	Interdisciplinarity Higher than Coauthors' Average	Interdisciplinarity Lower than Coauthors' Average	Coauthors with interdisciplinarity above college median	Coauthors with interdisciplinarity below college median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
log(Variety)	0.548***	0.563***	0.649***	0.532***	0.594***	0.544***	0.575***	0.493***	0.603***	0.508***
	(0.017)	(0.030)	(0.041)	(0.016)	(0.044)	(0.015)	(0.020)	(0.029)	(0.027)	(0.021)
log(Balance + 1)	−4.541***	−4.637***	−4.096***	−4.645***	−5.328***	−4.449***	−5.270***	−3.379***	−4.673***	−4.585***
	(0.217)	(0.402)	(0.399)	(0.217)	(0.633)	(0.197)	(0.262)	(0.330)	(0.324)	(0.273)
log(Disparity + 1)	−1.102***	−2.011***	−2.921***	−0.964***	−1.263†	−1.256***	−1.422***	−1.250**	−1.797**	−1.199***
	(0.252)	(0.561)	(0.593)	(0.254)	(0.760)	(0.237)	(0.358)	(0.431)	(0.583)	(0.297)
log(Number of Authors)	0.444***	0.449***	0.568***	0.413***	0.527***	0.430***	0.439***	0.486***	0.431***	0.472***
	(0.014)	(0.026)	(0.033)	(0.014)	(0.042)	(0.012)	(0.017)	(0.021)	(0.019)	(0.019)
International Collaboration	0.035*	0.039			0.049	0.036*	0.037†	0.028	0.063**	0.007
	(0.017)	(0.028)			(0.043)	(0.016)	(0.020)	(0.027)	(0.022)	(0.023)
log(H-index + 1)	0.119***	0.136***	0.184**	0.101***	−0.171	0.153***	0.093**	0.148***	0.107**	0.122***
	(0.024)	(0.034)	(0.059)	(0.021)	(0.161)	(0.020)	(0.035)	(0.035)	(0.034)	(0.032)
Variety = 1	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Fields of Study Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Researcher Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Number of Researchers	3,926	2,104	2,599	5,801	131	6,074	3,570	4,080	3,566	3,886
Observations	34,614	11,070	8,761	37,395	5,137	41,019	25,832	15,386	20,224	20,994
R^2	0.469	0.519	0.598	0.486	0.366	0.490	0.467	0.560	0.467	0.503
Adjusted R^2	0.400	0.404	0.425	0.391	0.346	0.401	0.380	0.400	0.352	0.389

Notes: This table presents OLS estimates of the effects of interdisciplinarity on citations, following equation 1, for subsamples divided by gender (columns 1-2), the presence of international collaborators (columns 3-4), productivity (columns 5-6), coauthors interdisciplinarity (columns 7-8), and the average coauthors interdisciplinarity measured at the college level (columns 9-10). Observations are at the paper-researcher level. The dependent variable is the total citations accrued in five years. All regressions include individual, year, and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: †p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

HP1a is confirmed for both of them.

In column (5) and (6), we estimate our model by considering alternatively articles written by star researchers (researchers in the upper 10th percentile of the H-index distribution within each year) and the rest of the sample. Once more, estimates are qualitatively similar between the two groups, and HP1a is confirmed for both of them. Prolific researchers who may engage in high-risk, high-reward publication strategies are exposed to the same effects of IDR than other researchers. Of course, this does not imply that their papers will accrue the same number of citations. This simply indicates that they face a similar dilemma, and differences across them are not to be attributed to a different effect that interdisciplinarity exerts on the scientific impact of their articles.

We continue our investigations by analyzing whether the interdisciplinary dilemma has an impact on investigators, depending on the characteristics of their co-authors in terms of IDR. In column (7) and (8), we estimate our model by considering alternatively articles written by researchers whose interdisciplinarity profile in that year was higher or lower than the average interdisciplinarity of their UF co-authors. We define interdisciplinarity profile of researchers as the maximum value of Variety registered for articles written by an investigator in one year. It may be the case that researchers face different constraints when they work with colleagues that produces papers with lower interdisciplinarity profile than when they go for a more interdisciplinary research team. However, again, estimates are qualitatively similar, which indicates that the effects of IDR on the number of citations are the same for both groups. Finally, in column (9) and (10), we estimate our model by considering alternatively articles written by researchers whose collaborators' average interdisciplinarity in a year was higher or lower than the average interdisciplinary of the college in which they are affiliated. Overall, we do not find evidence that the effects of IDR are driven by co-authors interdisciplinarity[24].

We further proceed by estimating equation (1) when considering a sample of researchers affiliated to a specific academic unit, in which the researcher's college and department are included. Results are reported in Table 6. In the estimations presented in this table, we alternatively consider researchers affiliated to: the College of Liberal Arts and Science (CLAS), column (1); the College

[24]As a robustness check, we have estimated the model in columns (7-9) in Table 5 adding to our specification a dummy variable registering if a paper was co-authored exclusively by UF investigators and interactions between this dummy and our interdisciplinarity measures. We find that our results are qualitatively unchanged. Results are available upon request.

Table 6: Interdisciplinarity effects on citations and college affiliation.

	Dependent variable: log(Citations + 1)			
	CLAS	ENG	HSC	IFAS
	(1)	(2)	(3)	(4)
log(Variety)	0.489***	0.493***	0.567***	0.555***
	(0.048)	(0.046)	(0.024)	(0.030)
log(Balance + 1)	−5.028***	−3.781***	−4.930***	−3.631***
	(0.572)	(0.770)	(0.283)	(0.373)
log(Disparity + 1)	−1.975**	−1.456*	−1.089**	−1.457*
	(0.636)	(0.563)	(0.337)	(0.658)
log(Number of Authors)	0.449***	0.317***	0.506***	0.343***
	(0.031)	(0.058)	(0.018)	(0.025)
International Collaboration	−0.048	0.023	0.070**	0.092**
	(0.040)	(0.041)	(0.026)	(0.030)
log(H-index + 1)	0.066	0.084	0.164***	0.040
	(0.069)	(0.073)	(0.030)	(0.035)
Variety = 1	YES	YES	YES	YES
Fields of Study Fixed Effects	YES	YES	YES	YES
Year Fixed Effects	YES	YES	YES	YES
Researcher Fixed Effects	YES	YES	YES	YES
Number of Researchers	665	389	2,215	1,198
Observations	5,267	4,946	18,260	9,825
R^2	0.503	0.360	0.474	0.478
Adjusted R^2	0.427	0.301	0.401	0.403

Notes: This table presents OLS estimates of the effects of interdisciplinarity on citations, following equation 1, for subsamples divided by academic unit affiliation. Column 1 estimates the effects for researchers affiliated to the College of Liberal Arts and Science, column 2 for those at the College of Engineering, column 3 for those at the Health Science Center and column 4 for those at the Institute of Food and Agricultural Sciences. Observations are at the paper-researcher level. The dependent variable is the total citations accrued in five years. All regressions include individual, year and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

of Engineering (ENG), column (2); the Health Science Center (HSC), column (3); and the Institute of Food and Agricultural Sciences (IFAS).[25] Results are qualitatively unchanged regardless of the affiliation considered, hinting that IDR has the same effect on the number of citations in all academic environments: i.e., researchers are subject to the same dilemma regardless of their affiliation, and HP1a is confirmed for all of them.[26]

[25]The colleges included in each academic unit are reported in Table A1 in the Appendix A.2.

[26]We also estimate equation (1) using alternative disciplinary subdivisions based on researchers' paper fields of study, individual main field of publication (measured as the field where the researcher published most of her papers),

Table 7: Heterogeneity analysis of interdisciplinarity effects on the diffusion of knowledge across fields.

	Dependent variable: log(Generality + 1)									
Samples	Men	Women	International Collaboration	Only US	Superstar	Non-Superstar	Interdisciplinarity Higher than Coauthors' Average	Interdisciplinarity Lower than Coauthors' Average	Coauthors with interdisciplinarity above college median	Coauthors with interdisciplinarity below college median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
log(Variety)	0.037^{***}	0.045^{***}	0.028^{***}	0.041^{***}	0.038^{***}	0.040^{***}	0.040^{***}	0.035^{***}	0.037^{***}	0.040^{***}
	(0.003)	(0.005)	(0.005)	(0.003)	(0.005)	(0.003)	(0.003)	(0.005)	(0.003)	(0.004)
log(Balance + 1)	0.244^{***}	0.221^{***}	0.309^{***}	0.217^{***}	0.260^{**}	0.228^{***}	0.265^{***}	0.236^{***}	0.280^{***}	0.220^{***}
	(0.029)	(0.047)	(0.042)	(0.029)	(0.080)	(0.026)	(0.034)	(0.046)	(0.035)	(0.040)
log(Disparity + 1)	-0.149^{***}	-0.432^{***}	-0.147	-0.195^{***}	-0.280^{*}	-0.198^{***}	-0.207^{***}	-0.230^{***}	-0.215^{**}	-0.231^{***}
	(0.042)	(0.078)	(0.092)	(0.042)	(0.109)	(0.040)	(0.056)	(0.064)	(0.078)	(0.051)
log(Number of Authors)	0.011^{***}	0.012^{***}	0.008^{***}	0.013^{***}	0.013^{***}	0.011^{***}	0.012^{***}	0.012^{***}	0.014^{***}	0.011^{***}
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)
International Collaboration	0.0003	0.003			0.001	0.001	−0.0003	0.001	0.001	0.001
	(0.002)	(0.003)			(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
log(H-index + 1)	−0.004	0.004	-0.009^{*}	−0.001	−0.021	−0.001	$-0.007^{\dagger}$	0.005	−0.006	0.001
	(0.003)	(0.004)	(0.005)	(0.003)	(0.014)	(0.002)	(0.004)	(0.005)	(0.004)	(0.004)
IMR	-0.073^{***}	$-0.064^{\dagger}$	-0.165^{***}	-0.048^{**}	−0.003	-0.067^{***}	-0.070^{***}	-0.070^{*}	-0.074^{**}	-0.051^{*}
	(0.016)	(0.033)	(0.048)	(0.017)	(0.040)	(0.016)	(0.021)	(0.028)	(0.025)	(0.023)
Variety = 1	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Fields of Study Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Researcher Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Number of Researchers	3,836	2,028	2,562	5,631	131	5,907	3,499	3,994	3,533	3,769
Observations	33,007	10,618	8,561	35,523	4,974	39,110	24,806	14,737	19,684	19,859
R^2	0.326	0.397	0.558	0.347	0.206	0.355	0.332	0.457	0.372	0.364
Adjusted R^2	0.236	0.252	0.366	0.223	0.180	0.240	0.221	0.253	0.233	0.213

Notes: This table presents second stage results from Heckman's two-steps estimation of the effects of interdisciplinarity on the diffusion of knowledge across fields, following equation 1, for subsamples divided by gender (columns 1-2), the presence of international collaborators (columns 3-4), productivity (columns 5-6), coauthors interdisciplinarity (columns 7-8), and the average coauthors interdisciplinarity measured at the college level (columns 9-10). Observations are at the paper-researcher level. The dependent variable is the logarithm of the generality index, defined in equation 5. All regressions include the Inverse Mills Ratio (IMR) to control for sample selection bias and individual, year, and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: †p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

Table 8: Interdisciplinarity effects on the diffusion of knowledge across fields and college affiliation.

	Dependent variable: log(Generality + 1)			
	CLAS	ENG	HSC	IFAS
	(1)	(2)	(3)	(4)
log(Variety)	0.026^{***}	0.037^{***}	0.037^{***}	0.051^{***}
	(0.005)	(0.008)	(0.004)	(0.006)
log(Balance + 1)	0.099^{*}	0.275^{**}	0.269^{***}	0.214^{***}
	(0.048)	(0.100)	(0.040)	(0.058)
log(Disparity + 1)	−0.101	$-0.187^{\dagger}$	-0.140^{*}	-0.463^{***}
	(0.070)	(0.100)	(0.063)	(0.105)
log(Number of Authors)	0.014^{***}	0.014^{*}	0.010^{***}	$0.006^{\dagger}$
	(0.003)	(0.006)	(0.002)	(0.003)
International Collaboration	-0.009^{*}	0.001	0.005^{**}	0.005
	(0.004)	(0.005)	(0.002)	(0.003)
log(H-index + 1)	−0.007	0.004	0.006	-0.015^{**}
	(0.007)	(0.008)	(0.003)	(0.005)
IMR	−0.038	−0.020	-0.088^{***}	$-0.065^{\dagger}$
	(0.031)	(0.055)	(0.022)	(0.039)
Variety = 1	YES	YES	YES	YES
Fields of Study Fixed Effects	YES	YES	YES	YES
Year Fixed Effects	YES	YES	YES	YES
Researcher Fixed Effects	YES	YES	YES	YES
Number of Researchers	620	385	2,151	1,172
Observations	5,024	4,741	17,389	9,368
R^2	0.386	0.243	0.285	0.335
Adjusted R^2	0.294	0.171	0.183	0.237

Notes: This table presents second stage results from Heckman's two-steps estimation of the effects of interdisciplinarity on the diffusion of knowledge across fields, following equation 1, for subsamples divided by academic unit affiliation. Column 1 estimates the effects for researchers affiliated to the College of Liberal Arts and Science, column 2 for those at the College of Engineering, column 3 for those at the Health Science Center and column 4 for those at the Institute of Food and Agricultural Sciences. Observations are at the paper-researcher level. The dependent variable is the total citations accrued in five years. All regressions include the Inverse Mills Ratio (IMR) to control for sample selection bias and individual, year and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

We now replicate our exercise by measuring scientific impact in terms of the diffusion of knowledge across disciplines, to test HP1b. Our results are reported in Table 7 and Table 8. Also in this case, we find evidence that researchers face the same dilemma when engaging in interdisci-

and also using department-level affiliation. All our results are qualitatively unchanged. Results are available upon request.

plinary work, regardless of their specific characteristics[27]. At the same time, when considering researchers' affiliation, we observe that the effect of Disparity is negative but no longer significant for the researchers at the College of Liberal Arts and Science (column (1)). This may suggest that researchers in social sciences, humanities, and hard sciences like physics are not penalized as much for combining dissimilar disciplines as more "applied" fields like engineering, health, and agricultural science.[28] In all the other cases, however, we still find evidence that Variety and Balance have a positive and statistically significant impact on the diffusion of knowledge across disciplines for all considered subgroups, while Disparity has a negative and statistically significant effect.

Finally, we investigate the validity of HP2 by comparing the results obtained when using different measures of scientific impact. Notably, we still find that Balance has strong but opposite effects on the number of citations and generality of knowledge. Papers with lower Balance have more citations but reach a less diverse audience of academics, regardless of the characteristics of the group considered. In other words, we find evidence that HP2 applies to all research profiles: i.e., all of them are subject to an interdisciplinary dilemma in their work.

Taken together, our results suggest that all scholars face the similar incentives and constraints in engaging in more interdisciplinary projects. Regardless of their characteristics or affiliation, the effects of IDR are large and widespread, and affect all research activities at the University of Florida.

7 Conclusion

Our results bring evidence to the idea that multiple logics within the academia might create contrasting incentives for scholars. In our study, we highlight that policies that govern hiring and evaluation within universities and policy that sustain interdisciplinarity incentivize behaviors that are, at least to a certain extent, incompatible. It is not always possible to act as to accumulate citations from published papers while combining knowledge from different domains: scholars are forced to trade-off between reputation and societal impact of their research.

[27] As a robustness check, we have estimated the model in columns (7-9) in Table 7 adding to our specification a dummy variable registering if a paper was co-authored exclusively by UF investigators and interactions between this dummy and our interdisciplinarity measures. We find that our results are qualitatively unchanged. Results are available upon request.

[28] Differences might be due to evaluation criteria that vary across sciences. For instance, Guetzkow et al. (2004) maintain that social sciences and humanities rely mainly on originality that, in their study, includes disciplinary variation.

Nowadays, the soaring amount of knowledge accumulated in published articles requires doctoral programs and post-doctoral training of longer duration (Jones, 2010). This, in turn, postpone first publishing (Conti and Liu, 2015) and the 'age of great achievement' (Jones et al., 2014; The National Academies, 1998). To compensate for such burden of knowledge, scientists often seek to (over)specialize in specific fields, making interdisciplinarity a necessary choice to ensure scientific communication and societal progress. Therefore, the need for appropriate incentives and coordinated policy is particularly urgent.

As the literature on IDR grows and gets more sophisticated, we encourage more investigations at the level of researchers to fully grasp the implications of choosing an interdisciplinary approach. Although our results are robust to various specifications and definitions of scholars' samples, further research is needed to corroborate the external validity of our analysis by including information on research activities in more than one university.

Acknowledgements

This research uses data from the Bureau of Economics and Business Research (BEBR) of the University of Florida. We thank the staff of the BEBR and the director Christopher McCarty for their important contribution to this work. We are also grateful for helpful comments from Ronald Rousseau, Jacques Mairesse, participants at the Regional Quantitative Analysis Group (AQR) seminar series of the University of Barcelona and at the WOEPSR 2022 workshop at KU Leuven.

References

Angrist, J., P. Azoulay, G. Ellison, R. Hill, and S. F. Lu (2020). "Inside job or deep impact? Extramural citations and the influence of economic scholarship." *Journal of Economic Literature* 58(1), 3–52. 10.1257/jel.20181508.

Arnold, A. et al. (2021). "Perspective: Promoting and fostering multidisciplinary research in universities." *Research Policy* 50(9), 104334. 10.1016/j.respol.2021.104334.

Battiston, F. et al. (2019). "Taking census of physics." *Nature Reviews Physics* 1(1), 89–97. 10.1038/s42254-018-0005-3.

Biancani, S., L. Dahlander, D. A. McFarland, and S. Smith (2018). “Superstars in the making? The broad effects of interdisciplinary centers.” *Research Policy* 47(3), 543–557. 10.1016/j.respol.2018.01.014.

Bresnahan, T. F. and M. Trajtenberg (1995). “General purpose technologies ‘Engines of growth’?” *Journal of Econometrics* 65(1), 83–108. 10.1016/0304-4076(94)01598-T.

Bromham, L., R. Dinnage, and X. Hua (2016). “Interdisciplinary research has consistently lower funding success.” *Nature* 534, 684–687. 10.1038/nature18315.

Carley, S. and A. L. Porter (2012). “A forward diversity index.” *Scientometrics* 90(2), 407–427. 10.1007/s11192-011-0528-1.

Cedrini, M. and M. Fontana (2018). “Just another niche in the wall? How specialization is changing the face of mainstream economics.” *Cambridge Journal of Economics* 42(2), 427–451. 10.1093/cje/bex003.

Conti, A. and C. C. Liu (2015). “Bringing the lab back in: Personnel composition and scientific output at the MIT department of biology.” *Research Policy* 44(9), 1633–1644. 10.1016/j.respol.2015.01.001.

Fontana, M., M. Iori, F. Montobbio, and R. Sinatra (2020). “New and atypical combinations: An assessment of novelty and interdisciplinarity.” *Research Policy* 49(7), 104063. 10.1016/j.respol.2020.104063.

Goel, R. K. and J. R. Faria (2007). “Proliferation of academic journals: Effect on research quantity and quality.” *Metroeconomica* 58(4), 536–549. 10.1111/j.1467-999X.2007.00285.x.

Gruber, M., D. Harhoff, and K. Hoisl (2013). “Knowledge recombination across technological boundaries: Scientists vs. engineers.” *Management Science* 59(4), 837–851.

Guetzkow, J., M. Lamont, and G. Mallard (2004). “What is originality in the humanities and the social sciences?” *American Sociological Review* 69(2), 190–212. 10.1177/000312240406900203.

Hackett, E. J. et al. (2021). “Do synthesis centers synthesize? A semantic analysis of topical diversity in research.” *Research policy* 50(1), 104069. 10.1016/j.respol.2020.104069.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001). “The NBER patent citation data file: Lessons, insights and methodological tools.” *NBER Working Paper* 8498. 10.3386/w8498.

Hamermesh, D. S. and G. A. Pfann (2012). “Reputation and earnings: The roles of quality and quantity in academe.” *Economic Inquiry* 50(1), 1–16. 10.1111/j.1465-7295.2011.00381.x.

Jamali, H., D. Nicholas, and E. Herman (2016). *Emerging reputation mechanisms for scholars: A literature-based theoretical framework of scholarly activities and a state-of-the-art appraisal of the social networking services used by scholars, to build, maintain and showcase their reputations.* Luxemburg: Joint Research Center Publications Office of the European Union. 10.2791/832930.

Jha, Y. and E. W. Welch (2010). "Relational mechanisms governing multifaceted collaborative behavior of academic scientists in six fields of science and engineering." *Research Policy* 39(9), 1174–1184. 10.1016/j.respol.2010.06.003.

Jones, B., E. J. Reedy, and B. A. Weinberg (2014). "Age and scientific genius." *NBER Working Paper* w19866. 10.3386/w19866.

Jones, B. F. (2010). "Age and great invention." *The Review of Economics and Statistics* 92(1), 1–14. 10.1162/rest.2009.11724.

— (2009). "The burden of knowledge and the "Death of the Renaissance Man": Is innovation getting harder?" *The Review of Economic Studies* 76(1), 283–317. 10.1111/j.1467-937X.2008.00531.x.

— (2021). "The rise of research teams: Benefits and costs in economics." *Journal of Economic Perspectives* 35(2), 191–216. 10.1257/jep.35.2.191.

Larsen, P. O. and M. von Ins (2010). "The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index." *Scientometrics* 84(3), 575–603. 10.1007/s11192-010-0202-z.

Leahey, E. and S. N. Barringer (2020). "Universities' commitment to interdisciplinary research: To what end?" *Research Policy* 49(2), 103910. 10.1016/j.respol.2019.103910.

Leahey, E., C. M. Beckman, and T. L. Stanko (2017). "Prominent but less productive: The impact of interdisciplinarity on scientists' research." *Administrative Science Quarterly* 62(1), 105–139. 10.1177/0001839216665364.

Leone Sciabolazza, V., R. Vacca, T. Kennelly Okraku, and C. McCarty (2017). "Detecting and analyzing research communities in longitudinal scientific networks." *PloS one* 12(8), e0182516. 10.1371/journal.pone.0182516.

Leone Sciabolazza, V., R. Vacca, and C. McCarty (2020). "Connecting the dots: Implementing and evaluating a network intervention to foster scientific collaboration and productivity." *Social Networks* 61, 181–195. 10.1016/j.socnet.2019.11.003.

Llopis, O., P. D'Este, M. McKelvey, and A. Yegros (2022). "Navigating multiple logics: Legitimacy and the quest for societal impact in science." *Technovation* 110, 102367. 10.1016/j.technovation.2021.102367.

Mazzucato, M. (2018). "Mission-oriented innovation policies: challenges and opportunities." *Industrial and Corporate Change* 27(5), 803–815. 10.1093/icc/dty034.

Microsoft Research (2020). Multi-Sense network representation learning in Microsoft Academic Graph. Microsoft Research. www.microsoft.com/en-us/research/project/academic/articles/multi-sense-network-representation-learning-in-microsoft-academic-graph/ (visited on 10/10/2021).

Porter, A. and I. Rafols (2009). "Is science becoming more interdisciplinary? Measuring and mapping six research fields over time." *Scientometrics* 81(3), 719–745. 10.1007/s11192-008-2197-2.

Ritzberger, K. (2008). "A ranking of journals in economics and related fields." *German Economic Review* 9(4), 402–430. 10.1111/j.1468-0475.2008.00447.x.

Rylance, R. (2015). "Grant giving: Global funders to focus on interdisciplinarity." *Nature News* 525(7569), 313. 10.1038/525313a.

Singh, J. and L. Fleming (2010). "Lone Inventors as Sources of Breakthroughs: Myth or Reality?" *Management Science* 56(1), 41–56. 10.1287/mnsc.1090.1072.

Sinha, A. et al. (2015). "An overview of Microsoft Academic Service (MAS) and applications." In: *Proceedings of the 24th International Conference on World Wide Web*, 243–246. 10.1145/2740908.2742839.

Smith, T. B., R. Vacca, T. Krenz, and C. McCarty (2021). "Great minds think alike, or do they often differ? Research topic overlap and the formation of scientific teams." *Journal of informetrics* 15(1), 101104. 10.1016/j.joi.2020.101104.

Squicciarini, M., H. Dernis, and C. Criscuolo (2013). "Measuring patent quality: Indicators of technological and economic value." *OECD Science, Technology and Industry Working Papers* 2013/03. 10.1787/18151965.

Sterzi, V., M. Pezzoni, and F. Lissoni (2019). "Patent management by universities: evidence from Italian academic inventions." *Industrial and Corporate Change* 28(2), 309–330. 10.1093/icc/dty070.

Stigler, G. J., S. M. Stigler, and C. Friedland (1995). "The journals of economics." *Journal of Political Economy* 103(2), 331–359. 10.1086/261986.

Stirling, A. (2007). "A general framework for analysing diversity in science, technology and society." *Journal of The Royal Society Interface*. 10.1098/rsif.2007.0213.

Subramanian, A. M., K. Lim, and P.-H. Soh (2013). "When birds of a feather don't flock together: Different scientists and the roles they play in biotech R&D alliances." *Research Policy* 42(3), 595–612. 10.1016/j.respol.2012.12.002.

Sun, Y., G. Livan, A. Ma, and V. Latora (2021). "Interdisciplinary researchers attain better long-term funding performance." *Communications Physics* 4(263). 10.1038/s42005-021-00769-z.

Tartari, V. and S. Breschi (2012). “Set them free: scientists’ evaluations of the benefits and costs of university–industry research collaboration.” *Industrial and Corporate Change* 21(5), 1117–1147. 10.1093/icc/dts004.

The National Academies (1998). *Trends in the Early Careers of Life Scientists*. Washington, DC: National Academies Press. 10.17226/6244.

— (2005). *Facilitating interdisciplinary research*. Washington, DC: National Academies Press. 10.17226/11153.

Trajtenberg, M., R. Henderson, and A. Jaffe (1997). “University Versus Corporate Patents: A Window On The Basicness Of Invention.” *Economics of Innovation and New Technology* 5(1), 19–50. 10.1080/10438599700000006.

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. 10.1016/j.joi.2010.06.004.

Wang, K. et al. (2019). “A review of Microsoft Academic Services for science of science studies.” *Frontiers in Big Data* 2. 10.3389/fdata.2019.00045.

Wang, Q. and J. W. Schneider (2020). “Consistency and validity of interdisciplinarity measures.” *Quantitative Science Studies* 1(1), 239–263. 10.1162/qss_a_00011.

Wuchty, S., B. F. Jones, and B. Uzzi (2007). “The increasing dominance of teams in production of knowledge.” *Science* 316(5827), 1036–1039. 10.1126/science.1136099.

Yegros-Yegros, A., I. Rafols, and P. D’Este (2015). “Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity.” *PLoS ONE* 10(8), e0135095. 10.1371/journal.pone.0135095.

Zeng, A. et al. (2017). “The science of science: From the perspective of complex systems.” *Physics Reports* 714-715, 1–73. 10.1016/j.physrep.2017.10.001.

A Data appendix

In this appendix, we report the procedure we followed to construct the database used in the analysis and some additional descriptive statistics.

A.1 Data collection

From the data collected by the Bureau of Economics and Business Research (BEBR) of the University of Florida (UF), we retrieved information concerning publication records, department affiliation, and gender for the universe of UF's researchers in the period 2008-2013. Each researcher is identified by a unique code (UFID). Raw publication records provide information regarding 34,851 scholarly works including journal name, article title, and publication year. Based on UF registered publications' information, we retrieved the publications' Digital Object Identifiers (DOI) from Crossref and Scopus databases.[29] This procedure allows us to identify researchers' academic output that was indexed in the largest and most common scholarly works' databases.

More specifically, we used an automated script to extract bibliographic metadata of UF publications available in the original dataset through Scopus Database API Interface and Crossref REST API.[30] The three main steps of this procedure are the following:

1. Get articles partial metadata based on publication title: From titles of publications in the UF records, the script – through queries to Scopus and Crossref APIs – collects publications matching our list of articles' titles and retrieves their metadata (DOI, journal name, publication title, publication year). We collect the first ten results of the queries for each title and store them in a new database.

2. Cleaning and processing article's title: The article titles in the raw data and in the data retrieved by API queries are cleaned and then processed. Cleaning consists in eliminating spaces, special characters, and punctuation. Processing consists in coercing characters to lowercase and comparing the raw (original) and newly extracted titles.

3. Title-DOI matching procedure: Matches are determined according to a fuzzy matching algo-

[29]The databases are available at the following webpage: Crossref and Scopus.

[30]Data collection using Scopus and Crossref occurred in 2018.

rithm implemented in the *fuzzywuzzy* text similarity package in Python.[31] The script considers a match if titles have a higher than 90% similarity ratio and the matching is unique. Matched publications and its respective metadata are assigned to the associated researcher. Unique matches with more than one DOI were manually checked and disambiguated. Publications without a unique match are dropped.

With this procedure, we were able to identify the DOIs of 28,239 publications of our original database. Using these DOIs, we collect the full metadata through Lens and Microsoft Academic Graph (MAG) databases.[32] Metadata from Lens API platform includes: IDs (Lens articles ID, Microsoft Academic Graph ID); publication type (journal article, book, working paper); list of citations; list of references; fields of study (computed by the MAG algorithm as described in Section 5); and authors' affiliations. We decided to focus on Lens database to collect citations and references data because it also provides their disciplines based on the natural language processing algorithms used by MAG. Furthermore, the Lens' scholarly citation data, contrary to Microsoft Academic Graph, indexes only publications of selected document types (journal article, book, working paper).[33] Publications missing references or missing fields of study are dropped. In addition, we restrict our sample to only journal articles. Our final database consists of 23,926 articles and their full metadata.

In the last data collection phase, we extracted from MAG a proximity measure between the fields of study using the functionality Network Similarity Package, as described in Section 5.[34] We collected similarity scores for all possible combinations between the 19 fields (first level of classification) and 292 subfields (second level of classification).

A.2 Additional descriptive statistics

Figure A1 shows the evolution of the total number of publications in our database (in the period 2008–2013). Table A1 reports the distribution of researchers across academic units and colleges, while A2 shows the distribution of citations by field of study. Table A3 shows summary statistics at the paper-researcher level. Finally, Table A4 reports the correlation between variables used in

[31]Documentation about *fuzzywuzzy* is available here: fuzzywuzzy.
[32]These databases are available at the following link: Microsoft Academic Graph and Lens.
[33]Data collection using the Lens occurred in 2019.
[34]Data collection Microsoft Academic Graph occurred in 2020.

our regression analysis.

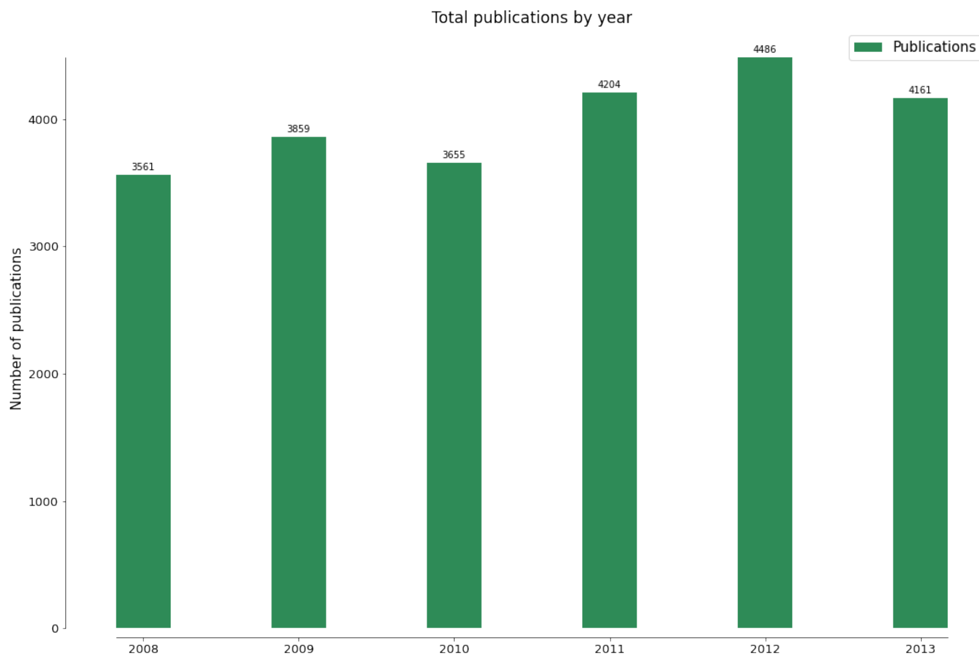

Figure A1: Number of total publications by year.

A.3 Motivating evidence: wages and grants

This section discusses data about wages and grants received by a sample of 3,481 UF's researchers. We exploit this data to conduct a preliminary investigation on researchers' trade-offs in pursing IDR (see Section 2).

To perform this analysis, we compute an aggregated indicator of interdisciplinarity profile at the researcher level, since the yearly wage and the number of grants refer to scholars. This indicator is equal to the maximum value of the number of cited fields of study (second level of MAG classification) found among the articles written by a researcher in a given year. We rely on MAG, instead, to define an indicator of scholar seniority: the variable academic age measures the time that a researcher has been active and is defined as the number of years between their first published work until the year of observation. Table A5 shows descriptive statistics for these variables, while Table A6 reports the results of our preliminary regression analysis.

Table A1: Number of researchers by academic unit and college.

Academic Units	**Colleges**	**Researchers**
Liberal Arts and Sciences	College of Liberal Arts and Sciences	665
Engineering	College of Engineering	389
Health Sciences	Medicine	1545
	Medicine-Jacksonville	175
	Public Health and Health Professions	166
	Pharmacy	140
	Dentistry	139
	Nursing	36
	Health Affairs	14
Food and Agricultural Sciences	Agricultural And Life Sciences	978
	Veterinary Medicine	218
	Institute of Food and Agricultural Sciences	2
Other	UF Students	946
	Uncategorized Departments	687

Notes: This table shows the distribution researchers affiliated to the academic units and colleges at the University of Florida (UF) from 2008 to 2013. The total number of researchers with a college affiliation is 5,130. Researchers that are not affiliated to any specific academic unit are counted in the category "Other". Researchers classified as students in the UF registry office are counted in "UF Students" and faculty affiliated to departments not belonging to any college are counted in "Uncategorized Departments".

Table A2: Distribution of citations by field of study (first level of classification).

Field of Study	**References**	**Citations**
Art	5490	1393
Biology	1181592	359734
Business	21831	6804
Chemistry	329228	101792
Computer science	88408	23706
Economics	64521	15417
Engineering	79673	28339
Environmental science	40883	14674
Geography	27317	6791
Geology	89619	25773
History	8581	1026
Materials science	54980	29450
Mathematics	101585	19532
Medicine	1071812	378978
Philosophy	10873	2082
Physics	225173	78280
Political science	10374	2585
Psychology	233354	62381
Sociology	38379	7401

Notes: This table shows the distribution of the documents of each field in the focal papers' references and which cited our focal paper (citations). The total number of documents referenced is 646,280 and the total number of citations is 366,024

Table A3: Summary statistics at the article-researcher level.

Variables	*Mean*	*SD*	*Min*	*50%*	*Max*	*Obs*
Nb. Citations	20.73	43.02	0	11	2,530	46,156
Generality	0.73	0.17	0	0.78	0.98	44,084
Variety	39.56	18.97	1	39	153	46,156
Balance	0.84	0.08	0	0.85	1	46,156
Disparity	0.69	0.06	0	0.71	0.94	46,156
Nb. of Authors	6.30	7.98	1	5	1,269	46,156
International Collab.	0.19	0.39	0	0	1	46,156
H-index	6.29	7.00	0	4	54	46,156

Notes: Nb. Citations is the total number of citations received in a 5 years time after the publication. Generality captures the degree of applicability of the knowledge codified in a paper on different fields of study. It is worth noting that Generality is not defined for papers with zero citations. International Collaboration is a dummy variable that assumes the value 1 when at least one co-author in the paper is affiliated to an institution outside the United States.

Table A4: Correlation table between variables used in regressions.

	Number Citations	Generality	Variety	Balance	Disparity	Number References	Number Authors	International Collaboration
Number Citations	1	0.18	0.18	-0.12	0.09	0.33	0.45	0.11
Generality	0.18	1	0.23	0.12	0.12	0.16	0.06	0.03
Variety	0.18	0.23	1	-0.05	0.57	0.67	0.07	0.04
Balance	-0.12	0.12	-0.05	1	0.10	-0.41	-0.08	-0.11
Disparity	0.09	0.12	0.57	0.10	1	0.28	0.06	0.02
Number References	0.33	0.16	0.67	-0.41	0.28	1	0.20	0.09
Number of Authors	0.45	0.06	0.07	-0.08	0.06	0.20	1	0.16
International Collaboration	0.11	0.03	0.04	-0.11	0.02	0.09	0.16	1

Table A5: Additional descriptive statistics on wages and grants.

Variables	*Mean*	*SD*	*Min*	*50%*	*Max*	*Obs*
Wage	127199.10	90411.25	100	99553.2	997465	11,160
Nb. Grants	0.55	1.25	0	0	21	18,005
Interdisciplinarity	46.34	20.69	1	46	153	18,005
Academic Age	15.54	11.44	1	13	53	14,606

Notes: These variable are available only for a subsample of UF's researchers. The variable Interdisciplinarity is equal to the maximum value of the number of cited fields of study found among the articles written by a researcher in a given year. Academic Age measures the time that a researcher has been active in a research field and is defined as the number of years between their first published work until the year of observation.

Table A6: Correlation between scholars' interdisciplinarity profile and academic achievements.

	Dependent variable:			
	log(Wage)		log(Nb. Grants + 1)	
	(1)	(2)	(3)	(4)
Interdisciplinarity	-0.001^{*}	-0.002^{***}	0.004^{***}	0.003^{***}
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Academic Age		0.018^{***}		0.010^{***}
		(0.001)		(0.001)
Constant	11.578^{***}	11.309^{***}	0.224^{***}	0.118^{***}
	(0.015)	(0.018)	(0.013)	(0.017)
Number of Researchers	3,481	2,785	3,481	2,785
Observations	11,160	9,444	11,160	9,444
R^2	0.0004	0.091	0.022	0.045
Adjusted R^2	0.0003	0.091	0.022	0.045

Notes: Estimated coefficients and standard errors (parentheses) obtained with ordinary least square estimations. The dependent variables are the logarithm of yearly wages (columns 1-2) and the number of awarded grants to a researcher in a year (columns 3-4). The variable Interdisciplinarity is equal to the maximum value of the number of cited fields of study found among the articles written by a researcher in a given year. Academic Age measures the time that a researcher has been active and is defined as the number of years between their first published work until the year of observation. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

B Robustness checks and first stage of Heckman correction

Table B1 allows comparing the estimations obtained through the use of OLS with those resulting from Poisson and negative binomial. As evident in the table, our results are robust to the different estimation approaches.

Table B1: Results using OLS, Poisson, and negative binomial to estimate the effect of IDR con the number of citations.

Dependent Variables:	log(Nb. of Citations+1)	Nb. of Citations	Nb. of Citations
	(1)	(2)	(3)
	OLS	*Poisson*	*Neg. Bin.*
log(Variety)	0.5499***	0.6920***	0.6211***
	(0.0146)	(0.0350)	(0.0226)
log(Balance + 1)	-4.552***	-4.410***	-4.526***
	(0.1910)	(0.4070)	(0.2432)
log(Disparity + 1)	-1.268***	-2.296***	-1.518***
	(0.2287)	(0.5492)	(0.3545)
log(Number of Authors)	0.4454***	0.5343***	0.4793***
	(0.0126)	(0.0351)	(0.0192)
International Collaboration	0.0376**	0.1133***	0.0615***
	(0.0149)	(0.0360)	(0.0227)
log(H-index + 1)	0.1266***	0.2715***	0.1506***
	(0.0196)	(0.0328)	(0.0233)
Variety = 1	YES	YES	YES
Fields of Study Fixed Effects	YES	YES	YES
Year Fixed Effects	YES	YES	YES
Researcher Fixed Effects	YES	YES	YES
Observations	46,159	45,974	45,974
Squared Correlation	0.47950	0.39767	0.31271
Pseudo R^2	0.21431	0.44458	0.09057
BIC	176,539.3	954,572.6	402,511.9
Over-dispersion			1.6172

Notes: This table presents OLS, Poisson, and negative binomial estimates of the effects of interdisciplinarity on the number of citations. Observations are at the paper-researcher level. The dependent variable is the logarithm of total citations accrued in five years in the first specification and the total number of citations in five years in the other specifications. All regressions include individual, year and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

Table B2, instead, reports the first stage of Heckman correction used to estimate the effect of IDR on Generality. Generality is, indeed, only defined for articles that receive at least one citation.

Table B2: First stage of the Heckman correction.

	Dependent variable:			
	Cited Paper			
	probit			
	(1)	(2)	(3)	(4)
log(Variety)	0.762***	0.653***	0.653***	0.645***
	(0.023)	(0.023)	(0.023)	(0.023)
log(Balance + 1)	−6.193***	−6.054***	−6.023***	−5.916***
	(0.327)	(0.331)	(0.332)	(0.332)
log(Disparity + 1)	−1.479***	−1.797***	−1.782***	−1.769***
	(0.315)	(0.324)	(0.324)	(0.324)
log(Number of Authors)		0.496***	0.491***	0.479***
		(0.021)	(0.021)	(0.021)
International Collaboration			0.055	0.040
			(0.037)	(0.037)
log(H-index + 1)				0.115***
				(0.016)
Constant	3.836***	3.571***	3.545***	3.424***
	(0.238)	(0.243)	(0.243)	(0.243)
Variety = 1	YES	YES	YES	YES
Fields of Study Fixed Effects	YES	YES	YES	YES
Year Fixed Effects	YES	YES	YES	YES
Researcher Fixed Effects	NO	NO	NO	NO
Number of Researchers	6,105	6,105	6,105	6,105
Observations	46,173	46,156	46,156	46,156
Log Likelihood	-6,555.652	-6,247.971	-6,246.868	-6,219.912
Akaike Inf. Crit.	13,195.300	12,581.940	12,581.740	12,529.830

Notes: This table presents first stage results from Heckman's two-steps estimation. Observations are at the paper-researcher level. Estimates stem from probit specifications with dependent variable being a dummy that assumes value 1 when a paper was cited in a 5 year time-window after the publication and 0 otherwise. All regressions include year and fields of study fixed effects. Standard errors are clustered at the researcher level. Significance levels: † p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

C Fields of study classification

Table C1 reports the conversion table, made by authors, between the first and the second level of fields of studies, as classified by MAG. The first level classify articles in 19 disciplines, while the second one has 292 possible values, corresponding to sub-disciplines. The table also include the ID used to represent fields of studies at the second level in the knowledge space (Figure 3).

ID	2^{nd} level	1^{st} level
0	Visual arts	Art
1	Classics	Art
2	Art history	Art
3	Literature	Art
4	Linguistics	Art
5	Communication	Art
6	Library science	Art
7	Humanities	Art
8	Zoology	Biology
9	Botany	Biology
10	Evolutionary biology	Biology
11	Computational biology	Biology
12	Cell biology	Biology
13	Molecular biology	Biology
14	Animal science	Biology
15	Astrobiology	Biology
16	Microbiology	Biology
17	Food science	Biology
18	Biotechnology	Biology
19	Biological system	Biology
20	Economic system	Business
21	Financial system	Business
22	Commerce	Business
23	Knowledge management	Business
24	Process management	Business
25	Marketing	Business
26	Public relations	Business
27	Advertising	Business
28	Accounting	Business
29	Operations research	Business
30	Management	Business
31	Operations management	Business
32	Management science	Business
33	Business administration	Business
34	Geochemistry	Chemistry
35	Computational chemistry	Chemistry
36	Physical chemistry	Chemistry
37	Organic chemistry	Chemistry
38	Stereochemistry	Chemistry
39	Environmental chemistry	Chemistry
40	Inorganic chemistry	Chemistry
41	Photochemistry	Chemistry
42	Combinatorial chemistry	Chemistry
43	Polymer chemistry	Chemistry
44	Analytical chemistry	Chemistry

ID	2^{nd} level	1^{st} level
45	Medicinal chemistry	Chemistry
46	Biochemistry	Chemistry
47	Nuclear chemistry	Chemistry
48	Chromatography	Chemistry
49	Radiochemistry	Chemistry
50	Toxicology	Chemistry
51	Pharmacology	Chemistry
52	Embedded system	Computer science
53	Distributed computing	Computer science
54	Computer network	Computer science
55	Artificial intelligence	Computer science
56	Pattern recognition	Computer science
57	Computer vision	Computer science
58	Machine learning	Computer science
59	Real-time computing	Computer science
60	World Wide Web	Computer science
61	Information retrieval	Computer science
62	Internet privacy	Computer science
63	Computer security	Computer science
64	Operating system	Computer science
65	Human–computer interaction	Computer science
66	Multimedia	Computer science
67	Natural language processing	Computer science
68	Data mining	Computer science
69	Programming language	Computer science
70	Theoretical computer science	Computer science
71	Algorithm	Computer science
72	Data science	Computer science
73	Database	Computer science
74	Bioinformatics	Computer science
75	Parallel computing	Computer science
76	Computer graphics (images)	Computer science
77	Computational science	Computer science
78	Speech recognition	Computer science
79	International economics	Economics
80	International trade	Economics
81	Market economy	Economics
82	Econometrics	Economics
83	Macroeconomics	Economics
84	Monetary economics	Economics
85	Economic policy	Economics
86	Positive economics	Economics
87	Neoclassical economics	Economics
88	Industrial organization	Economics
89	Finance	Economics
90	Natural resource economics	Economics
91	Environmental economics	Economics
92	Keynesian economics	Economics
93	Political economy	Economics
94	Development economics	Economics
95	Economic history	Economics
96	Agricultural economics	Economics
97	Economy	Economics
98	Financial economics	Economics
99	Labour economics	Economics
100	Demographic economics	Economics
101	Law and economics	Economics

ID	2^{nd} level	1^{st} level
102	Economic growth	Economics
103	Public economics	Economics
104	Microeconomics	Economics
105	Classical economics	Economics
106	Mathematical economics	Economics
107	Welfare economics	Economics
108	Computer hardware	Engineering
109	Electronic engineering	Engineering
110	Electrical engineering	Engineering
111	Systems engineering	Engineering
112	Software engineering	Engineering
113	Control engineering	Engineering
114	Control theory	Engineering
115	Environmental engineering	Engineering
116	Mechanics	Engineering
117	Manufacturing engineering	Engineering
118	Industrial engineering	Engineering
119	Mechanical engineering	Engineering
120	Engineering drawing	Engineering
121	Aerospace engineering	Engineering
122	Aeronautics	Engineering
123	Construction engineering	Engineering
124	Engineering management	Engineering
125	Geotechnical engineering	Engineering
126	Civil engineering	Engineering
127	Pulp and paper industry	Engineering
128	Structural engineering	Engineering
129	Agricultural engineering	Engineering
130	Optoelectronics	Engineering
131	Computer architecture	Engineering
132	Architectural engineering	Engineering
133	Chemical engineering	Engineering
134	Risk analysis (engineering)	Engineering
135	Reliability engineering	Engineering
136	Computer engineering	Engineering
137	Transport engineering	Engineering
138	Process engineering	Engineering
139	Biochemical engineering	Engineering
140	Petroleum engineering	Engineering
141	Automotive engineering	Engineering
142	Telecommunications	Engineering
143	Forensic engineering	Engineering
144	Remote sensing	Engineering
145	Marine engineering	Engineering
146	Simulation	Engineering
147	Mining engineering	Engineering
148	Nuclear engineering	Engineering
149	Biomedical engineering	Engineering
150	Atmospheric sciences	Environmental science
151	Meteorology	Environmental science
152	Climatology	Environmental science
153	Environmental resource management	Environmental science
154	Environmental planning	Environmental science
155	Agricultural science	Environmental science
156	Waste management	Environmental science
157	Agronomy	Environmental science
158	Horticulture	Environmental science

ID	2^{nd} level	1^{st} level
159	Hydrology	Environmental science
160	Soil science	Environmental science
161	Environmental protection	Environmental science
162	Ecology	Environmental science
163	Agroforestry	Environmental science
164	Water resource management	Environmental science
165	Geomorphology	Environmental science
166	Forestry	Environmental science
167	Earth science	Environmental science
168	Oceanography	Environmental science
169	Fishery	Environmental science
170	Environmental health	Environmental science
171	Regional science	Geography
172	Economic geography	Geography
173	Geodesy	Geography
174	Physical geography	Geography
175	Cartography	Geography
176	Petrology	Geology
177	Mineralogy	Geology
178	Paleontology	Geology
179	Crystallography	Geology
180	Archaeology	History
181	Ancient history	History
182	Genealogy	History
183	Metallurgy	Materials science
184	Composite material	Materials science
185	Ceramic materials	Materials science
186	Nanotechnology	Materials science
187	Polymer science	Materials science
188	Combinatorics	Mathematics
189	Discrete mathematics	Mathematics
190	Pure mathematics	Mathematics
191	Algebra	Mathematics
192	Statistics	Mathematics
193	Mathematics education	Mathematics
194	Actuarial science	Mathematics
195	Mathematical analysis	Mathematics
196	Applied mathematics	Mathematics
197	Topology	Mathematics
198	Calculus	Mathematics
199	Mathematical optimization	Mathematics
200	Arithmetic	Mathematics
201	Geometry	Mathematics
202	Psychiatry	Medicine
203	Orthodontics	Medicine
204	Dentistry	Medicine
205	Medical emergency	Medicine
206	Emergency medicine	Medicine
207	Ophthalmology	Medicine
208	Optometry	Medicine
209	Endocrinology	Medicine
210	Internal medicine	Medicine
211	Nursing	Medicine
212	Family medicine	Medicine
213	Intensive care medicine	Medicine
214	Radiology	Medicine
215	Nuclear medicine	Medicine

ID	2^{nd} level	1^{st} level
216	Physical therapy	Medicine
217	Physical medicine and rehabilitation	Medicine
218	Cancer research	Medicine
219	Oncology	Medicine
220	Medical education	Medicine
221	Gerontology	Medicine
222	Virology	Medicine
223	Immunology	Medicine
224	Pediatrics	Medicine
225	Veterinary medicine	Medicine
226	Pathology	Medicine
227	General surgery	Medicine
228	Surgery	Medicine
229	Nuclear magnetic resonance	Medicine
230	Genetics	Medicine
231	Cardiology	Medicine
232	Anesthesia	Medicine
233	Obstetrics	Medicine
234	Gynecology	Medicine
235	Neuroscience	Medicine
236	Gastroenterology	Medicine
237	Traditional medicine	Medicine
238	Physiology	Medicine
239	Audiology	Medicine
240	Urology	Medicine
241	Andrology	Medicine
242	Dermatology	Medicine
243	Anatomy	Medicine
244	Theology	Philosophy
245	Aesthetics	Philosophy
246	Engineering ethics	Philosophy
247	Epistemology	Philosophy
248	Environmental ethics	Philosophy
249	Astronomy	Physics
250	Astrophysics	Physics
251	Molecular physics	Physics
252	Chemical physics	Physics
253	Quantum electrodynamics	Physics
254	Quantum mechanics	Physics
255	Seismology	Physics
256	Geophysics	Physics
257	Particle physics	Physics
258	Nuclear physics	Physics
259	Atomic physics	Physics
260	Classical mechanics	Physics
261	Mathematical physics	Physics
262	Theoretical physics	Physics
263	Condensed matter physics	Physics
264	Optics	Physics
265	Biophysics	Physics
266	Computational physics	Physics
267	Statistical physics	Physics
268	Thermodynamics	Physics
269	Medical physics	Physics
270	Engineering physics	Physics
271	Acoustics	Physics
272	Law	Political science

ID	2^{nd} level	1^{st} level
273	Public administration	Political science
274	Clinical psychology	Psychology
275	Psychotherapist	Psychology
276	Social psychology	Psychology
277	Developmental psychology	Psychology
278	Pedagogy	Psychology
279	Cognitive psychology	Psychology
280	Applied psychology	Psychology
281	Psychoanalysis	Psychology
282	Criminology	Psychology
283	Cognitive science	Psychology
284	Religious studies	Sociology
285	Social science	Sociology
286	Gender studies	Sociology
287	Socioeconomics	Sociology
288	Media studies	Sociology
289	Ethnology	Sociology
290	Anthropology	Sociology
291	Demography	Sociology

Table C1: Conversion table between the second level (292 sub-disciplines) and the first level (19 disciplines) of fields of study. The table also reports node IDs used in the knowledge space (Figure 3).