KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

Pathways to Novelty in Science and Entrepreneurship



Dissertation presented to obtain the degree of Doctor in Business Economics

by

Andrés Madariaga Espinoza

Number 943 2025

KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

Pathways to Novelty in Science and Entrepreneurship

Dissertation presented to obtain the degree of Doctor in Business Economics

by

Andrés Madariaga Espinoza

Number 943 2025

Daar de proefschriften in de reeks van de Faculteit Economie en Bedrijfswetenschappen het persoonlijk werk zijn van hun auteurs, zijn alleen deze laatsten daarvoor verantwoordelijk.

Since the dissertations in the series published by the Faculty of Economics and Business are the personal work of their authors, only the latter bear full responsibility.

Examination Committee

Supervisor

prof. Dr. Stijn Kelchtermans	KU Leuven

Co-Supervisor

prof. Dr. Cindy Lopes-Bento	KU Leuven

Doctoral committee

prof. Dr. Fabian Gaessler	Universitat Pompeu Fabra
prof. Dr. Aldo Geuna	Università di Torino
prof. Dr. Arvids A. Ziedonis	Boston University & KU Leuven

Chair

prof. Dr. Sandra Rousseau	KU Leuven

Acknowledgements

People often think that writing the acknowledgements is the easiest part of a dissertation. I have come to realize that it can be surprisingly challenging to put into words all the gratitude and good fortune one has accumulated over four very special years.

First, I want to thank my supervisor, Stijn Kelchtermans. Like many prospective students, when I first contacted him about a PhD position, I was not entirely convinced about making such a significant decision and only had a vague sense of wanting to engage full-time in research. His timely and precise responses to my initial concerns—a relentless and defining attribute as an advisor—suggested that this could be the right time and place. I am now certain that I have been very fortunate to have had his mentorship (and, as you know, I do not use this word lightly). His detailed and accurate feedback, awareness of the challenges associated with each stage of the doctoral process, and clever sense of humour make him an exemplary doctoral advisor. Stijn has been extremely generous in sharing his time, expertise, and resources, while providing ample autonomy for me to develop my own ideas. He has been flexible in accommodating requests and responsive to different concerns, including my poor attempts to self-sabotage our research when work in progress did not look promising or results were not as expected. Through additional collaborative tasks, such as supervising master's theses and dealing with university administration, I have also come to know him as someone who genuinely enjoys the academic vocation, with all its perks and perils, while remaining authentic and grounded. I particularly appreciate his support in allowing me to work remotely during several weeks when I

needed sunlight and connection back home, and for encouraging me to go on a research visit. I hold the highest esteem for Stijn's dedication and for the rare balance he maintains between seriousness and lightness in research. It would be hard for me to overstate how much I have learned from him and the difference he has made in this journey. I am deeply grateful.

I want to express my sincere gratitude to Cindy Lopes Bento, my co-supervisor. Cindy has always been enthusiastic, approachable, and supportive of my work. She has been generous in sharing her thoughts, making time in her packed schedule, and providing access to data, without which my projects would not have been possible. After all, a doctoral student with ideas but no data can hardly see the light at the end of the tunnel. I particularly value Cindy's practical sense in conducting research and her spot-on (and lightning-fast) eye for interpreting regression outputs. I suspect it will take me a few more years (and a few more papers) to master that skill. More importantly, I appreciate Cindy as a good listener and have come to know her as someone who genuinely cares about her students' development and well-being. This was especially important during the early stages of my PhD.

I am also indebted to Arvids A. Ziedonis. Involved from the very beginning, I have always found my interactions with Arvids to be insightful and enriching. His careful reading of numerous drafts, thought-provoking comments, and clarifying questions have greatly improved my work. In particular, I believe my writing skills and attention to detail have benefited immensely from his feedback. I am also grateful to Arvids for finding the time and energy to be available, often before sunrise in the U.S., XL cup of coffee in hand, ready to jump into our video calls and discuss research. In addition to being a constant source of support throughout my doctoral journey, I have had the pleasure of meeting Arvids at several international conferences. One key takeaway from our exchanges is his reminder that an essential part of academic life consists of "talking to people," a point that may seem trivial but that I have come to appreciate as quite profound, especially coming from such a seasoned scholar. I am certain we will continue

to cross paths and talk to (hopefully) interesting people.

Furthermore, I want to thank Fabian Gaessler and Aldo Geuna, both of whom kindly agreed to join the examination committee. Fabian was especially supportive and kind in his response to my out-of-the-blue email asking about the possibility of hosting me during my academic visit to Barcelona. Thanks to him, I not only escaped the harsh Belgian winter but also enjoyed the pleasant atmosphere at Pompeu Fabra, its numerous top-notch seminars, and the Michelin-worthy cantina on the Ciutadella campus (I'm exaggerating, but they do serve a pretty decent primer y segundo plato). I came to know Fabian not only as an insightful researcher, truly sharp in his comments, but also as someone empathetic and genuinely enjoyable to talk with, even beyond research matters. I am equally grateful to Aldo, whom I first met at a conference in Vienna, where he was assigned to discuss a very early version of my work. I admit I was (mildly) underprepared for my first-ever academic conference and shared the draft with him just a couple of nights before presentation day (do not try this at home!). Despite this, he was kind and encouraging, remarking that we were onto something "novel and interesting, but needing quite some work." He is not only a knowledgeable scholar but also someone who provides substantive comments while maintaining a constructive spirit. Perhaps it is also time to disclose that one-third of my family is named "Aldo," which gave me an extra sense of trustworthiness, if not randomness. Jokes aside, I am especially grateful that Aldo took the time during his sabbatical to serve on the committee.

This work was made possible by financial support from the KU Leuven Impuls Fund (grant no. IMP/20/002), which I gratefully acknowledge. Although I was not directly involved in preparing the proposal or securing the award, I initially sought to pursue the ideas developed there. My dissertation later deviated considerably from that initial path, which I now believe (and hope) was for the better (key outcomes still to emerge; I will keep you all posted). I thank KU Leuven for providing an excellent and privileged environment for PhD students, with access to the necessary resources and generous time to focus on research. Likewise, the MSI Department has offered a

stimulating academic home, where I have found peers and friends with whom to share our (mostly harmless) struggles over papers, coding, and econometrics. Without any particular expertise and often with a healthy degree of disagreement, our lunch discussions have ranged from global politics and the weather to exotic holiday plans, cooking, (academic) celebrities, sports, and many other "hot topics." I am grateful that we found time to enjoy and grow together. Because our department is such a dynamic place, naming people individually would inevitably be unfair, so I raise a cheer to all of you who feel you were part of these moments, especially those with whom I have shared offices, conferences, summer schools, walks to Arenberg, happy hours, overpriced dinners, cinema sessions, Stapleton's nights, and some memorable late wine evenings on the fourth floor.

Most of all, I am grateful to my parents and family. A Aldo y Bruna, quienes siguen siendo mis primeros grandes maestros. Gracias por enseñarme a confiar, a equilibrar trabajo con disfrute, por animarme a seguir mis intuiciones y forjar un camino propio, y sobre todo, por no desperdiciar ninguna oportunidad de hacerme sentir querido. To each of my siblings, Alduca, Yiyi, and Pepe, who are always attentive and accommodating to me, taking the time to update me on the latest gossip from home. I appreciate your constant care and the distinct perspectives you bring; you are truly complementary sources of inspiration. To Domingo and Matilde, who have reminded me that time passes not in vain and vigorously insisted that I should keep my initial promise of "only four years away for study;" your visit was unforgettable. A Bruferva y a la Mari, gracias por ser parte de cada hito, por regalonearme y por derrochar esfuerzo, valentía y comunidad. A mi familia extendida, especialmente Espinocitas, Hugo, Carmen, Tere, Laurita, Ernita, Sra. Elena, Malu, Titines; los siento a todos parte del viaje.

A mis amigos y amigas, nuevos y antiguos, a quienes considero familia elegida. Soy muy afortunado de disfrutar y compartir con uds. en lugares y contextos que son muy numerosos para mencionar. Gracias por recibirme siempre con brazos abiertos y retomar la amistad justo donde había quedado. De manera especial, agradecido con quienes hemos compartido cumpleaños

y fechas especiales viviendo como afuerinos. To my adoptive Leuven fam, sharing our free time and (un)common problems has been comforting and enriching. A José Adaros, quien ha sido un apoyo crucial antes y durante este proceso.

Finally, I would like to express my deepest gratitude to my partner Fran. She has been incredibly supportive, understanding, flexible, and available, despite geographic distance. This journey would not have been as adventurous, rewarding, and memorable without you. Gracias de corazón por tu complicidad. No creo que hubiera podido lograrlo sin tu tremenda compañía.

Thank you all sincerely,

Andrés Madariaga Espinoza
Leuven, October 2025

Contents

Examination Committee **i**

Acknowledgements **iii**

List of Tables **xi**

1 General Introduction **1**

2 Deviations from Grant Proposals and Scientific Novelty **9**

2.1 Introduction 10

2.2 Theoretical Background 12

2.3 Methodology 15

2.4 Results 31

2.4.1 Antecedents of Deviation 31

2.4.2 Consequences of Deviation 36

2.5 Conclusion 39

2.6 Tables and Figures 42

3 Interdisciplinary Advisors and PhD Research **59**

3.1 Introduction 60

3.2 Institutional Context 64

3.3 Data 67

3.4 Empirical Strategy 71

3.5 Results 77

3.5.1 Novelty 77

3.5.2 Productivity 79

3.5.3 Additional Analyses 80
3.5.4 Exploring Mechanisms 83
3.6 Discussion and Conclusion 87
3.7 Tables and Figures 90

4 Framing Novel Entrepreneurial Ideas 109
4.1 Introduction 110
4.2 Theory and Hypotheses 113
4.3 Methodology 119
4.4 Results 127
4.5 Conclusion 133
4.6 Tables and Figures 138

General Conclusions 151

Appendices 159
A Appendix Chapter 2 160
B Appendix Chapter 3 191
C Appendix Chapter 4 211

List of Tables

2.1 Summary statistics, publication-grant pairs (N=1,341) 42
2.2 Summary statistics, grants (N=460) 44
2.3 The antecedents of deviation and grant openness (OLS) . . . 45
2.4 The consequences of deviation on scientific impact and novelty 46
2.5 The consequences of extensive deviation on scientific impact and novelty (CFA) – Panel A 47
2.6 The consequences of extensive deviation on scientific impact and novelty (CFA) – Panel B 48

3.1 Share of Students with ID Advisor by Faculty (student-level) 91
3.2 Descriptive statistics by Advisor Type (student-level) 91
3.3 Summary statistics (student,year-level) 92
3.4 Sorting into Advisors' Interdisciplinary Engagement 93
3.5 Covariate Balance Before Weighting 94
3.6 Covariate Balance After Weighting 95
3.7 PhD Students' Research Novelty 96
3.8 PhD Students' Research Productivity 97
3.9 PhD Students' Novelty by Research Orientation 98
3.10 PhD Students' Productivity by Research Orientation 99
3.11 PhD Students' Research Collaboration 100
3.12 PhD Students' Interdisciplinarity 101

4.1 Summary Statistics 138
4.2 Differences in Means, Above- and Below-Median Framing . . 139
4.3 Analysis of Rejection Probability 140
4.4 Analysis of Rejection Probability, Split by Evaluator Congestion 141

Chapter 1

General Introduction

Economic change is propelled by novel ideas (Schumpeter, 1951; Mokyr, 2002). From the polymerase chain reaction, which transformed DNA analysis and enabled us to understand diseases and develop new treatments, to the PageRank algorithm, which reshaped how we search and access information at scale, discoveries now commonplace in our lives were once little more than ideas on paper.[1] When such ideas move from concept to deployment, they do more than solve technical problems: they can reconfigure production, create new markets, and raise our living standards. For novel ideas to follow this trajectory is far from inevitable, however. It depends on choices by researchers and entrepreneurs, and on organizations and policies that enable the recognition of, and support for, novelty.

Recent evidence suggests that generating ideas that lead to breakthrough innovations has become more difficult over time (Bloom et al., 2020; Park et al., 2023). While the reasons behind this trend are multiple and contested (e.g., Gordon, 2016), there is more consensus regarding the implications. Namely, that our ability to sustain economic welfare and confront contemporary social and environmental challenges may be threatened (Schot and Steinmueller, 2018). This dissertation is therefore concerned with the following overarching question: What pathways allow novel ideas to flourish in

[1] Anecdotal accounts trace the conception of the polymerase chain reaction to a camping trip (Mullis, 1990), and PageRank was originally a doctoral project (Page and Brin, 1998).

science and entrepreneurship?

I study these pathways in two closely connected domains. The first is science. Scientific research is inherently uncertain (Arrow, 1962; Nelson, 1959), and is increasingly pursued through research collaboration (Jones, 2009; Wuchty et al., 2007), often involving individuals with distinct expertise who work together in the production of knowledge. The second is technology-based entrepreneurship. Typically, ideas with potential commercial application need to acquire resources to develop into products and services (Gans and Stern, 2003). These domains are distinct but strongly connected. Through scientific research, new knowledge is produced and new researchers are trained, both critical inputs for corporate R&D efforts and high-growth entrepreneurship (Azoulay et al., 2019; Stephan, 2015). Entrepreneurial action shapes what and how ideas develop into products and services that reach markets (Audretsch et al., 2006). Together, these two domains span the journey from the generation to the deployment of novel ideas. The dissertation's three chapters examine different points along this journey. Namely, Chapters 2 and 3 are situated within scientific research and the training of new scientists. Chapter 4 focuses on the evaluation of entrepreneurial ideas. The remainder of the introduction provides an overview of the three chapters.

In Chapter 2, we build on the notion that funders seek to reduce the uncertainty of scientific research by requiring detailed plans and expected results, yet they rarely verify whether published work closely adheres to the ideas outlined in the grant proposal (Azoulay and Li, 2020). While prior studies show that targeted funding opportunities can redirect research efforts (e.g., Myers, 2020), less is known about how research direction changes after the grant award. Recent evidence suggests that projects aimed at specific areas often yield findings in unexpected domains (Aslan et al., 2024), but it remains unclear what drives such deviations and whether deviation is associated with higher or lower scientific impact. The chapter addresses these gaps by quantifying *deviation* in terms of the semantic distance between the abstract of a grant proposal and the abstracts of subsequent publications

that acknowledge the grant. Specifically, we employ natural language processing techniques on over 1,300 publications resulting from 460 research grants awarded to scientists from a large European research university between 2000-2018, to document deviations between grants and the resulting publications. Our empirical strategy unfolds in two steps. First, we examine antecedents of deviation with a focus on new team members — those who join after the grant award — as a conduit for deviation. Second, we relate deviation to scientific impact using a control function approach (Wooldridge, 2015). Three findings stand out from our exercise. The addition of new team members after the grant award is associated with greater deviation from the proposal. The association is stronger when these newcomers are affiliated with external organizations or when their prior work is more distant from the content of the proposal. By contrast, first-time collaborations with those newcomers are associated with relatively less deviation, which is consistent with coordination challenges that constrain the introduction of new directions. In the second step, deviation is positively associated with novelty and scientific impact, which is consistent with a role for knowledge recombination in creative outcomes (e.g., Simonton, 1999). Our key contribution is to identify conditions under which deviation is positively associated with scientific impact and novelty, and derive implications for the grant funding system.

Chapter 3 focuses on academic training and interdisciplinary research. We assess how advisors' engagement in interdisciplinary centers affects PhD research outcomes. Interdisciplinary research centers have proliferated in universities and are promoted as vehicles for cross-field collaboration and scientific creativity (Jacobs and Frickel, 2009). Prior work portrays interdisciplinarity as a mode of research that brings both benefits and costs at the individual level (Leahey, 2018; Fontana et al., 2022). Because centers primarily seek to cultivate relationships among faculty (Biancani et al., 2014), these initiatives may alter the environment in which doctoral training takes place. Ample evidence attests to the role advisors play in shaping doctoral training (e.g., Corsini et al., 2022); however, whether and how advisors who engage in interdisciplinary research influence the research outcomes of their PhD stu-

dents remains unclear. We study the establishment of two interdisciplinary research centers rooted in the biomedical sciences at a large research university. We examine advisors' engagement with these centers as a footprint of their commitment to interdisciplinarity. The data combine administrative records on advisors and PhD students with publication data for approximately 1,000 advisors and 3,000 graduates from 2007 to 2023. The analysis focuses on student output from the start of the PhD through two years after graduation, employs scientist–year panel data, and estimates within-advisor effects by comparing students whose advisor engaged with a center to students of the same advisor before engagement. Our findings reveal a trade-off in how advisors' interdisciplinary engagement shapes doctoral training. Students whose advisors participate in interdisciplinary research centers publish fewer papers during their PhD, indicating a productivity penalty, but show modest gains in scientific novelty and citation impact, which appear to be associated with collaborating with more distant coauthors. These patterns suggest that exposure to interdisciplinarity may stimulate creative recombination and broader intellectual reach, even as it slows short-term productivity. The effects are particularly pronounced in fundamental research settings, as opposed to clinical research, where integrating different disciplinary methods and theories can be especially challenging. This chapter contributes to research on doctoral training by showing that the challenges and benefits of interdisciplinary research cascade into the formative stages of academic careers through advisors. Moreover, we offer insights for science policy, tempering expectations that institutional initiatives to stimulate collaboration can seamlessly benefit doctoral research.

Chapter 4 turns to the evaluation of entrepreneurial ideas. We study how contextual factors shape support for novel ideas submitted to an Accelerator program. While identifying high-potential ideas is critical for investors and support programs (Gompers et al., 2020), evaluating novelty is cognitively demanding and inherently subjective (Quy and Huy, 2007). Prior work shows that how ideas are articulated, and specifically the linguistic choices (referred to as 'framing'), influence evaluation outcomes (Snihur et al., 2022).

However, less is known about how the evaluation context moderates the effectiveness of framing. The chapter integrates two contextual features. The first is contextual novelty, defined as the atypicality of a venture's idea relative to the ideas submitted in the same cohort (Cutolo and Ferriani, 2024). The second is congestion, defined as the degree to which submission volumes exceed levels predicted by industry or time trends. Congestion can reduce the cognitive bandwidth evaluators devote to each idea and can shift which cues they attend to (Piezunka and Dahlander, 2015). The empirical setting is a prominent European Accelerator program that screens thousands of early-stage technology-based ventures and funds ideas with the potential to create new markets or disrupt existing ones. We examine more than 70,000 funded and rejected proposals from 2014 to 2020. We quantify the framing emphases in these proposals using a vocabulary of promotional adjectives that signal feasibility and innovativeness, which prior studies have linked to the success of novel scientific ideas (Peng et al., 2024). Moreover, we measure contextual novelty using natural language processing to quantify semantic distance, and we measure congestion from the program's rolling submissions and cutoff dates (Ferguson and Carnabuci, 2017). Three results stand out from our analyses. Evaluators reward feasibility-oriented framings, and ventures benefit when they combine feasibility with innovativeness. The benefit of feasibility framing does not extend to contextually novel ideas on average. However, under high congestion evaluators are less likely to reject contextually novel ideas, which suggests that institutional mission can dominate familiarity bias when evaluators are under strain. The interaction between framing and context shows that feasibility framing reduces rejection for contextually novel ideas only when congestion is high. Our key contribution is to refine theory on strategic framing by showing how the evaluation context conditions the effectiveness of framing choices for novel ideas in a real-world setting.

References

Arrow, K. J. (1962): *Economic Welfare and the Allocation of Resources for Invention*, Princeton University Press.

Aslan, Y., O. Yaqub, B. N. Sampat, and D. Rotolo (2024): "Unexpectedness in medical research," *Research Policy*, 53, 105075.

Audretsch, D. B., M. C. Keilbach, and E. E. Lehmann (2006): *Entrepreneurship and Economic Growth*, vol. 9780195183511, Oxford University Press.

Azoulay, P. and D. Li (2020): "Scientific Grant Funding," *NBER Working Paper Series*.

Azoulay, P., J. S. G. Zivin, N. L. Danielle, and B. N. Sampat (2019): "Public R&D Investments and Private-sector Patenting: Evidence from NIH Funding Rules," *Review of Economic Studies*, 86, 117–152.

Biancani, S., D. A. McFarland, and L. Dahlander (2014): "The semiformal organization," *Organization Science*, 25, 1306–1324.

Bloom, N., C. I. Jones, J. V. Reenen, and M. Webb (2020): "Are Ideas Getting Harder to Find? †," *American Economic Review*, 110, 1104–1144.

Corsini, A., M. Pezzoni, and F. Visentin (2022): "What makes a productive Ph.D. student?" *Research Policy*, 51.

Cutolo, D. and S. Ferriani (2024): "Atypicality: Toward an Integrative Framework in Organizational and Market Settings," *Academy of Management Annals*, 18, 157–209.

Ferguson, J. P. and G. Carnabuci (2017): "Risky recombinations: Institutional gatekeeping in the innovation process," *Organization Science*, 28, 133–151.

Fontana, M., M. Iori, V. L. Sciabolazza, and D. Souza (2022): "The interdisciplinarity dilemma: Public versus private interests," *Research Policy*, 51, 104553.

Gans, J. S. and S. Stern (2003): "The product market and the market for "ideas": commercialization strategies for technology entrepreneurs," *Research Policy*, 32, 333–350.

Gompers, P. A., W. Gornall, S. N. Kaplan, and I. A. Strebulaev (2020): "How do venture capitalists make decisions?" *Journal of Financial Economics*, 135, 169–190.

Gordon, R. J. (2016): *The Rise and Fall of American Growth*, Princeton University Press.

Jacobs, J. A. and S. Frickel (2009): "Interdisciplinarity: A Critical Assessment," *Annual Review of Sociology*, 35, 43–65.

Jones, B. F. (2009): "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?" *Review of Economic Studies*, 76, 283–317.

Leahey, E. (2018): "The Perks and Perils of Interdisciplinary Research," *European Review*, 26, S55–S67.

Mokyr, J. (2002): *The Gifts of Athena: Historical Origins of the Knowledge Economy*, Princeton University Press.

Mullis, K. B. (1990): "The Unusual Origin of the Polymerase Chain Reaction," *Scientific American*.

Myers, K. (2020): "The Elasticity of Science †," *American Economic Journal: Applied Economics*, 103–134.

Nelson, R. R. (1959): "The Simple Economics of Basic Scientific Research," *Journal of Political Economy*, 67, 297–306.

Page, L. and S. Brin (1998): "The anatomy of a large-scale hypertextual Web search engine," *Computer Networks and ISDN Systems*, 30, 107–117.

Park, M., E. Leahey, and R. J. Funk (2023): "Papers and patents are becoming less disruptive over time," *Nature*, 613, 138–144.

Peng, H., H. S. Qiu, H. B. Fosse, and B. Uzzi (2024): "Promotional language and the adoption of innovative ideas in science," *Proceedings of the National Academy of Sciences of the United States of America*, 121, e2320066121.

Piezunka, H. and L. Dahlander (2015): "Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing," *Academy of Management Journal*, 58, 856–880.

Quy, C. Z. and N. Huy (2007): "How Entrepreneurs Use Symbolic Management to Acquire Resources," *Administrative Science Quarterly*, 52, 70–105.

Schot, J. and W. E. Steinmueller (2018): "Three frames for innovation policy: R&D, systems of innovation and transformative change," *Research Policy*, 47, 1554–1567.

Schumpeter, J. A. (1951): *The Theory of Economic Development: An inquiry into profits, capital, credit, interest, and the business cycle*, Harvard University Press.

Simonton, D. K. (1999): *Origins of Genius: Darwinian Perspectives on Creativity*, Oxford University Press.

Snihur, Y., L. D. W. Thomas, R. Garud, and N. Phillips (2022): "Entrepreneurial Framing: A Literature Review and Future Research Directions," *Entrepreneurship Theory and Practice*, 46, 578–606.

Stephan, P. E. (2015): *How economics shapes science*, Harvard University Press.

Wooldridge, J. M. (2015): "Control Function Methods in Applied Econometrics," *The Journal of Human Resources*, 50, 420–445.

Wuchty, S., B. F. Jones, and B. Uzzi (2007): "The increasing dominance of Teams in Production of Knowledge," *Science*, 316, 1033–1036.

Chapter 2

Deviations from Grant Proposals and Scientific Novelty[1]

[1]Joint work with Stijn Kelchtermans, Cindy Lopes-Bento, and Arvids A. Ziedonis. We are grateful to Pascal Dengis for facilitating access to the Flanders Research Information Space. We also gratefully acknowledge helpful comments from participants at the OIS conference in Vienna, Eu-SPRI in Brighton, WOEPS in Torino, TPRI's Works-in-progress seminar, DRUID24, and the AOM Meeting 2024. A preliminary short version of this work was published online as: Madariaga Espinoza, A., Kelchtermans, S., Lopes-Bento, C., & Ziedonis, A. A. (2024). *Do funded research projects deviate from grant proposals, and does it matter?* Academy of Management Proceedings, 2024(1). https://doi.org/10.5465/AMPROC.2024.16873abstract

2.1 Introduction

It is widely acknowledged that uncertainty is inherent in the conduct of scientific research (Arrow, 1962; Nelson, 1959). Many funding agencies seek to reduce such uncertainty in grant-funded research projects by requiring a detailed explanation of the intended research and possible results (Rosenman, 1988; Azoulay and Li, 2020). Although funders typically demand ex post reporting on project results, they often do not require, nor do they verify, that results align closely with the intentions outlined in the funded proposal (Li, 2017; Myers, 2020).

The prior literature has examined how targeted funding opportunities (Myers, 2020) and funding restrictions on specific topics (Blomfield and Vakili, 2023), may steer research efforts into new directions. Changes in research direction occurring after the grant award — and the role of the grant proposal as an explicit research plan — have received less attention, however, despite recent evidence that funded research projects aimed at specific areas (e.g., diseases or conditions) often yield findings in unexpected areas (Aslan et al., 2024; Yaqub et al., 2024). Moreover, surprisingly little is known about the factors underpinning these deviations, and it remains unclear whether research that deviates from its initial plan produces scientific outcomes with more or less impact.

In this paper, we seek to shed light on these issues by identifying the prevalence of deviations of funded research from the initial proposal. We further examine the role that the addition of members who join the research team after funding has been awarded (new team members) may play in redirecting the project's research effort. While evidence suggests that grant applicants often benefit from team members' contributions and become more productive over time (Ayoubi et al., 2017; Davies et al., 2022), the specific influence of new team members on the research direction has been largely overlooked. Moreover, it is unclear whether deviation from the original, extensively prepared plan results in more impactful science.[2]

[2]In this study we do not consider why a researcher may join or leave a research team, rather, we investigate whether such changes in team composition are correlated with de-

To address these questions, we process publication and grant abstract data using natural language processing techniques and calculate a continuous measure of deviation[3] between the abstract of the grant proposal and the abstracts of subsequent publications that acknowledge the grant. Our unit of analysis is a grant-publication pair and our empirical approach consists of two stages. First, we examine the antecedents of deviation, with a focus on the role of new team members. In the second stage, we assess scientific impact as a function of deviation, using a control function approach (Wooldridge, 2015).

Our findings suggest that the addition of new team members post-grant is positively associated with the extent of deviation from the original proposal. This relationship is amplified when new team members are affiliated with external organizations or when prior research by those who join is more distant to the content of the grant proposal. Conversely, research projects involving first-time collaborations with new team members exhibit relatively less deviation. The latter association suggests that coordination challenges may constrain the team's ability to introduce new ideas that would cause it to deviate from the initial research plan. We do not find evidence that new team members from an external organization or conducting research that is more distant from the grant proposal further increases the coordination costs associated with first-time collaborators, however.

In the second part of our analysis, we find a positive association between deviation and scientific impact, suggesting that deviation from grant proposals, when spurred by knowledge recombination, is associated with increased scientific creativity (Simonton, 1999, 2021). This finding is robust to alternative measures of deviation, to variations of our main sample, and to alternative operationalization of key variables. Overall, by examining both antecedents and consequences of research outcomes deviating from their grant proposals, our analyses contribute to our understanding of how changes in team composition affect the direction of research. As we discuss in the con-

viation from the original grant proposal.

[3]Throughout the text, we italicize *deviation* whenever referring to our variable, rather than the general concept.

cluding section, our findings have important implications for the review of applications for grant-funded research and monitoring practices in science funding policy.

2.2 Theoretical Background

Why Research Outcomes May Deviate from Proposed Research

Numerous factors may cause executed research to differ or deviate from that outlined in funding proposals. First, scientists may explore alternative research directions when they fail to produce valuable findings (Nelson, 1959). Second, scientists may be eclipsed in the pursuit of a discovery by competing teams (Dasgupta and David, 1994), causing them to reorient their research. Third, the direction of a research project may evolve as intermediate findings reveal unforeseen opportunities. Such serendipitous discoveries may steer research in directions other than that originally planned (Merton and Barber, 2011; Yaqub, 2018). If the original team encounters roadblocks or identifies alternative research directions, it may recruit new members to overcome these challenges.

The composition of scientific teams can be fluid as members join or depart during the execution of research projects (Wang and Hicks, 2015; Jaravel et al., 2018). Previous work on knowledge recombination (e.g., Fleming and Sorenson, 2004) has found that team composition is crucial for fostering creativity through "bridging connections" that provide access to new information (Fleming and Marx, 2006). As new team members introduce diverse practices, unique resources, and novel perspectives, they can inspire fresh thinking by the original team members (Uzzi and Spiro, 2005). The resulting approaches may lead to deviation from the initially proposed research.

Scientists routinely acquire external knowledge through collaborations with new colleagues (Singh and Agrawal, 2011; Slavova et al., 2016), and external collaborators are frequently sought for skills and resources unavailable within an organization (Corley et al., 2006; Heinze and Kuhlmann, 2008). We thus expect the addition of external team members to positively corre-

late with deviation from the research proposal. Moreover, as bridging disparate knowledge sources is often linked to enhanced creativity (Audia and Goncalo, 2007; Kaplan and Vakili, 2015), we anticipate that adding team members pursuing more distant prior research increases the likelihood of introducing ideas or approaches that differ from those in the research proposal.

Cognitive processes may impede the assimilation of new ideas, however (Bercovitz and Feldman, 2011; Onal Vural et al., 2013; Reagans et al., 2016). Novel ideas are often met with rejection, especially when they arise from individuals whose expertise is distant from the team members' core knowledge base (Mueller et al., 2012; Rietzschel et al., 2010). Institutional frictions may also impede communication and consensus-building processes when individuals from different organizations unite (Cummings and Haas, 2012; Cummings and Kiesler, 2007), thus limiting the extent to which the performed research deviates from its original plan.

Moreover, unlike internal new team members who may have existing relationships and share an institutional background, new team members from an external organization may face greater challenges in adapting to the approach in an ongoing research project, potentially limiting their influence on the direction of the project and consequently reducing deviation from the grant proposal. Such would be the case in the "not-invented-here syndrome," where researchers resist external ideas, even when evidence suggests the external ideas are valuable and/or superior to those of existing team members (Antons and Piller, 2014; Peck et al., 2015). This resistance underscores the challenge of introducing outsiders to revise existing ideas (Hussinger and Wastyn, 2016; Lichtenthaler and Ernst, 2006).

Scientists may face unforeseen challenges when executing the plan, projects may encounter unexpected obstacles that were not anticipated at the time of initial grant proposal, such as lack of access to critical research materials (Walsh et al., 2007; Thompson et al., 2018; Tham, 2023) or adverse events (floods, fires, etc.) damaging specialized equipment or infrastructure (Baruffaldi and Gaessler, forthcoming). In such cases, deviation may occur as an adaptive response to unanticipated problems rather than a result of creative

thinking. Such a reorientation, while not a result of scientific creativity, may still affect the overall direction and success of the research project. Finally, funding agency priorities may also influence researchers' adherence to the original research plan. Funders may stimulate scientific exploration by rewarding scientists' accomplishments rather than the merits of a project proposal, or by maintaining flexible monitoring practices regarding deliverables (Azoulay et al., 2011). Funders may also seek to advance research on specific objectives (e.g., a particular disease or condition), such as NIH Requests for Applications examined by Myers (2020).[4] When grants are awarded based on a specific set of objectives, researchers may be more constrained in changing research direction.

Deviation and Scientific Impact

Researcher creativity is commonly viewed as central to the production of valuable scientific knowledge (Aghion et al., 2008; Stephan, 1996). For funded research, the merit of scientists' creative work is often assessed through the peer review of proposals, a crucial step in securing funding (Azoulay and Li, 2020). Whether deviation from the original, extensively prepared, and scrutinized grant proposal is associated with greater or lower scientific impact remains unclear, however.

On one hand, deviation that reflects scientific creativity, often resulting from the recombination of existing knowledge (Simonton, 2003; Weisberg, 1998), may receive greater peer recognition due to greater novelty of resulting research (Chai and Menon, 2019). On the other hand, deviation that departs significantly from a scientist's area of core expertise and prior line of research may lead to less scientific impact (Hill et al., 2025; De Rassenfosse et al., 2022). Furthermore, adverse events results in the loss of specialized equipment may encourage scientists to explore new research topics that may generate lesser scientific impact (Baruffaldi and Gaessler, forthcoming).

[4]Myers (2020) studies the role of thematic grants (as opposed to "open grants," in which the objectives are initiated by researchers) in steering researchers toward new research topics. In contrast to our paper, he does not explicitly examine the alignment between the content of the submitted proposal and the research conducted after the grant is awarded.

Thus, while research reorientation may generate more consequential breakthroughs, it may also result in decreased recognition and scientific impact.

2.3 Methodology

Research Context and Data

We examine funded research projects conducted by academic scientists at KU Leuven (KUL) in Belgium.[5] We combine several sources into a unique dataset containing publication – research grant pairs to identify teams of awardees or grant-holders (scientists who were jointly awarded project funding) and identify additional researchers included as coauthors in the publications (post-grant award coauthors).

This setting is conducive for empirical research on scientific deviation from grant proposals for two reasons. First, project-based funding is a cornerstone of research budget allocation across all academic disciplines at KUL. Scientists routinely submit grant proposals to obtain necessary funding for equipment and to hire technicians, Ph.D. students, and postdoctoral researchers. Second, all funded projects are centrally reported to the Flemish government by the university-wide research coordination office, thus allowing us to reliably identify and link researchers, funding records, and publication data.

We identify the organizational affiliation of each researcher at the time of publication and grant award, as well as past collaborations among grant holder(s) and coauthors joining after the grant award. Our data originates from three sources. From university records, we obtained personnel data across all scientific fields represented at the university (social sciences and

[5]KU Leuven is a research-intensive university with recruitment, promotion, and research policies conforming to international practices in the pursuit of research excellence and scientific impact. KUL is a founding member of the League of European Research Universities (LERU) and a member of international university associations including the Coimbra Group and UNA Europa, among others. The university ranked 42nd in the Times Higher Education (THE) ranking in 2022, is the highest-ranked institution in Belgium according to THE, and has been consistently ranked among the top 50 universities globally in recent years.

humanities, natural sciences and engineering, and biomedical and life sciences) in the period 2000-2018. Second, we obtained data on publications and grant funding from the Dimensions database of Digital Science.[6] Funding information in Dimensions is provided directly by the funding organizations, especially when funders are partners of Digital Science. Our third source is the official public funding records in Flanders, Belgium (the "FRIS data"), available from the year 2008 onwards. The FRIS database includes all funded research in our setting, where funding is distributed through competitive calls from various sources, including regional and federal sources in Belgium (e.g., the Research Foundation Flanders and the Belgian Science Policy Office), the European Union (e.g., the ERC), and other international sources such as funding agencies outside Belgium.[7]

Estimation Sample

We began by curating a list of 6,907 KUL researchers and retrieving publication records for each researcher since 1980. For each publication, we obtain metadata including the title, abstract, coauthors and their affiliation(s), publication date, disciplines (fields of research),[8] references, acknowledged funding information, and citation metrics as of December 2022. We obtain similar publication data for all coauthors of our initial set of KUL researchers.

From the publication corpus of 202,354 publications, 36,994 (18%) cite a supporting grant. For each grant, we obtain information such as the grant title, abstract (i.e., a summary of the grant proposal), funder details, grant start date and end date, researchers that were awarded the grant, and their af-

[6]Data on grants reported in the Dimensions database are provided directly by partner funding bodies and complemented by public sources, covering more than 600 funders worldwide including governments, nonprofit organizations, and international agencies. Dimensions further complements these data by processing grant codes mentioned in the acknowledgement section of publications, and then matched to the Dimensions grant database, in line with prior work in the literature (Grassano et al., 2017).

[7]A limitation of the FRIS data is that linkages to funded project's output are incomplete. We use the FRIS data primarily to complement information on grants retrieved from Dimensions (e.g., to validate the original team of grant holders and fill in missing values on grant abstracts, start and end dates, etc.).

[8]Fields of Research provided by Dimensions follow the Australian and New Zealand Standard Research Classification (ANZSRC 2020), composed of 22 disciplines.

filiation(s) from the Dimensions database. Also from Dimensions, we capture all grants awarded to our curated list of researchers and their co-awardees. We match these records with data from the FRIS database based on grant title, funder details, and researcher names. Of the grant-supported publications, 9,183 publications acknowledge grants awarded to (at least one) KUL researcher(s) in the years 2000-2018.[9] After initial cleaning of the data, we collect full information for 1,107 grants linked to 6,196 publications, representing 67% of the grant-supported publications in our study period.

We apply three restrictions to this sample of publication and grant pairs to operationalize our analysis. First, to ensure a clear point of comparison of the deviation of a given publication from the initial grant proposal, we include only publications supported by a single grant.[10] Previous studies indicate that the coexistence of multiple funding sources creates synergies and tensions between different project objectives that may alter the research process, and that are difficult to disentangle in the resulting publication output (Rotolo et al., 2023). Moreover, multiple funding sources can influence both the sources of new ideas and the scientific impact of publications (Hottenrott and Lawson, 2014), thus may confound the effect of changes in team composition that we aim to uncover.

Second, as our focus lies on research that is supported by grant awards, we exclude publications published within a year of the grant's start date. Instances where the grant proposal describes research that has largely been performed may blur the temporal separation between the research plan outlined in the grant proposal and its resulting publications. In turn, this could obscure the "true" deviation between the proposed research and its publication output, potentially leading to an artificially low deviation rather than

[9]The remaining publication-grant pairs are deemed out of scope since they have been awarded on a date before our observation period (2000-2018) or are linked to a group of scientists that is entirely external to KU Leuven.

[10]See Appendix A.1 for supplementary data description. In supplementary analyses, we extend our sample and include publications supported by multiple grants (Appendix A.3, Table A.3.3). As expected, we find a positive and statistically significant coefficient for the number of supporting grants mentioned by a given publication. We also find supporting evidence on the positive correlation between the presence of new team members and deviation from grant proposals.

describing new research.[11]

Third, to maintain empirical and theoretical tractability, we restrict our analysis to teams of fifteen or fewer scientists.[12] This restriction fits with our knowledge recombination logic that applies to small to medium-sized teams where individual contributions are more significant (Wu et al., 2019). Our choice of this team size threshold is consistent with prior research, particularly in studies investigating knowledge recombination processes within scientific teams (e.g., Haeussler and Sauermann, 2020).

After applying these restrictions, our estimation sample includes 460 grants awarded to 976 researchers, linked to 1,341 publications authored by 4,193 researchers. While our primary analyses are based on this sample of grant-publication pairs, we conduct various robustness tests using different sample variations and utilize the full corpus of publications and grants to calculate individual-level measures (e.g., researchers' productivity prior to the focal grant). Additional details and summary statistics are provided in the variables section.

Empirical Strategy

We employ a two-stage estimation procedure in our analysis. We first examine the relationship between the presence of new team members (antecedents of deviation) and the extent of deviation, disentangling two mechanisms: knowledge recombination and coordination challenges. In the second step, we examine the relationship between deviation and scientific impact employing the control function approach (CFA) outlined by Wooldridge (2015). This two-step procedure allows us to address omitted variable bias that may render deviation endogenous to scientific impact.

Our intuition behind the CFA approach is that unobserved research difficulties may cause OLS estimates of deviation on scientific impact to exhibit

[11] We further substantiate our chosen threshold in the Appendix A.1. In a supplementary analysis, we modify the one-year threshold and only exclude publications appearing within two months from the grant start date. Our findings are largely insensitive to this adjustment (Appendix A.5, Table A.5.6).

[12] Seventy-six observations, or 3.5% of the remaining publications, exceed our team size threshold of fifteen coauthors.

downward bias under two conditions: (i) research difficulties negatively affect scientific impact, and (ii) there is a positive correlation between deviation and research difficulties. Consistent with our theory, we anticipate projects encountering greater difficulty to deviate more and have lower scientific impact, all else being equal.

To implement the CFA approach, we estimate the determinants of deviation in the first stage equation including an "instrumental proxy" (Blundell et al., 2013) that purges the (unobserved) dependence between deviation and scientific impact. In doing so, we exploit the linguistic characteristics of the grant proposal to measure the research project's initial potential to change direction (i.e. the grant openness) and use these characteristics as an exclusion restriction. A key assumption in our econometric framework is that the linguistic characteristics of the grant proposal do not systematically correlate with unobserved research difficulties that may affect scientific impact. While a direct econometric test of this assumption is infeasible, our first-stage results support the notion that the openness of the grant proposal strongly explains the variation in our endogenous explanatory variable, *deviation*.

In the first stage, we capitalize on the chronological order of events: a team of applicants initially prepares the grant proposal and the award decision is subsequently made. Findings from the conducted research are then published, and resulting publications may involve coauthors that were not originally members of the team that submitted the grant.

While our study focuses on funded research,[13] unobserved quality differences in scientific teams and funding characteristics may confound our assessment of the antecedents of deviation. Regarding quality differences among scientific teams, it is well documented that researchers tend to select team members with similar levels of scientific quality (Banal-Estañol et al., 2018; Rivera et al., 2010), which may be positively associated with introduc-

[13] All funded projects in our study passed rigorous scrutiny by grant review panels thus demonstrating feasibility and scientific rigor in the proposal. To illustrate the level of scrutiny in our research context, for the main funding source in our sample (the FWO "research project"-grants) show an average success rate of 23.4% over the period 2010-2018, comparable to that of the NIH R01-Equivalent grants during the same years (19.4%). (Figures collected from the FWO annual reports and NIH Data Book).

ing valuable new research ideas. We mitigate the concern that unmeasured quality differences among research teams may explain deviation from a grant proposal, rather than the mere addition of new members to the team, by controlling for the research performance of both grant holders and new team members using their publication and citation rates prior to the grant award.

Funding characteristics, such as monitoring practices or grant requirements, may also affect the potential for deviation. On one hand, some funding sources may be unforgiving of failures (Azoulay et al., 2011), leading researchers to acknowledge grants in publications that only have a tenuous relation with the funded project, to justify having received the grant. Such publications would show an "artificially high" degree of deviation since they are not rooted in the grant proposal to begin with.[14] On the other hand, some funders may "reject" deliverables that deviate too far from the original proposal (Price Nicholson, 2018), thereby limiting the extent of deviation.[15]

Throughout our regression analyses, we include funding source fixed effects to account for the fact that the freedom to deviate from the grant proposal may depend on the funding source. In supplementary analyses we employ grant fixed effects to further account for unobservable characteristics

[14]We do not find evidence for such "over-acknowledgment behavior." For example, researchers who hold a single grant at a given time may be inclined to acknowledge that grant in all their publications, potentially including publications that are unrelated to the funded project, thereby increasing the average deviation between publications and the grant for such researchers. Conversely, researchers holding multiple grants simultaneously may be better able to acknowledge the most relevant grant for each publication, resulting in a lower average deviation between publications and grants for such researchers, all else equal. However, we do not find such systematic differences in deviation between researchers holding single versus multiple grants. To further mitigate concerns regarding grant-reporting behavior, we control for grant productivity (the number of publications that acknowledge the same grant) and include a dummy variable for researchers with multiple ongoing grants beyond the one acknowledged in a focal publication-grant pair in all regressions.

[15]An additional possibility is that funding sources may pose different requirements on applicants regarding additional activities (e.g., organizing workshops) and involvement with partners and networks outside the research team, which are not measured in our data. Some funding sources require building a consortium involving multiple partners. While the members of a consortium may not be mentioned as holders of the focal grant nor be involved in the resulting publications as coauthors, the consortium may act as a source of ideas (e.g., through joint workshops, engagement activities, etc.) and therefore act as a conduit for deviation.

at the grant level.[16] We explore the role of knowledge recombination by analyzing team size (*authors count*), the presence of new members in the team (*share of new team members*), the internal versus external affiliations of new team members (using dummy variables), and whether new team members' prior work is distant from the grant proposal (*distant new team members*). To assess coordination challenges, we examine whether grant holders and new team members collaborated for the first time (*first collaboration*) and introduce interaction terms between first collaboration and both the affiliations of new team members and distant new team members. In our control function approach, we also account for *grant openness*, which we measure using the presence of curiosity-driven expressions and the length of the grant abstract. Detailed descriptions of these variables are provided below.

Outcome Variables

Deviation. We apply natural language processing (NLP) techniques to construct a measure of the deviation between a grant proposal and an associated publication, based on the abstracts. In doing so, we use pre-trained language models to create vector representations of the abstracts and to measure the "semantic similarity" between each publication and its corresponding grant. We utilize the software package SciBERT, a variant of the BERT (Bidirectional Encoder Representations from Transformers) language model developed by Google (Devlin et al., 2019) and trained on the full text of publications from SemanticScholar.org (Beltagy et al., 2019). SciBERT is optimized to interpret the technical language and scientific terminology present in scholarly literatures, making it well-suited for tasks such as language understanding, text classification, and other NLP applications.[17]

[16] We report and discuss these results in Appendix A.4. While these alternative estimations highlight the influence of funding heterogeneity on the extent of deviation, they also support our main arguments, particularly the knowledge recombination mechanism and, to a lesser extent, the role of coordination challenges.

[17] Text preprocessing (e.g., lowercasing, removal of punctuation or special characters) is generally not required when using pretrained transformer-based language models such as SciBERT, since these models use internal tokenization methods that automatically handle casing, punctuation, and uncommon words consistently (Beltagy et al., 2019; Devlin et al., 2019). To verify this, we implemented additional standard preprocessing steps

We measure *deviation* as one minus the cosine similarity of the abstract of a grant and its resulting publication. Prior research has shown that abstracts provide a reliable representation of the main concepts, findings, and contributions contained in full texts (Hofstra et al., 2020). Cosine similarity is a distance measure of the angle between two vectors projected in a multi-dimensional space, irrespective of the size of the vectors. It has been extensively used to compare semantic similarity in a range of applications (e.g., Chandrasekaran and Mago, 2021). Deviation can range from 0 to 1 and equals 0 when the publication is semantically equivalent to the grant proposal. In our sample, mean *deviation* is 0.29, with a standard deviation of 0.07, and a range of 0.07 to 0.54. Economics and the Health Sciences show a relatively higher mean *deviation* of 0.32, suggesting a greater tendency to pivot from the grant proposal than in other disciplines. In contrast, fields such as Law and Legal Studies and Language and Communication Studies exhibit lower mean *deviation* of 0.22 and 0.23, respectively.[18]

To illustrate the ability of our measure to capture the extent of deviation from grant proposal, we present two examples of grant-publications pairs, together with their measure of *deviation* in the Appendix A.2, Table A.2.3. In the Example 1 we show a publication-grant pair with a relatively low *deviation* (0.07), i.e. the publication belongs to the bottom percentile of *deviation* and closely aligns with the goals outlined in the grant. The stated goal of the grant is to gain fundamental knowledge into the process and formulation parameters influencing electrospraying to improve drug delivery methods.[19] Both the grant and publication emphasize the challenges of low solubility in drugs, the use of amorphous solid dispersions, and the exploration of electrospraying for particle preparation. The publication identifies key parameters of this process such as flow rate, electric potential difference, distance between

(including lowercasing, tokenizing, removing special characters, and removing stop words, as in Arts et al., 2021); the correlation between the embeddings with and without "manual preprocessing" was 0.99, confirming that additional preprocessing is unnecessary.

[18]We report the distribution of deviation across our sample in the Appendix A.2, Figure A.2.1, and show summary statistics by Fields of Research in the Appendix A.2, Table A.2.2.

[19]Electrospraying is an electro-hydrodynamic process that uses an electrostatic force to spin or spray a polymer solution into ultrafine droplets (Rostamabadi et al., 2021).

the nozzle and collector, solvent type, and polymer concentration, that are critical for the design of electrospraying methods for amorphous solid dispersions. A casual reading suggests that the results reported in the publication largely mirror the plans described in the grant proposal.

In Example 2, we present a publication-grant pair characterized by a relatively high *deviation* (0.35) —above the 75th percentile— indicating a greater disparity between the grant proposal and the resulting publication. The grant application proposes developing new materials from rod-shaped cellulose and starch-based particulates by controlling their arrangement through chemical changes and studying how they interact with metal nanoparticles. The publication, however, focuses on how a specific protein (bovine serum albumin, BSA) interacts with different cellulose particles, looking at how individual amino acids behave on their surfaces. Both deal with cellulose modification and interactions, but the grant emphasizes creating materials and self-assembly, while the publication zooms in on protein–cellulose interactions. Importantly, the grant proposal did not anticipate the study of BSA, let alone any other animal proteins, indicating a notable deviation from the original research plan.

Novelty. To further investigate whether knowledge recombination drives deviation and the subsequent research, we explore the relationship between deviation and scientific novelty. While a citation-based measure of scientific impact may capture the role of new team members in disseminating research results, it mainly reflects how widely a piece of research is recognized and built upon by the scientific community rather than indicating how innovative the work is. A publication's novelty, on the other hand, pertains to the originality of the research, focusing on whether the work introduces new concepts, methodologies, or areas of inquiry. The degree of novelty of a paper is not affected by diffusion processes but rather is expected to be positively related to deviation driven by knowledge recombination, all else equal. Research can be highly impactful without introducing new concepts or disrupt prevailing paradigms, particularly in fields where incremental advancements are highly valued. Separating novelty from citation-based impact allows us to differen-

tiate between these two concepts and better understand the antecedents and consequences of deviation.

We measure the novelty of scientific research at the time of publication by adapting the text-based "new word" measure developed by Arts et al. (2025).[20] This measure identifies terms and concepts in papers that appear for the first time from 1880 to 2023 across the entire scientific corpus. Specifically, we define a novelty dummy variable that equals one if a given publication introduces at least one new word to the entire corpus of science, and this word has been used again in later publications at least once. This approach allows us to capture the novelty of research output at the time of publication and to assess whether deviation from the original grant proposal leads to the introduction of novel concepts into the scientific literature.

Scientific impact. In terms of scientific impact, our primary dependent variable is the *field citation ratio*, provided by the Dimensions database. Citation-based metrics, albeit imperfect, are widely used in bibliometric studies and the research assessment of scientific articles, individuals, and institutions (Moher et al., 2018). Indeed, credit to authors in the form of citations constitutes a key currency in science and is associated with funding and hiring decisions (Stephan, 2015).

The *field citation ratio* is calculated at the article-level by dividing the number of citations a paper has received by the average number of citations received by all publications in the same year and the same 4-digit Fields of Research category from the Dimensions database.[21] The values are centered at unity so that a publication with a value of two has received twice as many citations as average for an article that belongs to the same Field of Research category published in the same year. Unlike raw citation counts, the *field citation ratio* accounts for variations in citation practices across different

[20]We link our publication sample to the open repository of Arts et al. (2025) using the doi of publications, which relies on the OpenAlex database. We link and generate the novelty dummy for approximately 95% of our sample (1,277 out of 1,341 publications).

[21]The Fields of Research 4-digit classification is comprised of 171 disciplines. For an article to have a valid field citation ratio, it must meet the following criteria: (a) the article is at least two years old and published after the year 2000, and (b) a threshold of 500 articles must be categorized in the same field of research for the year in which the article was published.

disciplines and allows for a comparison of articles published in different years (Waltman, 2016).

In our estimation sample, the mean *field citation ratio* is 7.86, with a minimum value of 0 (no citations) and a maximum value of 112, while 45 publications have missing values (3.3%).[22] Our regression analyses log the *field citation ratio* to account for its highly skewed distribution. We complement this measure of scientific impact with a dummy variable for publications that fall in the top 10% or bottom 10% of the citation distribution within each Field of Research and year of publication to analyze whether larger *deviation* is associated with disproportionate scientific recognition.

Independent Variables

Author count. We anticipate that a larger team size, defined as the total number of researchers involved in a publication, will be associated with a greater deviation, as a higher number of team members would be expected to increase the likelihood of any one person introducing a new idea after the proposal's finalization.[23]

Share of new team members. In line with our knowledge recombination perspective, we expect the presence of new team members to positively affect the deviation of the publication from the grant proposal. As grants are often used to train junior scientists (e.g., Ph.D. students and/or Postdocs), these individuals may constitute a non-negligible proportion of the coauthors in a publication. In contrast to more senior scientists who typically enjoy substantial freedom to pursue their own research agendas, the role of junior

[22] In supplementary analyses of the consequences of deviations, we use raw citation count as of December 2022 (citations) as a dependent variable and find consistent results. These results are available in the Appendix A.5, Table A.5.4.

[23] In 4.2% of the observations in our analytic sample, we observe that the grant holders have no new team members (i.e., no additional coauthors joining after the grant). A test of means indicates there is a significantly lower degree of deviation when the team composition remains unchanged after the grant award. Moreover, within the cases where team composition remains unchanged, we observe a significantly larger deviation in publications with multiple coauthors compared to publications from solo authors. We report the corresponding mean, standard deviation, and t-statistics in the Appendix A.2, Table A.2.4 and Table A.2.5.

scientists in deviation is unclear ex ante. On the one hand, they may be expected to "execute the plan", i.e., stick close to the grant proposal. On the other hand, they may conduct research beyond the grant proposal boundaries in agreement with the principal investigator. To define the variable unambiguously, we use the share of new team members with an academic age of at least three years (since her/his first publication year), relative to the total number of new team members in the publication.[24]

Affiliations of new team members. Using authors' affiliation listings in the publication and grant data, we classify the affiliations of new team members into three categories: (i) *Internal*, where all the new team members are affiliated with KU Leuven; (ii) *Mixed*, where at least one new team member (but not all team members) is affiliated with an external institution (i.e., other than KU Leuven); and (iii) *External*, where all new team members are affiliated with external institutions. We expect external new team members to be more likely to introduce new ideas relative to the original grant proposal but also to induce higher coordination costs compared to internal new team members, as discussed in our theory section. For mixed teams, we anticipate intermediate coordination costs, higher than for *Internal* teams but lower than for *External* teams.

Alignment of new team members' prior work with grant proposal. We measure how closely new team members' prior work aligns with the research described in the grant proposal. We expect this alignment to

[24]Our chosen threshold of academic age (i.e., 3 years since first publication) is inspired by the early-career scientists definition adopted by the European Research Council. The rationale is that these scientists manage their own research funding and decide on their own research agendas, a point that is typically reached several years after completing postgraduate training (National Research Council, 2005). Fine-grained distinction between the precise role of team members is not possible: our data does not include any information on personnel, staff, and doctoral/postdoctoral researchers linked to a particular grant. Moreover, we have no information about the status of external researchers in their institutions. Consequently, we are unable to measure these roles unambiguously and rely on academic age calculated from publication data to identify junior researchers. In supplementary analyses of the antecedents of deviations, we also examine: (i) variations in the academic age threshold to two or four years, (ii) the inclusion of new team members both above and below an academic age of three years (i.e., including junior new team members), and (iii) the share of new team members relative to the total number of coauthors, rather than just new team members.

be correlated with the likelihood that new team members will introduce new ideas to the team. We extend our procedure for calculating *deviation* by computing the distance (inverse cosine similarity) between new team members' pre-grant publication abstracts and the focal grant proposal abstract, including all their publications in the three years preceding the grant.[25] We then calculate the average distance across all coauthors who joined the team as a co-author after the grant award for each publication-grant pair, hypothesizing that a larger distance should positively influence *deviation*. To aid interpretability, we identify whether the average distance between new team members' prior work and the grant proposal for a given publication-grant pair is substantial. Specifically, in our regression analysis, we include a dummy variable, *distant new team members*, defined as whether new team members are on average in the upper quartile of the distance distribution considering new team members across all grant-publication pairs in the sample.

First collaboration with new team members. We define a dummy variable that equals 1 when a publication entails at least one first-time collaboration between a grant holder and a researcher joining after the grant award. The first-time collaborator can be either from KU Leuven (internal) or outside KU Leuven (external).[26] In line with our theoretical argument, we expect first-time collaborators to have a negative influence on *deviation* to the extent that they create additional coordination costs for the team.[27]

Grant openness. Drawing on the literature that has developed linguistic indicators to analyze scientific writing (Millar et al., 2022; Van Den Besse-

[25]We also compute distance for new team members using a 5-year window, and our results remain robust, as the two variables show a very high pairwise correlation (0.98).

[26]In supplementary analyses (Appendix A.3, Table A.3.6), we utilize two alternative versions of the first collaboration dummy: (i) we replace the dummy for cases where all new team members are first-time collaborators with the grant holder(s) on the publication, rather than at least one; (ii) we exclude first-time-ever researchers from the calculation. Our rationale is that while the presence of a single first-time collaborator or a first-time-ever researcher may have a negligible influence on team dynamics, the most significant impact of first-time collaborations may occur when all new team members are first-time collaborators and/or when focusing on established researchers.

[27]The LIWC dictionary contains validated psychosocial constructs that have been recently shown to influence the success rates in grant applications (Van Den Besselaar and Mom, 2022). Specific curiosity-related terms in the list of 76 entries include 'wonder*,' 'explor*,' 'look for*,' etc., where '*' denotes a wildcard search term.

laar and Mom, 2022), we measure the degree of flexibility within the research proposal, as this may affect the project's potential to deviate from its initial plan. Our reasoning is that some proposals, motivated to advance fundamental knowledge, are open-ended in nature and have a broad scientific scope, thus allowing for exploratory approaches and a wider range of potential outcomes. In contrast, other proposals may be more narrowly focused, with a limited number of planned experiments that restrict the ability to pursue alternative research directions. We use two measures to capture grant-level heterogeneity in the potential to deviate: (i) expressions linked to curiosity in the grant proposal, which may signal higher propensity to reorient the research direction; and (ii) the word count in the grant abstract, since longer abstracts may signal more detailed and specific research plans that constrain the potential for deviation. To measure curiosity, we use the software package Linguistic Inquiry and Word Count (LIWC22) to analyze the grant abstract text. Our curiosity proxy corresponds to the percentage of words in the grant abstract that match one of 76 curated word entries reflecting the authors' search for —or interest in— new knowledge.

Control Variables

Researcher prior performance. The insights and discovery moments that drive scientific progress are often associated with scientists' creativity, intuition, and/or talent. More talented scientists are better able to generate more novel ideas and are more capable in bringing these novel approaches to success (Simonton, 1999). Prior studies have also shown that top research performance is highly persistent over time (Kelchtermans and Veugelers, 2013; Petersen et al., 2012). We proxy for research talent by the prior research performance, distinguishing between the performance of grant holders and new team members. We use the (coauthor-weighted) yearly average number of citations per publication received in the three years preceding the grant acknowledged by the focal publication (*GH citation rate*), and the (coauthor-weighted) yearly average number of publications in the three years preceding

the grant acknowledged by the focal publication (*GH publication rate*).[28] We further include a dummy variable to capture whether any of the grant holders have other active grants that are not acknowledged in the focal publication (*GH multiple funding*). Finally, we add a dummy variable for instances where none of the grant-holders are present in a focal publication, which occurs in about 9% of the publications in our sample (*GH absence*).

Time to publication. We define the variable time to publication as a continuous variable, calculated as the number of days above the (field-normalized) average time to publication from the grant start date, converted to years. A longer time to publication may be positively associated with deviations from grant proposals as it may reflect adjustments and adaptations of more complex research or potential roadblocks that were encountered in the research process (e.g., lack of access to research tools).

Funding source. Consistent with the intuition that funding characteristics may influence the ability of funded projects to deviate from the grant proposal, we include a set of dummies for the main funding agencies in our sample: Science Foundation Flanders (FWO), federal funding in Belgium (Belspo), the European Research Council (ERC), other funding sources within the EU, and other funding sources outside the EU.[29]

Grant productivity. We include the (field-normalized) number of resulting publications that acknowledge the grant as an additional ex-post measure of the scope of the grant, which may imply an increased probability of a publication deviating from the grant proposal.

Fields of Research. We control for the scientific domain of the publication based on the Fields of Research (FoR) in the Dimensions data, where each publication is assigned at least one of 22 FoR codes, following the ANZRC 2020 classification system. We add a category for cases where mul-

[28] Our results are robust to using a 5-year window of prior research performance instead of a 3-year window for both the publication and citation measures (not reported).

[29] While our data also distinguishes funding at the more fine-grained level program-level (within a funding agency), the distribution of grant funding sources closely aligns with the funding agency level in our estimation sample. For example, within the publications supported by grants awarded by the FWO, 90% (299 out of 329) correspond to the 'FWO Research projects' program.

tiple FoR codes are assigned. We follow the widely accepted view that the norms and organization of scientific research differ by domain (some domains are more prone to open-ended research, while experimental fields often expect grant proposals to present a detailed plan Nelson (2016)). As previously mentioned, descriptive statistics suggest that the prevalence of deviations from grant proposals varies across disciplines in our estimation sample.

Embeddedness in prior research and ***Year of publication.*** Bibliometric studies have shown that the scientific impact of a publication can be affected by characteristics such as the body of knowledge it builds on or its vintage (Vieira and Gomes, 2010). We account for these publication-level traits by including the field-normalized number of references in the bibliography and the year of publication.

We report the main descriptive statistics in Tables 2.1 and 2.2 and highlight a few key observations here. The summary statistics at the publication-grant level in Table 2.1 show that, on average, publications in our estimation sample list 1.2 grant holders and 3.9 researchers joining as co-authors after the grant award.[30] The average grant in our study is awarded to 2.5 scientists and produces 13.2 publications (Table 2.2). Further characteristics of grant-holders and new team members are included in Appendix A.1. On average, grant holders have had longer academic careers than new team members when they co-publish, as measured by the years since their first publication (18.1 vs. 11.3 years). They also demonstrate higher research productivity, with an average of 12.7 publications compared to 7.4 publications for new team members, and collaborate with more coauthors (17.5 vs. 8.9). Despite these differences, both groups exhibit a similar average yearly citation rate per publication in the three years preceding the grant award (7.6 vs. 7.0).

[30]These publications are distributed over a range of disciplines, including engineering and natural sciences (36%), social sciences and humanities (16%), clinical and biomedical sciences (24%), and multidisciplinary publications (24%). Approximately one-third of the grants are funded by the Research Foundation Flanders (FWO), 24% are funded by other funding agencies within the EU, 19% by funds outside the EU, 13% are funded by the ERC and 12% by the federal government in Belgium (Belspo). Due to space limitations, we omit summary statistics by field of research and funding source.

We report the pairwise correlations between the main variables used in the regression analyses in Table A.1.3.

2.4 Results

2.4.1 Antecedents of Deviation

We report our results pertaining to the antecedents of deviation in Table 2.3. In all models, we report OLS estimates using deviation as the dependent variable, with standard errors clustered at the grant level.[31] We begin with a baseline specification where deviations are estimated using only the control variables (Table 2.3, Model 1). Subsequently, we introduce the variables capturing the grant openness (Table 2.3, Model 2), and then progressively introduce the knowledge recombination variables (Table 2.3, Models 4-6) and the coordination challenge variables (Table 2.3, Models 6-9). We focus our discussion on these two mechanisms of interest, with brief interpretations of key control variables and supplementary analyses provided at the end of this section.

Knowledge Recombination

Consistent with our logic that new team members —coauthors joining after the grant award— facilitate knowledge recombination, potentially driving deviation between proposed and conducted research, we find evidence that a greater share of new team members in the team (*share of new team members*) is positively associated with *deviation* (Table 2.3, Model 4). Moreover, *external* new team members (i.e., those who are not affiliated to KU Leuven) (Table 2.3, Model 5) and those whose prior research is distant from the ideas in the grant proposal (*distant new team members*) further contribute to deviation (Table 2.3, Model 6).

The overall team size (*authors count*) shows a positive albeit weakly

[31] As a robustness check, we use fractional regression logit instead of OLS to account for the bounded nature of deviation (ranging from 0 to 1). Our findings remain robust to this alternative estimation method (Appendix A.3, Table A.3.2 and Table A.3.9).

statistically significant correlation (at the 10% level) with *deviation* (Table 2.3, Model 3), which is neutralized when *share of new team members* is added to the regression. This suggests that a simple increase in headcount of a team has a limited influence on deviating from the proposal, but team members joining after the grant award seem instrumental in the knowledge recombination process.[32]

The estimated coefficients for the share of *new team members* (Table 2.3, Models 5-9) remain positive but become weakly significant once we introduce the dummy variables capturing the affiliation of new team members. Interestingly, the *external* new team members exhibit a consistently positive and statistically significant influence on *deviation*, relative to the baseline category of internal new team members, while mixed teams do not systematically differ from the baseline (Table 2.3, Models 5-9). This suggests that teams that recruit new members exclusively from outside their organization tend to deviate more than teams whose new team members are drawn from the same organization as the grant-holders.

Finally, we also find a consistently positive and statistically significant coefficient for *distant new team members* (Table 2.3, Models 6-9). We find that the coefficient is twice that of *external* new team members (Table 2.3, Model 9). This estimate implies that when the grant-holders bring in new team members whose prior research ranks in the upper quartile of average distance from the grant proposal (within our sample), *deviation* increases by approximately 7% on average, all else equal. Taken together, these estimates show that external ideas and distance in idea space play distinct roles in explaining deviation. Finally, it is noteworthy that the positive coefficients associated with these knowledge recombination variables do not depend on the inclusion of the *grant openness* controls and remain robust to different operationalizations and alternative metrics of deviation from grant proposals.[33]

[32]We obtain similar results when substituting a log-transformation of authors count for the raw number of authors and squared authors count (not reported).

[33]We report supplementary analyses of the antecedents of deviation without controlling for the grant openness in the Appendix A.3, Tables A.3.1-A.3.2. Different operationalizations can be consulted in the Appendix A.3, Tables A.3.4-A.3.5. Our insights on the

Coordination Challenges

We investigate whether coordination challenges interfere with the knowledge recombination process. First, we expect *deviation* to be lower when new team members lack prior research collaborations with the grant holders. We further anticipate that the challenges that come with first-time collaborations play out more strongly for new team members, i.e., reduce their ability to steer the research away from the original grant proposal, if they are external or have more distant expertise. External team members may face greater coordination challenges due to their lack of experience with the research team and organizational context while new team members whose prior research is distant from the grant proposal may face difficulties in legitimizing their perspectives to the original team.

In Model 7 of Table 2.3, we introduce our *first collaboration* dummy, and in Models 8-9, we add two interactions: between *first collaboration* and new team members' affiliation (*internal* remains the baseline category), and between *first collaboration* and *distant new team members*.

As shown in Table 2.3, Model 7, we find a negative and significant coefficient for *first collaboration* indicating a decrease in deviations from the grant proposal. This suggests that when original team members —the grantholders— engage in collaborative work for the first time with new team member, they tend to conform more closely to the initially proposed research plan. Introducing this *first collaboration* dummy does not affect the estimated coefficients of the knowledge recombination variables, however.

While we anticipated that grant holders might be more inclined to reject ideas originating from external new team members, our findings suggest such coordination challenges have limited impact on the influence of new collaborators, regardless of their organizational background or distance from the grant ideas. The interactions between the *first collaboration* dummy and new team members' affiliations, or between *distant new team members* and *first collaboration* are not statistically significant (Table 2.3, Models 8-9). The effect of first collaboration disappears, however, when we include these in-

knowledge recombination mechanism remain robust.

teractions. We surmise that new team members may be added only if the grant-holders perceive that the benefit of embarking on a first-time collaboration exceeds the associated coordination costs.[34]

Grant Openness

We account for the heterogenous potential to change direction across research proposals (*grant openness*) by controlling for expressions of curiosity in the grant abstracts and the grant abstract length (in linear and quadratic form). Our results consistently identify *curiosity* as a positive and statistically significant factor influencing deviation (Table 2.3, Models 2-9). The positive coefficients for *curiosity* suggest that more exploratory and open-ended projects are more prone to deviate from the initial plan. This finding aligns with our expectations and is consistent with previous studies showing that the creative intent of research activities is a contextual factor that correlates with effective knowledge recombination (Franzoni et al., 2018).

In contrast, *grant abstract word count* exhibits a negative and statistically significant association with deviation. One interpretation of this finding is that a longer grant abstract is indicative of a more comprehensive and detailed research proposal, specifying a project with clearly laid out research lines, a more extensive description of planned experiments, reliance on specific methodologies, or the continuation of a previous project. The positive coefficient for *grant abstract word count* squared indicates a concave relationship, meaning deviations decrease at a decreasing rate as the length of the proposal increases.

Regarding the grant holder's prior research performance, we find that while *GH prior citation rate* does not appear to affect deviation, the coefficient for *GH prior publication rate* is consistently positive and statistically

[34]We report results of alternative operationalizations of the first collaboration dummy in Appendix A.3, Table A.3.6. These supplementary analyses suggest that the coordination costs associated with first collaborations are primarily driven by first-time-ever researchers. When first-time-ever researchers are excluded from the first collaboration dummy, the effect disappears (e.g., Table A.3.6, Model 4). These additional analyses confirm the limited evidence on heightened coordination costs induced by diverse organizational backgrounds or distance of new team members' prior research.

significant (Table 2.3, Models 1-9). This result suggests that researchers with more publications are more likely to deviate, regardless of the scientific impact or visibility of their publication output. This finding is consistent with prior research indicating that research productivity in terms of publication volume is positively associated with a higher likelihood of exploring new fields while research quality is not (Kelchtermans et al., 2021).

We also find a negative and statistically significant relationship between grant-holders who hold active grants in addition to the one acknowledged in the focal publication (*multiple funding*) and deviation (Table 2.3, Models 3-9). This finding implies that researchers with multiple ongoing projects are less likely to deviate from the original grant acknowledged in the focal publication. Although researchers with multiple funded grants may capitalize on opportunities to cross-fertilize ideas across projects, potentially resulting in deviation, previous studies have highlighted difficulties in effectively leveraging synergies between multiple grants, often due to time constraints when managing multiple funding streams (Hottenrott and Lawson, 2017). This result is also consistent with a previously documented association between greater funding success and lower probability of exploring new fields (Kelchtermans et al., 2021).

For publication-specific factors, we find that *embeddedness in prior research* has a negative and statistically significant effect on *deviation*, i.e., the more extensively a publication draws on existing literature, the smaller the deviation from the grant proposal. We find little evidence that publication delays significantly affect deviation, however, as the coefficient for *time to publication* is not statistically significant in any of the models in Table 2.3.

As expected, we also find significant systematic differences in *deviation* across funding sources. Specifically, we find that funding from federal (Belspo) and European (non-ERC) sources are associated with higher levels of *deviation*, relative to the baseline. We conjecture these effects are due to the higher importance that these funding sources place on networking and consortium-building, resulting in projects in which the grant holders have a relatively weaker connection with the project proposal. We explore the role of

funding source heterogeneity in more detail in Appendix A.4. These analyses suggest that the relative role of knowledge recombination and coordination challenges varies depending on the funding source and that differences can typically be rationalized based on the specific features of the funding source.[35]

2.4.2 Consequences of Deviation

We present our findings on the relationship between deviation and scientific impact using the two-step control function approach in Table 2.4. We use as dependent variable the (log of) the *field citation ratio* (Models 1-3), our *top 10* dummy defined as publications above the 90th percentile of total citations within field and year of publication (Models 4-6), and our *novelty* dummy identifying publications that introduced at least one new word to the entire corpus of science, provided this word has been used at least once more subsequently (Models 7-9). These specifications build on the previous section on the antecedents of deviation, introducing the knowledge recombination and coordination challenge variables stepwise. We also introduce grant-level, publication-level, and researcher-level controls, including prior research performance for both grant holders and new team members. Additionally, in Tables 2.5 and 2.6 we report results from an extended analysis examining whether a larger deviation (defined as publications with deviation falls between the 75th and 90th percentiles) is associated with higher scientific impact. We also explore the relationship between large deviation and two additional dependent variables: the total citation count as of December 2022 (*citations*) and the likelihood of falling within the 10% of least cited publications within the field and year in our data (*bottom 10*).

Our two-step control function estimates uncover a positive and statistically significant relationship between deviation and scientific impact.[36] This

[35]For example, in the case of ERC funding -the most selective scheme in Europe- where grant holders are typically scientists of exceptional prominence, we find relatively stronger evidence for the knowledge recombination mechanism, but little to suggest that coordination challenges play a major role. This is consistent with prior studies suggesting that collaboration tends to disproportionately benefit prominent researchers (e.g., Agrawal et al., 2015).

[36]We estimate naïve OLS regression models and fail to find a significant relationship

positive effect is sizeable and consistent across all specifications, including both citation-based impact measures (Table 2.4, Models 1-6) and *novelty* (Table 2.4, Models 7-9). More specifically, we find that a 1 percent increase in *deviation* is associated with approximately a 1.81 percent increase in the *field citation ratio* (Table 2.4, Model 3). While this effect may seem small in absolute terms, its relative magnitude is substantial. The estimated coefficient suggests that the impact of a 1 percent increase in *deviation* is about 7-8 times greater than that of a 1 percent increase in the grant-holder's citation rate within the three-year window preceding the grant start (e.g., 1.81/0.24, Table 2.4, Model 3). Furthermore, the estimated coefficients imply that increasing *deviation* by 0.05 units (e.g., moving from the median value of 0.29 to the 75th percentile value of 0.34) is associated with a 30 percent increase in the field citation ratio, all else being equal. Similarly, our results consistently demonstrate a positive effect of deviation on both the *top 10* (Table 2.4, Models 4-6) and *novelty* dummy variables (Table 2.4, Models 7-9). This suggests that deviation not only has an incremental effect on citation average but is also consistent with research that contributes disproportionately to the advancement of scholarship as measured by *top 10* and *novelty*. This effect of deviation on scientific novelty supports our knowledge recombination argument, where the addition of new team members plays a key role in shifting the research direction after the grant award by introducing fresh perspectives to the initial plan.

To provide additional insight into the role of large as opposed to marginal deviation, we re-estimate our CFA models in Table 2.4, replacing the continuous *deviation* with a dummy variable capturing large *deviation*. We employ two dummies: one for publications with a *deviation* value above the 75th percentile (Table 2.5, Models 1-5) and another for *deviation* above the 90th

(Appendix A.5, Table A.5.1 and Table A.5.2). The downward bias in the OLS estimates is consistent with the CFA model primarily capturing the unobserved "research difficulties" aspect that may confound the relationship between deviation and scientific impact. Our conjecture that a naïve OLS regression would produce downward biased estimates of the effect of deviation on scientific impact is further supported by the first-stage residuals included in the CFA model, which are consistently statistically significant across all specifications.

percentile (Table 2.6, Models 6-10). The results in Tables 2.5 and 2.6 show that, while the effect on the *field citation ratio* and *citations* is no longer statistically significant, we still find a positive and statistically significant result for both the *top 10* and *novelty* dummy variables. Moreover, we find no statistically significant relationship between more extensive deviations and publications bottom 10 (i.e., no increased likelihood of publication "flops"). The lack of significance for the *citations* and *field citation ratio* measures, compared to the statistically significant results for the *top 10* and *novelty* dummies, further suggests that extensive *deviation* may not consistently increase the average impact of the published output, as measured by higher overall (field-adjusted) citations. Instead, it is more likely to contribute to breakthrough research that receives increased attention from the scientific community or introduces novel concepts. We interpret this evidence to be consistent with the idea that deviation can be regarded as a sign of scientific creativity, often stemming from knowledge recombination.

Regarding the knowledge recombination variables, our CFA estimates indicate a negative and statistically significant residual effect of the *share of new team members* (Table 2.4, Models 1-6), and of the distant new team members dummy variable (Table 2.4, Models 1-9) on scientific impact. Importantly, these statistically significant relationships in the second stage reflect an influence on scientific impact beyond their role in explaining deviation, as both variables are significant predictors in the first-stage equation. We conjecture that the negative coefficients in the second-stage regression reflect collaboration costs not directly captured by our coordination costs metric. This is further supported by our failure to find a consistent and statistically significant relationship between the *first collaboration* dummy, as well as the interaction terms, and scientific impact (Table 2.4, Models 1-9).

Regarding the grant holder's prior research performance, we find a positive influence of their prior citation rate on the *field citation ratio* (Table 2.4, Models 1-3), but no significant effect on the *top 10* or *novelty* dummy variables. Interestingly, our findings also indicate that the prior research performance of new team members does not have a direct influence on scientific

impact, as we fail to find statistically significant coefficients for new team members' prior citation rate or publication rate. This result is consistent with prior research documenting that researchers tend to collaborate with others who have similar citation profiles (Davies et al., 2022; Rivera et al., 2010).

As expected, we also find statistically significant differences in scientific impact across funding sources.[37] Finally, also in line with existing evidence (e.g., Ahlgren et al., 2018), our CFA results consistently show the positive influence of *embeddedness in prior research*, as measured by the extension of the reference list in the publication, on citation-based impact (Table 2.4, Models 1-6). Overall, our set of analyses provide consistent evidence that deviation from grant proposals is associated with greater scientific impact.[38]

2.5 Conclusion

We analyze the role played by new team members on deviation from grant proposals, controlling for other characteristics, with the aim of deriving implications of such deviation on scientific impact. Using natural language processing techniques, we quantify the degree of alignment between planned and conducted research (i.e., deviation) and we document the influence of team members joining post grant award, as antecedents of deviation. We find

[37]Specifically, we find that publications supported by Belspo show a consistently lower scientific impact, relative to the FWO baseline, which likely reflects the specificities of different funding sources, in this case producing high-impact research versus networking and collaboration promotion across regions, as it is the case with the IAP scheme funded by Belspo discussed in the Annex D.

[38]In supplementary analyses, we revisit the analyses of scientific impact using citations as an alternative dependent variable by estimating Poisson regressions and find consistent results (Appendix A.5, Table A.5.4). We further confirm our findings in variations of our estimation sample. Specifically, we estimate the effect of deviation on the (ln) field citation ratio in two extended versions of our main sample. In the Annex E, Table A.5.5, we include publication-grant pairs linked to highly productive grants, defined as the top 1% of the distribution and equivalent to those producing more than 75 publications, that were previously excluded of our main estimation sample. In the Annex E, Table A.5.6, we incorporate publication-grant pairs where the publication emerges shortly after the grant start date, adjusting the previously employed threshold from one year to two months. Across these variations, we find a consistently positive and statistically significant association between deviation and scientific impact, that is supported by our CFA estimates.

that a larger presence of team members joining the research team after the grant award is positively associated with deviation from the grant proposal.

Our second analysis employs a control function approach that exploits the linguistic features of grant proposals to understand deviation, and we establish a positive relationship between deviation and scientific impact. Our findings suggest that teams willing to deviate from the original research proposal tend to produce research of higher scientific impact, as measured by citations, and are more likely to introduce new concepts to the existing body of scientific knowledge.

In our investigation of mechanisms, we find consistent support for the knowledge recombination perspective and limited support for the moderating role of coordination challenges. In doing so, we find a positive effect associated with new team members affiliated with external organizations and whose prior research is more distant relative to the grant proposal. Conversely, research teams initiating their first collaboration exhibit a relatively lower degree of deviation, which we interpret as reflecting higher coordination costs. However, we find little consistent evidence supporting the notion that teams engaging in a first collaboration with team members affiliated with external organizations or more distant prior research relative to the grant, experience heightened coordination challenges.

Our results raise questions about the efficiency and effectiveness of the resource-intensive peer review process in allocating science funding and the close monitoring of research outputs of grant-funded projects. Our findings suggest that strict monitoring of adherence to initial proposals may not be beneficial under certain conditions, such as curiosity-driven changes in research direction. While grant proposals provide valuable information on the merit of a research idea and the teams' capabilities, our findings regarding deviation and its positive role on scientific impact suggests a reevaluation of the investment in scrutinizing proposals. Our results contribute to the ongoing debate on the net benefit of detailed project evaluation (Herbert et al., 2013; Ioannidis, 2011), complementing documented challenges in the ex-ante selection of research projects (Boudreau et al., 2016; Lane et al., 2022).

Our analysis contains several limitations. First, we are unable to identify a causal assessment of team changes in the analysis of deviation antecedents. Future research could investigate exogenous changes within funded research teams, such as the exit of team members or unexpected events like the unexpected death of a grant holder (Azoulay et al., 2019; Khanna, 2021). Second, while our focus on team changes offers conceptual and empirical tractability, our study remains largely descriptive regarding other factors influencing deviation, such as serendipitous discoveries or competition-induced deviation. Lastly, although our single-institution focus provides an ideal setting to track "paper trails" of change in team composition and its effect on grant-funded research direction, enhancing the generalizability of our findings calls for future research to examine research teams in other contexts.

Our study advances understanding of the role that new team members play in shaping funded research by demonstrating their influence on deviation from original grant proposals and, ultimately, on scientific impact. Further research should build on these insights to unpack the mechanisms underlying such effects. In particular, future work could clarify the channels through which new team members exert influence—for example, whether they contribute through broader networks or bring a specific skill that the original team required. It would also be valuable to distinguish between opportunity-driven deviations, which reflect novel insights not foreseen at the grant application stage (Yaqub, 2018), and necessity-driven deviations, such as responses to competition or being "scooped" by other teams (Hill et al., 2025). Moreover, given that many grant-supported publications draw on multiple funding sources and principal investigators often treat grants as part of a broader research portfolio, future studies could explore how overlapping funding streams shape research direction and deviation. Finally, examining how teams' capabilities and strategic orientations moderate the relationship between deviation and scientific impact would provide a deeper understanding of when and why deviation enhances scientific progress.

2.6 Tables and Figures

Table 2.1: Summary statistics, publication-grant pairs (N=1,341)

Variable	Description	Mean	SD	Min	Max
Deviation	Degree of deviation between publication and grant abstract (SciBERT)	0.29	0.07	0.07	0.54
Citations	Raw citation count as of December 2022. Sourced from the Dimensions database	41.23	54.32	0	617
Field Citation Ratio	Relative citation performance vs. similarly aged articles in the same field. 41 missing values	7.86	10.47	0	111
Novelty (0/1)	1 if publication introduces at least one new word (used at least twice). 64 missing values	0.07	0.25	0	1
Top 10 (0/1)	1 if publication belongs to top 10% of citation distribution (by FoR & year)	0.10	0.29	0	1
Bottom 10 (0/1)	1 if publication belongs to bottom 10% of citation distribution (by FoR & year)	0.14	0.35	0	1
Authors count	Number of authors listed in the publication	5.27	2.97	1	15
Grant holders (GH)	Number of authors awarded the grant funding acknowledged by the publication	1.16	0.82	0	10
New team members	Authors not awarded the grant funding acknowledged in the publication	3.98	2.93	0	14
Share of new team members	Share of new team members with academic age $\geq$ 3 years	0.68	0.33	0	1
Internal (0/1)	All new team members affiliated with same org. as GH	0.26	0.44	0	1
External (0/1)	All new team members affiliated with different orgs. from GH	0.41	0.49	0	1

Continued on next page

Table 2.1 – *continued from previous page*

Variable	Description	Mean	SD	Min	Max
Mixed (0/1)	New team includes both internal and external members	0.28	0.45	0	1
Distant new team members (0/1)	Mean distance of new team members' prior pubs vs. grant abstract in top 25%	0.25	0.43	0	1
First collaboration (0/1)	At least one first collaboration among GH and new team	0.57	0.49	0	1
GH publication rate (log)	Coauthor-weighted pubs. by GH in 3 yrs before grant	0.86	0.54	0	2.31
GH citation rate (log)	Coauthor-weighted yearly citations per pub. by GH in 3 yrs before grant	0.12	0.21	0	1.33
GH multiple funding (0/1)	1 if GH has other active grants not acknowledged in focal pub.	0.39	0.49	0	1
GH absence (0/1)	1 if none of the GH are part of the authors list	0.09	0.29	0	1
New team pub. rate (log)	Coauthor-weighted pubs. of new members in 3 yrs before grant	0.11	0.19	0	1.54
New team cit. rate (log)	Coauthor-weighted yearly citations of new members in 3 yrs before grant	0.36	0.37	0	2.14
New team multiple funding (0/1)	1 if new members hold grants not acknowledged in focal pub.	0.36	0.48	0	1
Embeddedness in prior research	References / avg. references in FoR*	0.98	0.57	0	3.71
Time to publication (field-norm.)	Years from grant start to pub. date (normalized by FoR)	0.58	1.07	0	6.05
Year	Year of publication	2014.28	3.58	2003	2018

Notes: *Field of Research (FoR) as provided by Dimensions (ANZSRC 2020).

Table 2.2: Summary statistics, grants (N=460)

Variable	Description	Mean	SD	Min	Max
Grant awardees count	Nr of researchers that were awarded a given grant	2.45	2.47	1	14
Publications per grant	Nr of publications that acknowledge a given grant	13.18	14.15	1	74
Grant abstract length	Count of words in the grant abstract (in hundreds of words)	3.38	2.62	0.54	15.44
Curiosity	% of words in the grant abstract that express curiosity (LIWC-22 metric)	0.72	0.96	0	9.64
Grant start	Year of grant start	2012	4.04	2000	2017
Funding source:					
Belspo	1 if the grant is funded by the Belgian Science Policy Office (Belspo)	0.12	0.32	0	1
ERC	1 if the grant is funded by the European Research Council (ERC)	0.13	0.34	0	1
FWO	1 if the grant is funded by the Research Foundation Flanders (FWO)	0.32	0.47	0	1
Other within the EU	1 if other national funding agencies in Europe fund the grant	0.24	0.43	0	1
Other outside the EU	1 if other funding agencies outside Europe fund the grant	0.19	0.39	0	1

Table 2.3: The antecedents of deviation and grant openness (OLS)

DV : *Deviation*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Grant openness									
Curiosity		0.006**	0.007***	0.006**	0.007***	0.007**	0.007**	0.006**	0.007**
		[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Grant abstract word count (in 100s)		-0.008**	-0.008**	-0.008*	-0.008**	-0.011***	-0.012***	-0.011***	-0.012***
		[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Grant abstract word count sq. (in 100s)		0.000*	0.001*	0.000*	0.000*	0.001**	0.001***	0.001**	0.001***
		[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Authors count (ln)			0.008*	0.006	0.006	0.005	0.009	0.008	0.009
			[0.004]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Share of new team members				0.015**	0.012*	0.013*	0.012*	0.011	0.012*
				[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
Affiliations of new team members (NTM)(base: Internal)									
Mixed (0/1)					0.002	0.003	0.003	-0.002	0.003
					[0.005]	[0.005]	[0.005]	[0.008]	[0.005]
External (0/1)					0.014**	0.014**	0.014**	0.019**	0.014**
					[0.006]	[0.006]	[0.006]	[0.008]	[0.006]
Distant new team members (0/1)						0.023***	0.023***	0.023***	0.028***
						[0.008]	[0.008]	[0.008]	[0.010]
First collaboration (0/1)							-0.009**	-0.008	-0.007
							[0.004]	[0.007]	[0.005]
First collab. x Affiliations of NTM (baseline: Internal)									
First collab. x Mixed								0.008	
								[0.010]	
First collab. x External								-0.008	
								[0.010]	
First collab. x Distant new team members									-0.009
									[0.009]
GH citation rate (ln; 3yrs pre-grant)	-0.013	-0.014	-0.014	-0.015	-0.015	-0.012	-0.011	-0.012	-0.011
	[0.016]	[0.015]	[0.015]	[0.015]	[0.016]	[0.015]	[0.015]	[0.015]	[0.015]
GH publication rate (ln; 3yrs pre-grant)	0.017**	0.016**	0.017**	0.016**	0.016**	0.017**	0.017**	0.017**	0.017**
	[0.008]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
GH multiple funding (0/1)	-0.008	-0.009	-0.010*	-0.010*	-0.011*	-0.012**	-0.012**	-0.011**	-0.011*
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
GH absence (0/1)	0.016	0.016	0.017	0.015	0.016	0.016	0.01	0.012	0.008
	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.013]	[0.013]	[0.013]
Time to publication (field-norm.)	-0.003	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Embeddedness in prior research	-0.019***	-0.019***	-0.018***	-0.019***	-0.018***	-0.018***	-0.018***	-0.018***	-0.018***
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Grant productivity (field-norm.)	-0.007	-0.005	-0.006	-0.006	-0.006	-0.01	-0.01	-0.009	-0.01
	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]
Funding source (Baseline= FWO)									
Belspo (0/1)	0.069***	0.081***	0.078***	0.077***	0.076***	0.068***	0.068***	0.068***	0.067***
	[0.013]	[0.017]	[0.017]	[0.017]	[0.017]	[0.016]	[0.017]	[0.017]	[0.016]
ERC (0/1)	0.004	0.008	0.008	0.008	0.006	0.01	0.011	0.01	0.01
	[0.011]	[0.011]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.011]	[0.010]
Other funder within the EU (0/1)	0.035***	0.043***	0.043***	0.041***	0.034***	0.032***	0.031***	0.033***	0.031***
	[0.008]	[0.009]	[0.009]	[0.009]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Other funder outside the EU (0/1)	0.017	0.021*	0.020*	0.019*	0.013	0.007	0.006	0.007	0.006
	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.011]	[0.011]	[0.011]
Constant	0.282***	0.301***	0.284***	0.280***	0.281***	0.287***	0.287***	0.287***	0.284***
	[0.034]	[0.036]	[0.037]	[0.036]	[0.036]	[0.035]	[0.034]	[0.036]	[0.035]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	1341	1341	1341	1341	1341	1341	1341	1341	1341
Total R-sq.	0.210	0.222	0.225	0.229	0.235	0.246	0.248	0.250	0.249
Adj. R-sq.	0.181	0.192	0.195	0.198	0.203	0.213	0.216	0.216	0.216
Partial R-sq. (Grant openness)		0.018	0.019	0.019	0.020	0.021	0.020	0.020	0.020

Notes: Deviation is measured as the inverse cosine similarity between the publication and grant abstract.GH denotes grant holders. Robust standard errors, clustered at the grant level, are shown in brackets. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table 2.4: The consequences of deviation on scientific impact and novelty

	Field Citation Ratio (ln)			Top 10 (0/1)			Novelty (0/1)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Deviation (ln)	1.841**	1.817*	1.808**	0.832**	0.835**	0.816**	0.389*	0.403**	0.383*
	[0.913]	[0.938]	[0.907]	[0.365]	[0.372]	[0.360]	[0.199]	[0.203]	[0.197]
Authors count (ln)	-0.042	-0.038	-0.041	-0.029	-0.027	-0.029	0.003	0.006	0.003
	[0.071]	[0.070]	[0.071]	[0.024]	[0.023]	[0.024]	[0.016]	[0.016]	[0.016]
Share of new team members	-0.213**	-0.199*	-0.212**	-0.075**	-0.071*	-0.074**	-0.002	0.000	-0.002
	[0.102]	[0.101]	[0.102]	[0.037]	[0.036]	[0.037]	[0.025]	[0.025]	[0.024]
Affiliations of new team members (NTM) (baseline: Internal)									
Mixed (0/1)	0.016	0.096	0.016	0.024	0.050	0.024	0.036**	0.052**	0.035**
	[0.067]	[0.107]	[0.067]	[0.022]	[0.034]	[0.022]	[0.017]	[0.026]	[0.017]
External (0/1)	-0.094	-0.034	-0.092	-0.023	-0.011	-0.022	0.004	-0.022	0.004
	[0.078]	[0.107]	[0.078]	[0.027]	[0.037]	[0.027]	[0.023]	[0.030]	[0.023]
Distant new team members (0/1)	-0.173*	-0.172*	-0.209*	-0.073**	-0.073**	-0.103**	-0.039*	-0.039*	-0.050*
	[0.092]	[0.092]	[0.126]	[0.030]	[0.030]	[0.040]	[0.021]	[0.021]	[0.026]
First collaboration (0/1)	0.096	0.188*	0.082	0.011	0.035	-0.001	0.020	0.004	0.015
	[0.064]	[0.097]	[0.070]	[0.022]	[0.032]	[0.024]	[0.018]	[0.027]	[0.019]
First collab. x Affiliation of NTM (baseline: Internal x First collab.)									
First collab x Mixed		-0.149			-0.048			-0.022	
		[0.124]			[0.044]			[0.036]	
First collab x External		-0.119			-0.025			0.047	
		[0.106]			[0.038]			[0.037]	
First collab x Distant new team members			0.064			0.052			0.018
			[0.105]			[0.037]			[0.031]
GH citation rate (ln; 3yrs pre-grant)	0.240*	0.238*	0.238*	0.053	0.054	0.052	0.056	0.058	0.056
	[0.132]	[0.133]	[0.132]	[0.042]	[0.042]	[0.041]	[0.042]	[0.043]	[0.043]
GH publication rate (ln; 3yrs pre-grant)	0.005	0.004	0.006	-0.042	-0.042	-0.041	-0.043**	-0.043**	-0.042**
	[0.087]	[0.088]	[0.087]	[0.029]	[0.029]	[0.029]	[0.020]	[0.020]	[0.020]
GH multiple funding (0/1)	0.068	0.067	0.064	0.042*	0.042*	0.039	-0.019	-0.018	-0.020
	[0.072]	[0.071]	[0.072]	[0.025]	[0.025]	[0.025]	[0.020]	[0.020]	[0.020]
GH absence (0/1)	0.025	0.024	0.036	0.000	-0.003	0.007	-0.063*	-0.075**	-0.060*
	[0.140]	[0.146]	[0.140]	[0.040]	[0.042]	[0.039]	[0.035]	[0.037]	[0.034]
NTM publication rate (ln; 3yrs pre-grant)	-0.117	-0.122	-0.113	-0.061	-0.062	-0.059	-0.019	-0.013	-0.018
	[0.137]	[0.135]	[0.137]	[0.041]	[0.040]	[0.040]	[0.040]	[0.040]	[0.040]
NTM citation rate (ln; 3yrs pre-grant)	0.025	0.028	0.026	0.017	0.018	0.018	0.001	0.004	0.002
	[0.084]	[0.085]	[0.084]	[0.026]	[0.026]	[0.026]	[0.022]	[0.022]	[0.022]
NTM multiple funding (0/1)	0.070	0.067	0.071	-0.017	-0.018	-0.016	0.053***	0.052**	0.053***
	[0.068]	[0.068]	[0.068]	[0.022]	[0.022]	[0.022]	[0.020]	[0.020]	[0.020]
Time to publication (field-norm.)	0.008	0.006	0.006	0.008	0.008	0.007	0.010	0.011	0.009
	[0.031]	[0.032]	[0.031]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]
Embeddedness in prior research	0.496***	0.498***	0.495***	0.138***	0.139***	0.138***	0.010	0.011	0.010
	[0.071]	[0.073]	[0.071]	[0.030]	[0.030]	[0.030]	[0.016]	[0.017]	[0.016]
Grant productivity (field-norm.)	0.005	0.003	0.003	0.016	0.015	0.015	-0.001	-0.003	-0.001
	[0.090]	[0.090]	[0.090]	[0.023]	[0.023]	[0.023]	[0.022]	[0.021]	[0.022]
Funding source (*Baseline: FWO*)									
Belspo (0/1)	-0.490**	-0.477**	-0.479**	-0.119	-0.117	-0.112	-0.013	-0.014	-0.010
	[0.225]	[0.227]	[0.222]	[0.086]	[0.086]	[0.084]	[0.051]	[0.051]	[0.051]
ERC (0/1)	0.172	0.176	0.175	0.051	0.053*	0.053*	0.023	0.024	0.024
	[0.112]	[0.112]	[0.111]	[0.031]	[0.031]	[0.031]	[0.019]	[0.019]	[0.019]
Other funder within the EU (0/1)	-0.137	-0.133	-0.133	-0.035	-0.036	-0.032	-0.039	-0.048	-0.038
	[0.136]	[0.144]	[0.135]	[0.052]	[0.055]	[0.052]	[0.030]	[0.031]	[0.029]
Other funder outside the EU (0/1)	0.037	0.040	0.038	0.030	0.030	0.030	0.011	0.006	0.011
	[0.101]	[0.102]	[0.101]	[0.030]	[0.030]	[0.030]	[0.026]	[0.027]	[0.026]
Constant	4.183***	4.060***	4.152***	1.088**	1.067**	1.079**	0.662**	0.682**	0.658**
	[1.360]	[1.388]	[1.358]	[0.526]	[0.534]	[0.523]	[0.324]	[0.330]	[0.324]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	1296	1296	1296	1341	1341	1341	1277	1277	1277
adj. R-sq	0.173	0.173	0.172	0.116	0.115	0.115	0.034	0.033	0.033
1st stage residual	-1.901**	-1.879**	-1.867**	-0.829**	-0.833**	-0.812**	-0.394**	-0.406**	-0.388**
	[0.920]	[0.944]	[0.915]	[0.367]	[0.374]	[0.362]	[0.198]	[0.202]	[0.196]

Notes: CFA: Estimates from a two-step control function approach; detailed first-stage results are available in Appendix E, Table E3. Deviation is measured as the inverse cosine similarity between the publication and grant abstract. GH denotes grant holders. NTM denotes new team members. Field Citation Ratio represents relative citation performance compared to similarly aged articles in the same research group (sourced from the Dimensions database, 41 missing values). Top 10 equals 1 if a publication is in the top 10% of the citation distribution by Field of Research and year of publication in our data. Novelty equals 1 if a publication introduces at least one new word to the corpus of science, and the word is used at least twice (adapted from Arts et al., 2023; 64 missing values). Robust standard errors, clustered at the grant level, are shown in brackets.

Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table 2.5: The consequences of extensive deviation on scientific impact and novelty (CFA) – Panel A

	Field Citation Ratio (ln) (1)	Citations (ln) (2)	Bottom 10 (0/1) (3)	Top 10 (0/1) (4)	Novelty (0/1) (5)
Percentile 75 of Deviation (0/1)	0.359	0.099	0.068	0.332**	0.199*
	[0.444]	[0.502]	[0.158]	[0.151]	[0.107]
Authors count (ln)	0.026	0.186***	-0.049**	-0.003	0.015
	[0.061]	[0.070]	[0.021]	[0.019]	[0.015]
Share of new team members	-0.121	-0.132	0.014	-0.037	0.012
	[0.089]	[0.101]	[0.035]	[0.032]	[0.021]
Affiliations of new team members (base: Internal)					
Mixed (0/1)	0.022	0.044	0.007	0.029	0.034**
	[0.067]	[0.080]	[0.027]	[0.022]	[0.017]
External (0/1)	-0.051	-0.046	0.031	-0.016	0.000
	[0.077]	[0.094]	[0.035]	[0.026]	[0.023]
Distant new team members (0/1)	-0.092	-0.052	-0.023	-0.044*	-0.027
	[0.078]	[0.093]	[0.034]	[0.023]	[0.019]
First collaboration (0/1)	0.050	0.035	0.021	-0.004	0.016
	[0.061]	[0.070]	[0.027]	[0.020]	[0.017]
GH citation rate (ln; 3yrs pre-grant)	0.186	0.269	-0.046	0.045	0.054
	[0.135]	[0.165]	[0.048]	[0.041]	[0.042]
GH publication rate (ln; 3yrs pre-grant)	0.077	0.111	-0.036	-0.022	-0.038**
	[0.079]	[0.094]	[0.029]	[0.024]	[0.019]
GH multiple funding (0/1)	0.032	0.023	0.006	0.035	-0.018
	[0.070]	[0.087]	[0.031]	[0.024]	[0.020]
GH absence (0/1)	0.060	0.125	-0.008	0.000	-0.062*
	[0.145]	[0.158]	[0.053]	[0.040]	[0.035]
NTM publication rate (ln; 3yrs pre-grant)	-0.032	0.025	-0.034	-0.015	0.005
	[0.128]	[0.131]	[0.050]	[0.035]	[0.039]
NTM citation rate (ln; 3yrs pre-grant)	0.033	0.016	0.001	0.023	0.008
	[0.086]	[0.095]	[0.035]	[0.026]	[0.024]
NTM multiple funding (0/1)	0.061	0.084	-0.028	-0.018	0.051**
	[0.068]	[0.079]	[0.029]	[0.022]	[0.020]
Time to publication (field-norm.)	-0.013	-0.019	0.008	0.005	0.009
	[0.031]	[0.036]	[0.015]	[0.010]	[0.011]
Embeddedness in prior research	0.406***	0.561***	-0.104***	0.112***	0.002
	[0.057]	[0.064]	[0.021]	[0.022]	[0.014]
Grant productivity (field-norm.)	-0.033	-0.060	0.010	0.012	0.000
	[0.089]	[0.096]	[0.030]	[0.022]	[0.021]
Funding source (Baseline: FWO)					
Belspo (0/1)	-0.213	-0.177	0.034	-0.038	0.014
	[0.170]	[0.190]	[0.061]	[0.055]	[0.042]
ERC (0/1)	0.234**	0.246**	-0.037	0.084***	0.039**
	[0.105]	[0.120]	[0.034]	[0.027]	[0.019]
Other funder within the EU (0/1)	0.019	-0.015	-0.009	0.023	-0.014
	[0.101]	[0.115]	[0.040]	[0.033]	[0.024]
Other funder outside the EU (0/1)	0.055	-0.060	-0.030	0.042	0.018
	[0.101]	[0.111]	[0.042]	[0.029]	[0.027]
Constant	1.388***	3.696***	0.159*	-0.265***	0.006
	[0.375]	[0.348]	[0.096]	[0.094]	[0.152]
Year FE	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y
N	1296	1341	1341	1341	1277
adj. R-sq	0.169	0.261	0.072	0.113	0.034
1st stage residual	-0.376	-0.139	-0.070	-0.340**	-0.190*
	[0.450]	[0.507]	[0.160]	[0.153]	[0.108]

Notes: CFA: Estimates from a two-step control function approach. Deviation is measured as the inverse cosine similarity between the publication and grant abstract. GH denotes grant holders. NTM denotes new team members. Field Citation Ratio represents relative citation performance compared to similarly aged articles in the same research group (sourced from the Dimensions database, 41 missing values). Citations represents total citation count as of December 2022 (sourced from the Dimensions database). Bottom 10 equals 1 if a publication is in the bottom 10% of the citation distribution by Field of Research and year of publication in our data. Top 10 equals 1 if a publication is in the top 10% of the citation distribution by Field of Research and year of publication in our data. Novelty equals 1 if a publication introduces at least one new word to the corpus of science, and the word is used at least twice (adapted from Arts et al., 2023; 64 missing values). Robust standard errors, clustered at the grant level, are shown in brackets. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.6: The consequences of extensive deviation on scientific impact and novelty (CFA) – Panel B

	Field Citation Ratio (ln) (6)	Citations (ln) (7)	Bottom 10 (0/1) (8)	Top 10 (0/1) (9)	Novelty (0/1) (10)
Percentile 90 of Deviation (0/1)	0.602	0.219	0.247	0.546**	0.236*
	[0.682]	[0.757]	[0.216]	[0.242]	[0.142]
Authors count (ln)	0.044	0.193***	-0.042**	0.016	0.024
	[0.061]	[0.072]	[0.021]	[0.020]	[0.016]
Share of new team members	-0.119	-0.133	0.010	-0.037	0.016
	[0.088]	[0.100]	[0.035]	[0.032]	[0.021]
Affiliations of new team members (base: Internal)					
Mixed (0/1)	0.029	0.046	0.008	0.032	0.036**
	[0.068]	[0.080]	[0.027]	[0.022]	[0.017]
External (0/1)	-0.049	-0.048	0.026	-0.014	0.005
	[0.076]	[0.092]	[0.034]	[0.025]	[0.023]
Distant new team members (0/1)	-0.110	-0.061	-0.035	-0.060**	-0.031
	[0.087]	[0.102]	[0.036]	[0.027]	[0.020]
First collaboration (0/1)	0.035	0.031	0.018	-0.017	0.009
	[0.060]	[0.069]	[0.026]	[0.020]	[0.016]
GH citation rate (ln; 3yrs pre-grant)	0.146	0.257*	-0.055	0.008	0.036
	[0.119]	[0.155]	[0.048]	[0.034]	[0.041]
GH publication rate (ln; 3yrs pre-grant)	0.057	0.103	-0.049	-0.035	-0.038*
	[0.089]	[0.100]	[0.031]	[0.027]	[0.019]
GH multiple funding (0/1)	0.032	0.025	0.010	0.035	-0.021
	[0.070]	[0.086]	[0.030]	[0.024]	[0.020]
GH absence (0/1)	0.065	0.124	-0.012	0.006	-0.056
	[0.141]	[0.155]	[0.053]	[0.039]	[0.035]
NTM publication rate (ln; 3yrs pre-grant)	-0.041	0.022	-0.035	-0.026	-0.003
	[0.128]	[0.129]	[0.049]	[0.035]	[0.039]
NTM citation rate (ln; 3yrs pre-grant)	0.028	0.015	0.002	0.019	0.004
	[0.085]	[0.094]	[0.034]	[0.026]	[0.023]
NTM multiple funding (0/1)	0.035	0.075	-0.037	-0.041*	0.041**
	[0.071]	[0.081]	[0.030]	[0.022]	[0.021]
Time to publication (field-norm.)	-0.026	-0.022	0.006	-0.004	0.004
	[0.026]	[0.031]	[0.014]	[0.009]	[0.011]
Embeddedness in prior research	0.395***	0.560***	-0.101***	0.102***	-0.008
	[0.047]	[0.055]	[0.018]	[0.020]	[0.011]
Grant productivity (field-norm.)	-0.015	-0.053	0.021	0.023	0.000
	[0.099]	[0.102]	[0.031]	[0.025]	[0.021]
Funding source (Baseline: FWO)					
Belspo (0/1)	-0.221	-0.186	0.012	-0.039	0.028
	[0.170]	[0.185]	[0.058]	[0.055]	[0.039]
ERC (0/1)	0.247**	0.251**	-0.029	0.095***	0.043**
	[0.104]	[0.120]	[0.035]	[0.027]	[0.020]
Other funder within the EU (0/1)	0.059	-0.002	0.003	0.062**	0.007
	[0.085]	[0.103]	[0.039]	[0.025]	[0.023]
Other funder outside the EU (0/1)	0.041	-0.062	-0.032	0.035	0.015
	[0.103]	[0.112]	[0.041]	[0.030]	[0.027]
Constant	1.552***	3.736***	0.181***	-0.118**	0.099
	[0.320]	[0.267]	[0.059]	[0.058]	[0.144]
Year FE	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y
N	1296	1341	1341	1341	1277
adj. R-sq	0.169	0.261	0.072	0.115	0.033
1st stage residual	-0.588	-0.227	-0.249	-0.543**	-0.232*
	[0.679]	[0.760]	[0.218]	[0.243]	[0.139]

Notes: CFA: Estimates from a two-step control function approach. Deviation is measured as the inverse cosine similarity between the publication and grant abstract. GH denotes grant holders. NTM denotes new team members. Field Citation Ratio represents relative citation performance compared to similarly aged articles in the same research group (sourced from the Dimensions database, 41 missing values). Citations represents total citation count as of December 2022 (sourced from the Dimensions database). Bottom 10 equals 1 if a publication is in the bottom 10% of the citation distribution by Field of Research and year of publication in our data. Top 10 equals 1 if a publication is in the top 10% of the citation distribution by Field of Research and year of publication in our data. Novelty equals 1 if a publication introduces at least one new word to the corpus of science, and the word is used at least twice (adapted from Arts et al., 2023; 64 missing values). Robust standard errors, clustered at the grant level, are shown in brackets. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

References

Aghion, P., M. Dewatripont, and J. C. Stein (2008): "Academic Freedom, Private-Sector Focus, and the Process of Innovation," *RAND Journal of Economics*, 39, 617–635.

Agrawal, A., J. McHale, and A. Oettl (2015): "Collaboration, Stars and the Changing Organization of Science: Evidence from Evolutionary Biology," in *The Changing Frontier: Rethinking Science and Innovation Policy*, ed. by A. Jaffe and B. Jones, University of Chicago Press, 75–102.

Ahlgren, P., C. Colliander, and P. Sjögårde (2018): "Exploring the Relation Between Referencing Practices and Citation Impact: A Large-Scale Study Based on Web of Science Data," *Journal of the Association for Information Science and Technology*, 69, 728–743.

Antons, D. and F. T. Piller (2014): "Opening the Black Box of "Not-Invented-Here": Attitudes, Decision Biases, and Behavioral Consequences," *Academy of Management Perspectives*, 29.

Arrow, K. J. (1962): "Economic Welfare and the Allocation of Resources for Invention," in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Princeton University Press.

Arts, S., J. Hou, and J. C. Gomez (2021): "Natural Language Processing to Identify the Creation and Impact of New Technologies in Patent Text: Code, Data, and New Measures," *Research Policy*, 50, 104144.

Arts, S., N. Melluso, and R. Veugelers (2025): "Beyond Citations: Measuring Novel Scientific Ideas and their Impact in Publication Text," *Review of Economics and Statistics*, 1–33.

Aslan, Y., O. Yaqub, B. N. Sampat, and D. Rotolo (2024): "Unexpectedness in Medical Research," *Research Policy*, 53, 105075.

Audia, P. G. and J. A. Goncalo (2007): "Past Success and Creativity over Time: A Study of Inventors in the Hard Disk Drive Industry," *Management Science*, 53, 1–15.

Ayoubi, C., M. Pezzoni, and F. Visentin (2017): "At the Origins of Learning: Absorbing Knowledge Flows from Within the Team," *Journal of Economic Behavior and Organization*, 134, 374–387.

Azoulay, P., C. Fons-Rosen, and J. S. Graff Zivin (2019): "Does Science Advance One Funeral at a Time?" *American Economic Review*, 109, 2889–2920.

Azoulay, P., J. S. Graff Zivin, and G. Manso (2011): "Incentives and Creativity: Evidence from the Academic Life Sciences," *RAND Journal of Economics*, 42, 527–554.

Azoulay, P. and D. Li (2020): "Scientific Grant Funding," Tech. Rep. 26889, National Bureau of Economic Research.

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

Baruffaldi, S. and F. Gaessler (forthcoming): "The Returns to Physical Capital in Knowledge Production: Evidence from Lab Disasters," *American Economic Journal: Applied Economics*.

Beltagy, I., K. Lo, and A. Cohan (2019): "SCIBERT: A Pretrained Language Model for Scientific Text," `https://github.com/allenai/scibert/`.

Bercovitz, J. and M. Feldman (2011): "The Mechanisms of Collaboration in Inventive Teams: Composition, Social Networks, and Geography," *Research Policy*, 40, 81–93.

Blomfield, M. and K. Vakili (2023): "Incentivizing Effort Allocation Through Resource Allocation: Evidence from Scientists' Response to Changes in Funding Policy," *Organization Science*, 34, 100–128.

Blundell, R., D. Kristensen, and R. L. Matzkin (2013): "Control Functions and Simultaneous Equations Methods," *American Economic Review: Papers & Proceedings*, 103, 563–569.

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

Chai, S. and A. Menon (2019): "Breakthrough Recognition: Bias Against Novelty and Competition for Attention," *Research Policy*, 48, 733–747.

Chandrasekaran, D. and V. Mago (2021): "Evolution of Semantic Similarity: A Survey," *ACM Computing Surveys*, 54, 41.

Choudhury, P., K. Doran, A. Marinoni, and C. Yoon (2024): "Loss of Peers and Individual Worker Performance: Evidence From H-1B Visa Denials," *Organization Science*.

Corley, E. A., C. P. Boardman, and B. Bozeman (2006): "Design and the Management of Multi-Institutional Research Collaborations: Theoretical Implications from Two Case Studies," *Research Policy*, 35, 975–993.

Cummings, J. N. and M. R. Haas (2012): "So Many Teams, So Little Time: Time Allocation Matters in Geographically Dispersed Teams," *Journal of Organizational Behavior*, 33, 316–341.

Cummings, J. N. and S. Kiesler (2007): "Coordination Costs and Project Outcomes in Multi-University Collaborations," *Research Policy*, 36, 1620–1634.

Dasgupta, P. and P. A. David (1994): "Toward a New Economics of Science," *Research Policy*, 23, 487–521.

Davies, B., J. Gush, S. C. Hendy, and A. B. Jaffe (2022): "Research Funding and Collaboration," *Research Policy*, 51.

De Rassenfosse, G., K. Higham, and O. Penner (2022): "Scientific Rewards for Biomedical Specialization Are Large and Persistent," *BMC Biology*, 20.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019): "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv*.

Fleming, L. and M. Marx (2006): "Managing Creativity in Small Worlds," *California Management Review*, 48, 6–27.

Fleming, L. and O. Sorenson (2004): "Science as a Map in Technological Search," *Strategic Management Journal*, 25, 909–928.

Franzoni, C., G. Scellato, and P. Stephan (2018): "Context Factors and the Performance of Mobile Individuals in Research Teams," *Journal of Management Studies*, 55, 27–59.

Grassano, N., D. Rotolo, J. Hutton, F. Lang, and M. M. Hopkins (2017): "Funding Data from Publication Acknowledgments: Coverage, Uses, and Limitations," *Journal of the Association for Information Science and Technology*, 68, 999–1017.

Haeussler, C. and H. Sauermann (2020): "Division of Labor in Collaborative Knowledge Production: The Role of Team Size and Interdisciplinarity," *Research Policy*, 49.

Heinze, T. and S. Kuhlmann (2008): "Across Institutional Boundaries? Research Collaboration in German Public Sector Nanoscience," *Research Policy*, 37, 888–899.

Herbert, D. L., A. G. Barnett, P. Clarke, and N. Graves (2013): "On the Time Spent Preparing Grant Proposals: An Observational Study of Australian Researchers," *BMJ Open*, 3.

Herzog, C., D. Hook, and S. Konkiel (2020): "Dimensions: Bringing Down Barriers Between Scientometricians and Data," *Quantitative Science Studies*, 1, 387–395.

Hill, R., Y. Yin, C. Stein, X. Wang, D. Wang, and B. F. Jones (2025): "The pivot penalty in research," *Nature*, 642, 999–1006.

Hofstra, B., V. V. Kulkarni, S. M. N. Galvez, B. He, D. Jurafsky, and D. A. McFarland (2020): "The Diversity–Innovation Paradox in Science," *Proceedings of the National Academy of Sciences*, 117, 9284–9291.

Hottenrott, H. and C. Lawson (2014): "Research Grants, Sources of Ideas and the Effects on Academic Research," *Economics of Innovation and New Technology*, 23, 109–133.

——— (2017): "Fishing for Complementarities: Research Grants and Research Productivity," *International Journal of Industrial Organization*, 51, 1–38.

Huisman, J. and J. Smits (2017): "Duration and Quality of the Peer Review Process: The Author's Perspective," *Scientometrics*, 113, 633–650.

Hussinger, K. and A. Wastyn (2016): "In Search for the Not-Invented-Here Syndrome: The Role of Knowledge Sources and Firm Success," *R&D Management*, 46, 945–957.

Ioannidis, J. P. (2011): "Fund People Not Projects," *Nature*, 477, 529–531.

Jaravel, X., N. Petkova, and A. Bell (2018): "Team-Specific Capital and Innovation," *American Economic Review*, 108, 1034–1073.

Kaplan, S. and K. Vakili (2015): "The Double-Edged Sword of Recombination in Breakthrough Innovation," *Strategic Management Journal*, 36, 1435–1457.

Kelchtermans, S., D. Neicu, and R. Veugelers (2021): "Off the Beaten Path: What Drives Scientists' Entry into New Fields?" *Industrial and Corporate Change*, 00, 1–27.

Kelchtermans, S. and R. Veugelers (2013): "Top Research Productivity and Its Persistence: Gender as a Double-Edged Sword," *The Review of Economics and Statistics*, 95, 273–285.

Khanna, R. (2021): "Aftermath of a Tragedy: A Star's Death and Coauthors' Subsequent Productivity," *Research Policy*, 50, 104159.

Lane, J. N., M. Teplitskiy, G. Gray, H. Ranu, M. Menietti, E. C. Guinan, and K. R. Lakhani (2022): "Conservatism Gets Funded? A Field Experiment on the Role of Negative Information in Novel Project Evaluation," *Management Science*, 68, 4478–4495.

Li, D. (2017): "Expertise Versus Bias in Evaluation: Evidence from the NIH," *American Economic Journal: Applied Economics*, 9, 60–92.

Lichtenthaler, U. and H. Ernst (2006): "Attitudes to Externally Organising Knowledge Management Tasks: A Review, Reconsideration and Extension of the NIH Syndrome," *R&D Management*, 36, 367–386.

Merton, R. K. and E. Barber (2011): *The Travels and Adventures of Serendipity*, Princeton University Press.

Millar, N., B. Batalo, and B. Budgell (2022): "Trends in the Use of Promotional Language (Hype) in Abstracts of Successful National Institutes of Health Grant Applications, 1985–2020," *JAMA Network Open*, 5, e2228676.

Millar, N., B. Mathis, B. Batalo, and B. Budgell (2023): "Trends in the Expression of Epistemic Stance in NIH Research Funding Applications: 1985–2020," *Applied Linguistics*, 1–18.

Moher, D., F. Naudet, I. A. Cristea, F. Miedema, J. P. A. Ioannidis, and S. N. Goodman (2018): "Assessing Scientists for Hiring, Promotion, and Tenure," *PLoS Biology*, 16.

Mueller, J. S., S. Melwani, and J. A. Goncalo (2012): "The Bias Against Creativity: Why People Desire but Reject Creative Ideas," *Psychological Science*, 23, 13–17.

Myers, K. (2020): "The Elasticity of Science," *American Economic Journal: Applied Economics*, 4, 103–134.

Nelson, R. R. (1959): "The Simple Economics of Basic Scientific Research," *Journal of Political Economy*, 67, 297–306.

——— (2016): "The Sciences Are Different and the Differences Matter," *Research Policy*, 45, 1692–1701.

Neumann, M., D. King, I. Beltagy, and W. Ammar (2019): "ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing," `https://github.com/allenai/`.

Onal Vural, M., L. Dahlander, and G. George (2013): "Collaborative Benefits and Coordination Costs: Learning and Capability Development in Science," *Strategic Entrepreneurship Journal*, 7, 122–137.

Peck, R. W., D. W. Lendrem, I. Grant, B. C. Lendrem, and J. D. Isaacs (2015): "Why Is It Hard to Terminate Failing Projects in Pharmaceutical R&D?" *Nature Reviews Drug Discovery*, 14, 663–664.

Peng, H., H. S. Qiu, H. B. Fosse, and B. Uzzi (2024): "Promotional Language and the Adoption of Innovative Ideas in Science," *Proceedings of the National Academy of Sciences*, 121, e2320066121.

Petersen, A. M., M. Riccaboni, H. E. Stanley, and F. Pammolli (2012): "Persistence and Uncertainty in the Academic Career," *Proceedings of the National Academy of Sciences*, 109, 5213–5218.

Price Nicholson, W. (2018): "Grants," *Berkeley Technology Law Journal*, 34.

Reagans, R., E. Miron-Spektor, and L. Argote (2016): "Knowledge Utilization, Coordination, and Team Performance," *Organization Science*, 1108–1124.

Rietzschel, E. F., B. A. Nijstad, and W. Stroebe (2010): "The Selection of Creative Ideas After Individual Idea Generation: Choosing Between Creativity and Impact," *British Journal of Psychology*, 101, 47–68.

Rivera, M. T., S. B. Soderstrom, and B. Uzzi (2010): "Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms," *Annual Review of Sociology*, 36, 91–115.

Rosenman, M. (1988): "Serendipity and Scientific Discovery," *The Journal of Creative Behavior*, 22, 132–138.

Rostamabadi, H., S. Reza Falsafi, M. Mahdi Rostamabadi, E. Assadpour, and S. Mahdi Jafari (2021): "Electrospraying as a Novel Process for the Synthesis of Particles/ Nanoparticles Loaded with Poorly Water-Soluble Bioactive Molecules," *Advances in Colloid and Interface Science*, 290.

Rotolo, D., M. Hopkins, and N. Grassano (2023): "Do Funding Sources Complement or Substitute? Examining the Impact of Cancer Research Publications," *Journal of the Association for Information Science and Technology*, 74, 50–66.

Simonton, D. K. (1999): *Origins of Genius: Darwinian Perspectives on Creativity*, Oxford University Press.

——— (2003): "Scientific Creativity as Constrained Stochastic Behavior: The Integration of Product, Person, and Process Perspectives," *Psychological Bulletin*, 129, 475–494.

——— (2021): "Scientific Creativity: Discovery and Invention as Combinatorial," *Frontiers in Psychology*, 12.

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

Slavova, K., A. Fosfuri, and J. O. De Castro (2016): "Learning by Hiring: The Effects of Scientists' Inbound Mobility on Research Performance in Academia," *Organization Science*, 27, 72–89.

Stephan, P. E. (1996): "The Economics of Science," *Journal of Economic Literature*, 34, 1199–1235.

——— (2015): *How Economics Shapes Science*, Harvard University Press.

Sunilkumar, P. and A. P. Shaji (2019): "A Survey on Semantic Similarity," in *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)*.

Tham, W. Y. (2023): "Science, Interrupted: Funding Delays Reduce Research Activity but Having More Grants Helps," *PLOS ONE*, 18, e0280576.

Thompson, N. C., A. A. Ziedonis, and D. C. Mowery (2018): "University Licensing and the Flow of Scientific Knowledge," *Research Policy*, 47, 1060–1069.

Uzzi, B. and J. Spiro (2005): "Collaboration and Creativity: The Small World Problem," *American Journal of Sociology*, 111, 447–504.

Van Den Besselaar, P. and C. Mom (2022): "The Effect of Writing Style on Success in Grant Applications," *Journal of Informetrics*, 16, 101257.

Vieira, E. S. and J. A. N. F. Gomes (2010): "Citations to Scientific Articles: Its Distribution and Dependence on the Article Features," *Journal of Informetrics*, 4, 1–13.

Walsh, J. P., W. M. Cohen, and C. Cho (2007): "Where Excludability Matters: Material Versus Intellectual Property in Academic Biomedical Research," *Research Policy*, 36, 1184–1203.

Waltman, L. (2016): "A Review of the Literature on Citation Impact Indicators," *Journal of Informetrics*, 10, 365–391.

Wang, J. and D. Hicks (2015): "Scientific Teams: Self-Assembly, Fluidness, and Interdependence," *Journal of Informetrics*, 9, 197–207.

Weisberg, R. W. (1998): "Creativity and Knowledge: A Challenge to Theories," in *Handbook of Creativity*, Cambridge University Press, 226–250.

Wooldridge, J. M. (2015): "Control Function Methods in Applied Econometrics," *The Journal of Human Resources*, 50, 420–445.

Wu, L., D. Wang, and J. A. Evans (2019): "Large Teams Develop and Small Teams Disrupt Science and Technology," *Nature*, 566, 378–382.

Yaqub, O. (2018): "Serendipity: Towards a Taxonomy and a Theory," *Research Policy*, 47, 169–179.

Yaqub, O., J. Coburn, and D. A. Q. Moore (2024): "Research-Targeting, Spillovers, and the Direction of Science: Evidence from HIV Research-Funding," *Research Policy*, 53, 105076.

Chapter 3

Interdisciplinary Advisors and PhD Students' Research[1]

[1]I am grateful for the comments and suggestions provided by participants at the PhD Workshop on the Economics of Innovation in Bordeaux on an earlier version of this work. I also thank Fan Qi for supporting the calculation of interdisciplinary indicators.

3.1 Introduction

Interdisciplinary[2] research centers aimed at promoting cross-field collaboration and scientific creativity have proliferated in contemporary universities (Leahey and Barringer, 2020; Jacobs and Frickel, 2009). While university administrators, funders, and policymakers continue to emphasize support for interdisciplinarity (Rylance, 2015; D'Este and Robinson-Garcia, 2023), prior studies often portray this mode of research as a challenging venture that creates both benefits and costs for individual scientists (Leahey, 2018; Rhoten and Parker, 2004). For example, interdisciplinary work can boost scientists' visibility at the expense of productivity (Leahey et al., 2017) and enhance the societal relevance of their research, while undermining their reputation among scholars (Fontana et al., 2022). Because interdisciplinary centers chiefly seek to stimulate research organization and reconfigure collaborations among faculty (Biancani et al., 2014; Boardman and Corley, 2008), these initiatives may also alter the environment underpinning academic training, presenting a unique set of opportunities and challenges for PhD students.

A growing body of literature attests to the crucial role of the research environment, particularly the mentorship provided by doctoral advisors, in successful academic training (Brostrom, 2019; Corsini et al., 2022; Wuestman et al., 2023). In this context, mentorship is primarily shaped by joint research activities involving advisors and students (Paglis et al., 2006; Roche, 2022). Advisors not only transfer essential skills and tacit knowledge to conduct research (Roche, 2022), but also supply social capital that affords long-lasting career opportunities (Pezzoni et al., 2012; Hottenrott and Lawson, 2015), thereby shaping a range of outcomes for their advisees. Empirical evidence demonstrates that advisors are key determinants of PhD students' research productivity (Corsini et al., 2022; Shibayama, 2019), post-graduation job placement (Rose and Shekhar, 2023), and creative performance (Wang and

[2]We follow the commonly-adopted broad definition of interdisciplinary research as the "integration of perspectives, information, data, techniques, tools, concepts, and/or theories from two or more disciplines (Leahey, 2018).

Shibayama, 2022; Ma et al., 2020).

While the benefits of effective mentorship are widely documented, other scholars also highlight that mentorship can have unintended or even problematic consequences for students' progress (Bargar and Mayo-Chamberlain, 1983; Hu et al., 2022; Tuma et al., 2021). Such adverse outcomes can arise when there are mismatched expectations, misdirection in students' research pursuits, or insufficient time to nurture a constructive relationship (Tuma et al., 2021). These insights align with foundational research on effective mentorship, which emphasizes that, beyond formal supervision, it entails professional development, adequate training, and psychosocial support (e.g., Anderson and Shannon, 1988). While it has been suggested that mentorship can be pivotal in successfully undertaking interdisciplinary research (Lyall and Meagher, 2012), the influence of advisors' engagement in interdisciplinary work on their PhD students' research remains unclear, a gap that this study seeks to address.[3]

We posit that advisors' engagement with an interdisciplinary center can create a fundamental tension for their PhD students. On the one hand, exposure to interdisciplinary research can expand students' collaboration with scientists from different areas of expertise. Doctoral students typically lack extensive ties to academic communities, making it difficult for them to initiate collaborations beyond their immediate research group (Wuestman et al., 2023). By lowering the search costs of identifying individuals with complementary expertise and shared interests (Boudreau et al., 2017), advisors who join interdisciplinary centers may help foster the formation of intellectually distant collaborations. Consistent with the knowledge recombination perspective, a more diverse set of collaborators can provide access to new information, fresh perspectives, and unique resources (Fleming et al., 2007; Uzzi and Spiro, 2005), thereby advancing creative performance. Moreover, a more extensive collaboration network can broaden the potential audience for their research, thereby generating more visibility and potentially leading

[3]In this study, we do not contrast interdisciplinary and monodisciplinary mentorship per se. Instead, we examine how advisors' engagement in interdisciplinary centers as a distinct organizational context may influence the research outcomes of their PhD students.

to a citation premium (Leahey et al., 2017; Vakili et al., 2021).

On the other hand, for scientific collaborations to effectively leverage researchers' distinct expertise, they require adequate communication procedures (Reagans et al., 2016) and a functional division of research tasks (Haeussler and Sauermann, 2020). These processes are time-intensive and cognitively demanding (Bercovitz and Feldman, 2011; Onal-Vural et al., 2013), and can be especially challenging in interdisciplinary teams, where differences in terminology, methods, and procedures tend to create friction (Cummings and Kiesler, 2005; Mansilla et al., 2016). Because engagement in interdisciplinary centers may demand additional time from advisors, this could potentially undermine the attention and support students receive. To the extent that such challenges arise during a critical stage of training, when students are honing their research skills and expertise, advisors' engagement in interdisciplinary work may negatively affect PhD students' productivity.

In this paper, we examine whether and how advisors who engage in interdisciplinary research centers influence the research outcomes of their PhD students, focusing on scientific productivity and novelty. Our empirical setting is the establishment of two interdisciplinary research centers at a large European university. We treat advisors' participation in these centers as a footprint of their engagement with interdisciplinary work during their students' training. We combine administrative records on advisors and PhD students with publication data, assembling a longitudinal dataset covering 1,036 advisors and 3,658 PhD graduates from 2007 to 2023. The analyses are conducted at the scientist–year level and focus on students' research output from PhD start up to two years after graduation. To examine the influence of advisors' interdisciplinary engagement, we estimate panel data models with advisor fixed effects, comparing students of the same advisor who graduated before and after the advisor joined an interdisciplinary center. This within-advisor design isolates variation attributable to the advisor's interdisciplinary engagement, holding constant advisor-specific mentorship style and the broader research environment. We also verify that there is no systematic sorting of select students into advisors' interdisciplinary engagement, lending

further credibility to our empirical approach.

Our findings are consistent with a trade-off in doctoral training under advisors who engage in interdisciplinary work. Students of interdisciplinary advisors publish fewer papers during their PhD, indicating a productivity penalty. At the same time, we uncover emerging evidence of gains in scientific novelty, as these students are more likely to introduce new scientific ideas, suggesting that exposure to interdisciplinarity can stimulate creative recombination. We also observe modest benefits in related dimensions: joint publications with advisors attract more citations, and students collaborate with coauthors whose expertise is more distant, indicating broader intellectual reach and visibility. Further analyses reveal heterogeneity across research orientations. The productivity penalty is concentrated among students in fundamental research fields, whereas no systematic differences emerge in translational settings. When extending the observation window to five years after graduation, we observe tentative signs of gains in productivity and novelty, although these likely reflect the selection dynamics of students who continue to publish up to five years after their PhD graduation. Finally, our results remain robust to alternative measures of advisors' interdisciplinary engagement.

Our study extends prior work on the trade-offs of interdisciplinary research in two key ways. While earlier work has primarily examined established scientists (e.g., Leahey et al., 2017; Biancani et al., 2018), we shift attention to doctoral students, exploring how advisors' interdisciplinary engagement shapes academic training at a formative career stage. This perspective reveals how the challenges and benefits of interdisciplinarity can unfold through advisors, influencing both research productivity and creative performance of their students. We broaden the notion of the trade-off itself: beyond the well-documented productivity–visibility tension (Leahey et al., 2017), we identify emerging gains in scientific novelty, as reflected in the introduction of new ideas (Arts et al., 2025). Finally, by showing that PhD students' research responds to interventions designed to foster collaboration among established researchers, with advisors serving as a key channel of

influence, we bridge research on the conditions that underpin successful academic training (Corsini et al., 2022; Graddy-Reed et al., 2021) and studies on the effects of research collaboration policies (Boudreau et al., 2017; Carayol et al., 2023).

3.2 Institutional Context

We study two interdisciplinary research centers rooted in the biomedical sciences at KU Leuven in Belgium: the Leuven Cancer Institute (LKI) and the Leuven Brain Institute (LBI). The biomedical sciences are particularly intensive in team knowledge production, with the apprenticeship model within the laboratory playing a central role (Conti and Liu, 2015; Stephan, 2015). While PhD students in these fields typically produce several publications before graduation (e.g., Freeman et al., 2001), the training system faces critical issues, including increasing exit rates from academic careers (Daniels, 2015; Alberts et al., 2014). This section outlines key features of our research context, including a characterization of the two centers.

KU Leuven is a large research-intensive university, is the highest-ranked institution in Belgium according to the Times Higher Education ranking, and has been consistently placed among the top 60 universities globally.[4] In recent years, promoting interdisciplinary research has been a core element of the university's strategic policies, with several interdisciplinary centers established under a common framework. Unlike other initiatives that rely on dedicated facilities and funding, the university policy seeks to foster collaboration among existing academic units by pooling their resources and expertise around specific topics and research lines (Jacobs and Frickel, 2009; Biancani et al., 2018). Concretely, the university administration offers official recognition in the organizational chart of the university, access to central administrative support (communications, ICT services, etc.), and modest incubation funding to cover a dedicated institute coordinator. As stated by the rector

[4] An overview of university rankings, including ARWU/Shanghai and others, can be found here. In 2024, KU Leuven ranked 1st for research budget awarded among higher education institutions in Horizon Europe.

at the time, "this is about a new, recognisable, inspiring way to organise important interdisciplinary scientific work, not about new buildings."[5]

Members of the interdisciplinary centers hold their primary appointments in university departments. Each center spans more than fifty departments and research units, drawing faculty from science and technology, the social sciences, and laboratories of the university hospital.[6] The affiliation to each of these centers is not automatic, nor mandatory for faculty members, and involves a light administrative procedure before it becomes recognized institutionally. For example, the LBI website emphasizes that prospective members should "express a genuine commitment to become an active member," and are invited to apply by sending a description of their current research projects, their fit, and how they expect to contribute to the center.[7]

The LKI aims to integrate cancer research, therapy, and care by enhancing the interactions between biomedical scientists, clinicians, nurses, and care workers. In terms of its origin, the LKI builds on decades of earlier efforts to foster cooperation among biomedical researchers working on cancer. While it was formally established as a center in 2006 with a focus on clinical activities, in 2015 it adopted an interdisciplinary model, integrating several university and hospital laboratories under a common strategic leadership. More recently, in 2020, it was recognized as one of the interdisciplinary centers officially supported by the university administration.[8] The LKI is organized around multiple research lines, covering fundamental research on cancer types and their mechanisms, to diagnosis, drug development, and experimental therapies, each involving multiple laboratories, and together they encompass over fifty distinct research units. While research projects and clinical trials take place within these research lines, a number of activities span multiple research lines, such as research symposia, thematic focus groups, and outreach events.

[5]KU Leuven (2019), press note available here

[6]KU Leuven is organized into three core group (Humanities and Social Sciences; Science, Engineering and Technology; and Biomedical Sciences) each of which spans multiple campuses and comprises more than two hundred academic departments and research units.

[7]As described in the LBI website, accessed July 2024 and available here.

[8]Based on the LKI website, available in this link.

The LBI seeks to advance research on the molecular basis of the brain, children's brain development, the prevention and treatment of brain diseases, and the development of brain research technologies. The LBI was established in 2018, gaining recognition from the university administration as an institutional interdisciplinary initiative shortly after. The members of the LBI include physicians, psychologists, physicists, engineers, and biologists, among other faculty members distributed across over forty academic units. The internal organization of the LBI relies on consortia blending laboratories and self-assembled teams who seek to develop joint research projects. As these consortia need to be endorsed by the center steering committee, they are expected to have regular meetings and commit to structural collaborations. In addition, the center organizes focus groups on cross-cutting topics of interest, work-in-progress seminars, and outreach events. We interpret that, compared to the LKI, the LBI is not only more recently established but also emerged in part as a response to contemporary demands for greater interdisciplinary collaboration within the university and beyond.[9] Moreover, while the LKI primarily relies on collaborations across existing research units and places its clinical mission at the forefront, the LBI promotes ad-hoc team structures and its clinical mission is more secondary.

Three features of this institutional context are conducive to our research objectives. First, the centers are comparable in size and research domain and both conform to the university's policy on interdisciplinary institutes. They are subject to the same evaluation criteria for the university-wide designation as an 'interdisciplinary research center' and report on comparable objectives on a four-year cycle. During periodic reviews, they are expected to document cross-departmental synergies, collaborations across research units, and the formation of new collaborative networks. Second, they recruit doctoral researchers from a comparable pool through equivalent procedures. University research units typically post project-based vacancies linked to specific advisors who have secured funding in advance. These advisors participate directly in recruitment and select among applicants who meet predefined

[9]Based on the LBI website, available in this link.

eligibility criteria, such as a master's degree with distinction, relevant laboratory or research skills, etc. Successful applicants are appointed to full-time, four-year positions, with doctoral training predominantly taking place in the advisor's academic unit. Third, because the centers' primary objective is to promote research collaboration, they do not run separate PhD training programs and instead influence graduate training indirectly. In this context, the advisors is the primary conduit through which students are exposed to interdisciplinary research.[10]

3.3 Data

We merge university records of PhD defenses with administrative funding and personnel data to create a researcher–year panel of 3,658 PhD graduates (2007–2023) linked to 1,036 advisors. Using each researcher's full name and organization affiliation details, we retrieve bibliometric data for these researchers and their coauthors from the Dimensions database of Digital Science (Herzog et al., 2020), and construct their publication histories from 2002 through 2024. We also collect metadata for each publication, including citation counts, article-level disciplinary classifications, and detailed information on every listed author, including their affiliation as reported in the publication and the curated organization identifiers.[11]

To identify advisors' membership in the research centers, we combine information on the center websites with affiliation information extracted from publication records. We focus on substantive interdisciplinary engagement, defined as the inclusion of the center in a publication's affiliation, which is

[10] In recent years, the centers have offered voluntary certifications for students who participate in center activities during their doctoral studies. While valuable, these schemes are peripheral to both the centers and students' activities, and are not considered in our analyses.

[11] The Global Research Identifier Database (GRID) ensures that we identify researchers who are affiliated with the KU Leuven. In complementary queries, we also capture all their publication records, irrespective of their KU Leuven affiliation (i.e., publications before or after being employed by the university). Furthermore, we exclude publications with 20 or more authors (equivalent to approx. the top 3% of retrieved records by author count) because the affiliation data are often imprecise, abbreviated, or merged across multiple individuals in these cases.

encouraged by both the centers and the university. Since the university's renewal policy emphasizes acknowledging center-enabled collaborations, but centers do not replace faculty departmental affiliations, and much faculty research may not be interdisciplinary, we do not expect researchers to mention the center in every publication. Our approach is therefore conservative and provides a lower bound on center engagement. Concretely, we search each publication's affiliation field using string matching and regular expressions to detect center names in English and Dutch, including acronyms and full names.[12] As of 2024, the LBI and LKI together listed 352 senior academic researchers on their websites. Of these, we identify 164 who both served as advisors and acknowledged the center in their publications during the study period, equivalent to 15.5% of all advisors in our dataset. These advisors supervised 544 PhD students (equivalent to 14.8% of graduates) and, on average, referenced the center in about 30% of their publications in a given year.

Advisors

We begin our descriptive analysis by comparing trends in research output for two groups of advisors: those who at any point joined an interdisciplinary center ("ID advisors") and those who did not ("non-ID advisors"). Figure 3.1 shows binned scatterplots with academic unit fixed effects to isolate within-acdemic unit variation in advisors' outcomes. In addition, we overlay a quadratic fit to visualize the trend by advisor type (Starr and Goldfarb, 2020). These trends are descriptive and intended to contextualize our analysis of PhD students' research. Consistent with prior work (Su, 2014), the plots indicate systematic differences across the two groups, especially in research productivity and collaboration.

In Figure 3.1a, ID advisors exceed non-ID advisors in coauthor-weighted publications, although the gap narrows over time: ID advisors productivity

[12]For example, we test *Leuven Cancer Institute*, *Leuven Kanker Instituut*, and *Kankerinstituut*, as well as *Leuven Brain Institute*, *Leuven Brein Instituut*, and *Breininstituut*. Because we always pair these strings with KU Leuven's organizational identifier (GRID-ID), the risk of false positives is negligible, which we confirm by manual checks.

appears to decelerate, while non-ID advisors follow a steady upward trajectory. Figure 3.1b reports average annual citations per publication; showing that ID advisors are consistently more cited, but the difference is relatively modest and stable.

Figure 3.1c plots collaborations with other center members, which is suggestive of engagement with interdisciplinary research. We observe that up to about 2018, ID and non-ID advisors collaborate at similar rates; in recent years, however, we see ID advisors collaborating more intensively with center-affiliated researchers. Finally, Figure 3.1d shows the number of new coauthors: ID advisors display a clear upward trend that outpaces the increase for non-ID advisors.

PhD Students

Table 3.1 reports the proportion of students whose advisor at some point engaged in interdisciplinary research across faculties. Medicine shows the highest proportion, with about one-third of students advised by an advisor that engages in interdisciplinary research, followed by Movement and Rehabilitation Sciences (25%) and Psychology and Education (18%). By contrast, Engineering, Pharmaceutical Sciences, and the Natural Sciences have much lower shares, below 5% each. Overall, about 15% of students in the sample are linked to an ID advisor.

In Table 3.2, we report summary statistics of PhD students and their advisors (both ID and non-ID Advisors) at the PhD entry year. Regarding student characteristics, we report the time to graduation (Phd Duration), the presence of a co-advisor, and a binary variable if the student reported any publications prior to starting the PhD. PhD students whose advisors ever joined an interdisciplinary center earned their PhD in a slightly longer time (4.24 vs. 4.09 years), they are more likely to have co-advisors (75% vs. 66%), and they are more likely to report publications before PhD entry (22% vs. 17%). We also report key characteristics of advisors at the time of PhD entry, including the number of PhD students they supervise, whether they hold the rank of full professor, their research productivity, the size of

their coauthor network, and the breadth of their research portfolio (further details on variables in the following section). Compared to non-ID advisors, ID advisors supervise fewer PhD students on average (1.80 vs. 2.14) and are more likely to be full professors (26% vs. 19%). Finally, consistent with the trends described above, ID advisors are on average more productive, collaborate with more coauthors, and conduct more diversified research.

Before turning to the main analyses, it is important to examine whether students systematically sort into advisors who are engaged in interdisciplinary research. Because advisors may join interdisciplinary centers at different stages of their careers, the composition of their advisees could change over time. To explore this possibility, we focus on advisors who eventually become affiliated with an interdisciplinary center and examine whether the characteristics of their PhD students differ before and after the first year of center affiliation.

Specifically, we test whether pre-doctoral attributes of students predict starting the PhD by their advisor's first year of center affiliation. If advisors selectively recruit students with stronger research experience or broader research interests once they engage in interdisciplinary work, our estimates of the effects of interdisciplinary exposure could partly reflect this sorting. We therefore analyze several proxies of students' pre-PhD research activity and orientation, measured within five years before PhD entry, while including fixed effects for the student's first publication year, cohort, academic unit, and advisor.

Panel (A) of Table 3.4 reports regression estimates for students' prior research experience and collaboration. Initially, we observe a positive, albeit weakly significant, association between having at least one pre-doctoral publication and beginning the PhD after the advisor's first affiliation year. This association, however, vanishes once advisor fixed effects are included. A similar pattern holds for the (log-transformed) number of pre-doctoral coauthors, suggesting that apparent differences are explained by advisor-specific characteristics rather than student sorting into center engagement. Panel (B) of Table 3.4 further examines students' pre-PhD research orientation,

captured through two indicators of interdisciplinarity: research breadth (see Variable Definitions section) and variety (the number of distinct disciplines represented in students' prior work). We find no significant relationship for either measure.

Overall, these results indicate that the composition of students remains relatively stable before and after advisors join an interdisciplinary center, particularly within advisors. The lack of evidence for systematic student sorting into advisors' interdisciplinary engagement suggests that later outcome differences are unlikely to be driven by selection effects, underscoring the importance of advisor fixed effects in isolating the role of advisors' interdisciplinary engagement on student outcomes.

3.4 Empirical Strategy

We examine whether advisors' engagement with interdisciplinary research centers is associated with their PhD students' research productivity and novelty, and if so, in which direction this association operates. Our unit of analysis is defined at the PhD student-year level, and we examine their outcomes from their year of entry to two years after graduation.[13] We combine panel fixed effects models with an entropy balancing procedure akin to matching (Hainmueller, 2011; Hainmueller and Xu, 2013), which we further describe below. Concretely, we estimate the following relationships:

$$Y_{ijt} = \beta d_{ij} + \theta A_{ij} + \gamma S_{it} + \mu_{(i)c} + \mu_{(i)k} + \lambda_t + e_{ijt} \tag{3.1}$$

$$Y_{ijt} = \beta d_{ij} + \alpha_j + \gamma S_{it} + \mu_{(i)c} + \mu_{(i)k} + \lambda_t + e_{ijt} \tag{3.2}$$

Our key variable of interest is d_{ij}, a binary variable equal to one if advisor j demonstrates engagement in interdisciplinary research within a five-year window around the start of student i's PhD (from five years before to five

[13]These additional years after graduation account for the time lag between research execution and publication (Bjork and Solomon, 2013). For example, a recent systematic review suggests that the average editorial handling times vary from a few months to two years across biomedical journals (Andersen et al., 2021). Our results are robust to using just one year of lag (results not reported).

years after PhD entry). By using a five-year window before and after PhD entry, we factor in that advisors' interdisciplinary engagement can influence students both through research conducted before the student starts her PhD and also through research in progress. Furthermore, the definition ties interdisciplinary exposure directly to the student's training period, aligning the measure with the developmental stage when mentorship and research environment should matter most.[14]

The crucial difference between the two equations is that, while in equation (1) we include advisor's j attributes measured at the time student i starts the PhD (A_{ij}), in equation (2) we replace it by advisor fixed effects (α_j), thus absorbing time-invariant advisor characteristics.[15] In other words, β in equation (2) captures the average difference in research outcomes (Y) between students of the same advisor, holding student attributes and other research environment characteristics constant, thereby allowing us to interpret the remaining variation more directly as reflecting differences in mentoring.

Specifically, we control for students' attributes (S_{it}), including characteristics at PhD entry (pre-doctoral publications and the presence of a co-advisor), and the scientific field(s) in which the student is active each year. We also include three sets of fixed effects: PhD entry cohort ($\mu_{(i)c}$) since students who start together may face similar training conditions; academic unit ($\mu_{(i)k}$) to absorb persistent features of the research environment such as infrastructure, culture, and organization; and year fixed effects (λ_t). Together, these variables capture the socialization process of PhD students

[14]In supplementary analyses, we also construct an intensity-based measure of advisors' interdisciplinary engagement, defined as the number of an advisor's publications that mention the center. We further experiment with alternative window lengths to test the sensitivity of our results and define an additional measure based on advisors' substantive interdisciplinary work. These analyses confirm the robustness of our findings when using the intensity-based measure, and the latter measure provides further confidence in the consistency of our results. All results are reported and discussed in Appendix B.2.

[15]To clarify, because A_{ij} captures advisors characteristics measured at the student's entry year, these values vary across students of the same advisor only if they begin their PhDs at different times (and, of course, across different advisors), but not within a given student. We can therefore not include both the advisor fixed effects (a_j) and A_{ij} in the same specification, as a_j would absorb all time-invariant advisor characteristics, making A_{ij} collinear and unidentified.

(Gardner, 2010), encompassing both formal and informal interactions with colleagues and faculty (e.g., through coursework or social activities) as well as the broader environment in which training occurs. Academic units and departments typically follow a disciplinary logic, transmitting shared scientific norms and practices to graduate students (Jacobs, 2013).

In addition, we control for several advisor attributes (A_{ij}) that may influence student outcomes and confound the effects of interdisciplinary exposure, including a proxy of their research ability, career stage, and collaboration network. Because our descriptive analyses reveal systematic differences between PhD students whose advisors engage in interdisciplinary work and those who do not, in all specifications, we use weights for the control group observations using entropy balancing. In doing so, we correct for the first three moments (mean, variance, skewness) so that the distribution of students' and advisors' attributes (i.e., S_{it} and A_{ij}) is as similar as possible across students with and without an ID advisor.

Entropy balancing offers two key advantages over conventional matching methods and has therefore gained popularity in observational studies (for example, Kaellberg and Waernbaum, 2023; Trunschke et al., 2024). First, unlike other techniques that require a common support and typically lead to excluding substantial portions of the data, entropy balancing techniques use all the observations in the sample to achieve comparable groups. Second, entropy balancing does not impose parametric assumptions about the determinants of treatment assignment; in our context, the (observable) student attributes that are associated with potential selection into ID advisors. We report all summary statistics at the student,year-level in Table 3.3, and the balance achieved across groups in Table 3.5 and Table 3.6.[16]

Outcome Variables

Novelty. Our focus on novelty connects to theories of knowledge recombination, which view creativity as arising from bridging otherwise disconnected ideas and accessing new perspectives and information (Simonton, 1999; Flem-

[16]Pairwise correlations are reported in the Appendix B, Table B.1.1.

ing et al., 2007). Interdisciplinary research may facilitate such recombination by fostering atypical combinations of knowledge domains, thereby increasing the likelihood of scientific discoveries (Uzzi et al., 2013; Fontana et al., 2020).

We measure scientific novelty at the time of publication using the text-based measures developed by Arts et al. (2025). These measures identify terminology first introduced in a focal publication and observed at least twice in the entire scientific literature (as recorded in the OpenAlex repository), capturing the scientific novelty of research at the time of publication and irrespective of subsequent citations.[17]

We aggregate these measures to the researcher–year level and construct four complementary variables: (i) a binary indicator equal to one if any publication in that year introduces at least one *new word*, (ii) a binary indicator for at least one *new phrase*, (iii) the average number of *new words* across all publications in that year, and (iv) the average number of *new phrases* across all publications in that year. The binary indicators capture whether a researcher produces any novel terminology in a given year, reflecting the extensive margin of novelty, while the average counts, which we log-transform to reduce skewness, capture the intensive margin, or how much novel terminology is introduced once novelty occurs. Together, these measures provide a complementary view of the frequency and depth of students' scientific novelty.

Following Arts et al. (2025), a *new word* is identified as a previously unseen term in the literature, while a *new phrase* corresponds to a novel multi-word expression, typically a noun phrase consisting of a central noun or pronoun (the head) and accompanying words (dependents) that refine or specify its meaning. Identifying new words is comparatively straightforward, as it involves detecting unique terms that have not appeared before, often corresponding to entirely new concepts or entities. A notable example is the first appearance of the term 'antiproton' in 1955, representing a fundamental advance in particle physics. In contrast, detecting new phrases requires

[17]We link our publication corpus to the open repository made available by Arts et al. (2025) using publication DOIs. The repository is available here, and it covers publications up until 2023.

identifying structured combinations of existing words that together form a coherent scientific idea extracted from text. For example, the 'polymerase chain reaction' first coined in 1986 and widely regarded as a methodological breakthrough. While new words signal the creation of novel concepts or terms, new phrases can capture more subtle, combinatorial forms of novelty that arise from integrating established concepts in new ways, closely aligning with theories of knowledge recombination in science (e.g., Uzzi et al., 2013).

Productivity. Our focus on productivity is motivated by prior research highlighting the trade-offs of interdisciplinary research. While interdisciplinary work can yield novel insights and increase visibility, it often comes at the cost of reduced research output (Leahey et al., 2017; Leahey, 2018). More broadly, the literature on the organization of scientific work emphasizes that collaboration choices frequently involve trade-offs between individual productivity and the potential for disproportionate credit or recognition (Bikard et al., 2015). In particular, participation in complex, interdisciplinary collaborations can depress short-term productivity due to heightened coordination and communication costs. Nonetheless, scientists may still pursue such collaborations either because they anticipate a "credit premium" if the project succeeds or because they are motivated by the opportunity to advance ambitious scientific goals (Bikard et al., 2015; Vakili et al., 2021).

We capture student productivity using five complementary measures: (i) total publication count; (ii) coauthor-weighted publication count, which divides each paper by the number of coauthors to account for team size; (iii) publications coauthored with the advisor, reflecting the intensity of mentoring during PhD training; (iv) average yearly citations per publication across the student's research output; and (v) average yearly citations per publication for the subset coauthored with the advisor. For (iv) and (v), we calculate, for each publication in year t, the cumulative citations accrued to date, divide by the number of years since t, and then average across the relevant set of publications.

Control Variables

We measure the following attributes of PhD students:

Pre-doctoral publications. Binary variable that equals one if the student has any publications prior to the PhD start. The presence of students' research skills and/or experience before their formal training may be an imperfect but useful signal of ability (Roche, 2022; Corsini et al., 2022).

Co-Advisors. Binary variable that equals one if the student has a co-advisor involved in his PhD. The co-promotors may share responsibilities in training, expand the sources of ideas and expertise that the student may draw on, and can also provide additional, access to other collaborators.

Scientific Field. We use a set of binary variables to capture the disciplines in which the researcher publishes in each year. We rely on the publication-level disciplinary classification obtained from Dimensions, which follows the Australian and New Zealand Standard Research Classification.[18] For parsimony, we use a compact version of the division-level fields of research, including Biological Sciences; Biomedical and Clinical Sciences; Chemical Sciences; Health Sciences; Engineering; Information and Computing Sciences; Psychology; and include a residual category for Other fields.

In addition, we measure the following characteristics of advisors:

Nr Advisees. We measure the number of students who begin their PhD under the same advisor in a given year, as a larger number of students may dilute the time and attention the advisor can devote to each student (e.g., Paglis et al., 2006).

Advisor Productivity and ***Advisor Coauthors***. Because more productive and better-connected advisors can provide richer training opportunities and resources (e.g., Corsini et al., 2022), We measure the citation-weighted number of publications and the number of unique coauthors, both within five years prior to PhD start.

Full Professor. Binary variable that equals one if the advisor is a full professor at the university by the PhD start year, and zero otherwise. This

[18] More details on the ANZRC system can be found here.

allows us to capture career-specific factors that may affect the engagement in interdisciplinary work and influence research outcomes. For instance, unlike junior colleagues, full professors face little career risk even if a research project fails (Franzoni and Rossi-Lamastra, 2017).

Advisor Research Breadth. We measure the knowledge breadth based on the distribution of the publication portfolio of the advisor across scientific fields (Teodoridis et al., 2019; Nagle and Teodoridis, 2020). Put differently, we capture the extent to which an advisor's expertise is concentrated in a few domains or spread across multiple areas. Formally, using the 171 group-level field categories of the ANZRC system (f), we define the breadth of a researcher i in year t in the following way:

$$ResearchBreadth_{it} = 1 - \sqrt{\sum_{f=1}^{171} \left(\frac{PublicationCount_{itf}}{PublicationCount_{it}} \right)^2} \tag{3.3}$$

The index lies within the range $[0, 1[$, as it never reaches one and is equal to zero when all publications of a given researcher in a year belong to the same field of research. Given that a single publication may be categorized in up to four group-level fields of research, we take a fractional approach so that a single publication that mentions two categories scores zero point five on each category. The interpretation of the research breadth is intuitive: the greater the breadth of knowledge (i.e., more diversified expertise), the closer to one the value of the index.

3.5 Results

3.5.1 Novelty

We present results on PhD students' research novelty in Table 3.7. For each of the outcome variables, we estimate equations (1) and (2) presented in the section 3.4, alternating between the specification that includes advisor attributes at PhD entry and the specification with advisor fixed effects. All estimated models include year, cohort, scientific field, and academic unit

fixed effects. While equation (1) captures cross-sectional correlations between advisor characteristics and student outcomes conditional on a comprehensive set of student-, advisor-, and research environment- controls, equation (2) improves our estimates by capturing within-advisor effects of engagement with interdisciplinary research. We implement the procedure developed by Guimaraes and Portugal (2010), which absorbs high-dimensional fixed effects to address redundancy or multicollinearity, and iteratively drops singleton observations.[19] Moreover, in all models we use the entropy balancing weights and we cluster standard errors at the advisor level to account for intra-cluster (within-advisor) correlation. Unless stated otherwise, we log-transform our dependent variables for ease of interpretation and comparability of estimated coefficients across models.

A notable contrast in our novelty results lies between the *new words* and the *new phrases*. Focusing on the coefficient of interest, advisors' engagement with an interdisciplinary center (d_{ij}), we find no robust evidence of an association with the introduction of *new words*. Concretely, in Table 3.7, Columns (1)–(2) and (5)–(6) show that d_{ij} is not significantly related to either the likelihood of introducing a new word or the average number of *new words*. By contrast, Columns (3)–(4) indicate a positive association between d_{ij} and the likelihood of introducing a new phrase, although this effect becomes only weakly significant once advisor fixed effects are included. Moreover, in Columns (7)–(8), we find a positive and statistically significant relationship between d_{ij} and the average number of new phrases, with the association becoming even stronger when accounting for advisor fixed effects. This pattern is consistent with the baseline frequencies reported by Arts et al. (2025), who show that only 7% of publications in their corpus introduce a new word, compared to 25% introducing a new phrase. In our sample, at the student-year level, the introduction of new phrases is also more than three times as common as the introduction of new words. Taken together, these results suggest that while the introduction of new words represents a rarer and arguably more fundamental form of novelty, and we find no effect, the in-

[19] This explains the discrepancy in the number of observations reported across model estimates, even when the outcome variable is the same.

troduction of new phrases is more prevalent and shows a positive relationship with advisors' interdisciplinary engagement.

3.5.2 Productivity

Turning to research productivity, we follow the same structure discussed above and present our results in Table 3.8. Table 3.8, Columns (1)–(4), shows a positive correlation between d_{ij} and research productivity (as measured by publications count, and less so when coauthor-weighted) in the baseline model. However, once we include advisor fixed effects, the coefficient becomes negative and statistically significant. This suggests that, while students of ID advisors appear slightly more productive in the cross-section, within-advisor estimates indicate that students are significantly less productive when their advisors engage in interdisciplinary research. In Columns (5)-(6), we find no statistically significant relationship for the number of publications coauthored with the advisor, suggesting the effect is not driven by mentorship intensity as proxied by advisor-student coauthorships. Turning to citations, Columns (7)–(8) show no statistically significant relationship for students' average yearly citations when considering their full output. However, Column (10) reveals a positive and statistically significant within-advisor association between interdisciplinary engagement and students' average citations when we restrict to advisor–student publications. This citation premium suggests that, although interdisciplinary work may come at the cost of lower research productivity, the projects on which students collaborate with their advisors can reach broader audiences and attract greater visibility, consistent with the idea of a trade-off between productivity and prominence Leahey et al. (2017).[20]

Taken together, these findings suggest a potential trade-off in doctoral training under interdisciplinary advisors. On the one hand, we find some evidence that students of interdisciplinary advisors are more likely to introduce new phrases, and that their joint work with advisors tends to receive

[20]In the Appendix B, Table B.2.9, we report estimates from Poisson fixed-effects regressions using raw counts as the dependent variable. Our results on students' publication productivity remain robust.

a citation premium, indicating gains in more incremental forms of scientific novelty and visibility. On the other hand, we observe a productivity penalty, with students publishing less overall when their advisors are engaged in interdisciplinary work.

3.5.3 Additional Analyses

Heterogeneity by Research Orientation

To examine heterogeneity across different research orientations, we classify each academic unit into three mutually exclusive categories: (i) *Fundamental/Basic*, (ii) *Translational/Clinical*, and (iii) *Hybrid*. The classification is based primarily on the unit's name and the description of its research activities on the university website.[21] We report a summary of the categorization of each unit into one of the three research orientations, and the names of units by each category in Appendix B, Table B.4.1 and Table B.4.2, respectively.

To illustrate, we provide one example for each category. The 'Adaptive Immunology Laboratory' exemplifies the *Fundamental/Basic* orientation, focusing on the molecular and genetic mechanisms underlying immune systems, core biological processes that can help us understand immune disorders. The 'GastroIntestinal Disorders' unit represents the *Translational/Clinical* orientation, explicitly combining pre-clinical and clinical expertise to investigate physiological and pathophysiological mechanisms and to identify new therapeutic targets for gastrointestinal diseases. Lastly, the 'Development and Regeneration' unit illustrates the *Hybrid* orientation, as they conduct research on the cellular and molecular mechanisms of tissue formation and repair while pursuing regenerative strategies for clinical application. While we acknowledge these distinctions are not always clear-cut, the primary mission and focus of each academic unit provide a reasonable first approximation of the research orientation that shapes doctoral training in our context.

We revisit our main estimates of PhD students' novelty and productivity

[21] The initial categorization was generated using OpenAI's ChatGPT (GPT-5, 2025 release), and each classification was subsequently validated by the author, including a review of the units' official KU Leuven webpages in case of doubt.

using specifications with advisor fixed effects and the number of publications by the advisor that mention the center within five years before the student's PhD start (*Advisor ID engagement*). Tables 3.9 and 3.10 report results by research orientation in three panels: Panel A, *Translational/Clinical*; Panel B, *Fundamental/Basic*; and Panel C, *Hybrid*. Focusing on the key regressor advisor's engagement with the interdisciplinary center, we observe substantial heterogeneity. In *Translational/Clinical* units (Panel A), effects on productivity and novelty are mostly insignificant, suggesting that an advisor's center affiliation neither systematically boosts nor depresses student outcomes, consistent with advisors already operating within structurally interdisciplinary agendas. In *Fundamental/Basic* units (Panel B), coefficients are consistently negative and statistically significant for productivity (especially citation-weighted publications and citations, including advisor-linked citations). We also find negative and significant effects for students' novelty (notably new phrases), indicating that advisors' interdisciplinary engagement is mainly adverse. We interpret this result as consistent with higher coordination costs when aiming to push boundaries within a field, where deeper specialization may be required (e.g., Jones (2009)). In *Hybrid* units (Panel C), novelty effects are insignificant, while productivity is mixed: publication output declines, but advisor-linked citations show a positive albeit weakly significant association. Overall, our heterogeneity analyses reveal that the influence of an advisor's interdisciplinarity engagement is most substantial and adverse in *Fundamental/Basic* research units. Moreover, it is worth noting that in this exercise, we do not uncover a systematic relationship between research novelty and research orientation, suggesting further work is needed to pinpoint contexts where novelty gains are more pronounced.

Extended Window of Analysis

We extend the observation window to include up to five years after PhD graduation. A longer horizon provides a more comprehensive view of students' research output, potentially capturing projects that, due to their complexity

or interdisciplinary nature, required more time to complete and publish.[22] However, this extended window also introduces limitations. In particular, students who remain in academia are more likely to be observed over a longer period in our panel, and they likely form a selected group that tends to be higher-performing and may have a distinct "taste for science" (Stephan, 2015). Furthermore, comparisons across cohorts become mechanically less meaningful, as more recent graduates have had less time to accumulate publications. To mitigate this imbalance, we limit the extended-window analysis to students who graduated by 2020, which represents 53% of all PhD students in our sample.[23]

We re-estimate our main models of PhD students' novelty and productivity using this restricted subsample, with detailed results reported in Appendix B (Tables B.2.10 and B.2.11). Regarding novelty, we find additional support for a positive relationship between advisor interdisciplinary engagement and the likelihood of introducing *new words*, while the association with the average number of *new words* is also positive but becomes statistically insignificant once advisor fixed effects are included. In contrast, we find no statistically significant association with *new phrases*. These results complement our baseline findings on novelty as measured by *new phrases*: when considering students' outcomes over a longer horizon, advisor engagement appears to be linked primarily to the introduction of *new words*—a rarer form of novelty—which can result from a complex and prolonged research process.

Turning to research productivity, the results for this subsample differ

[22] A related concern is that the limited post-graduation window in our main analysis may primarily capture the continuation of the advisor's research agenda rather than the student's independent work. On average, about 70% of students' yearly publications are coauthored with their advisor, and students appear as corresponding authors in only around 20% of cases (Appendix B, Figure B.2.2). These shares slowly decline after graduation. Complementary analyses show that students of more interdisciplinary advisors are slightly more likely to publish as corresponding authors, suggesting greater research independence; however, the citation premium reported in our baseline results does not extend to these publications, indicating that it may be driven by the advisor's visibility and networks (Appendix B, Table B.2.12).

[23] Appendix Figure B.2.1 shows the distribution of observations by graduation year, illustrating a sharp decline in observations beginning two years post-graduation.

markedly from our baseline findings. We find consistently positive and statistically significant effects on publication output across all three measures of productivity, while there is no evidence of a citation premium. This suggests that interdisciplinary engagement may promote greater publication quantity over a more extended period, even if these outputs do not immediately translate into higher-impact publications. Over time, as students' research roles become more autonomous from those of their advisors, the initial citation premium may dissipate, reflecting a gradual shift from advisor-driven visibility to the slower accumulation of independent impact.

These extended-window results should be interpreted with caution, nonetheless. Because they apply only to graduates who continue publishing up to five years after their PhD graduation, they may partly reflect attrition bias; students who persist in academic publishing likely differ systematically from those who exit earlier. Moreover, later publication outcomes may capture different stages of career development, as some students transition into research roles with increased responsibilities, such as postdoctoral positions. Taken together, the findings from this subsample underscore the need for further research on how time horizons shape the trade-offs we uncover in our baseline result.

3.5.4 Exploring Mechanisms

Having established that advisors' interdisciplinary engagement is associated with varying outcomes in students' productivity and novelty, it is important to explore why these effects occur. In particular, we aim to understand whether the observed trade-offs reflect differences in collaboration structures or in the intellectual scope of students' research. In what follows, we elaborate and examine two potential mechanisms that may underlie this trade-off: (1) whether interdisciplinary advisors enable distinct research collaboration patterns, and (2) whether students' research becomes more interdisciplinary.

Research Collaboration

One way an advisor's engagement with an interdisciplinary research center can affect PhD students is by to facilitate the formation of new collaborations, and possibly collaborations with researchers whose expertise is more distant (Boudreau et al., 2017). To understand if PhD students whose advisors engage in interdisciplinary work are associated with distinct patterns of research collaborations, we revisit our main specifications using a different set of outcome variables. Namely, we test the relationship between d_{ij} and four research collaboration variables: (i) a binary indicator that equals one if the student collaborates with another member of an interdisciplinary center in a given year, capturing the extent to which students connect with members of center; (ii) the number of new coauthors by year, reflecting students' ability to expand their collaboration networks; (iii) the average distance in collaborators' expertise, which captures the extent to which students engage with cognitively distant researchers, consistent with the knowledge recombination process; and (iv) the standard deviation of collaborator distance. The value of (iii) and (iv) lies in unpacking the differences between students who, on average, work with more distant collaborators and those who participate in more heterogeneous teams that combine both nearby and far expertise.

To measure distance in collaborators' expertise, we compute, for each pair of researchers who collaborate in a given publication year, the Jaccard distance between their sets of active research fields over the five preceding years.[24] This measure captures differences in their research knowledge bases. The distance ranges from 0, indicating identical prior field expertise, to 1, indicating no overlap at all. We then aggregate these distances for each student by year level by calculating the mean and standard deviation across each student's collaborations.

We present results on PhD students' research collaboration patterns in Table 3.11, following the same analytical approach as in our main results. Columns (1)–(2) show that students with ID advisors are more likely to collaborate with ID center members, though this relationship disappears once

[24]This leads to the exclusion of collaborators with no research activity in that window.

advisor fixed effects are included. Turning to Columns (3)–(4), we find that students of ID advisors appear to form more new coauthor relationships in cross-sectional analysis, but the within-advisor estimates reveal a negative and statistically significant effect. One possible interpretation is that once advisors join an interdisciplinary center, they may become "spread too thin" across collaborations, leaving them less able to facilitate new collaborative ties for their advisees than they were before. Regarding collaborator expertise, Column (6) demonstrates a positive and significant association between ID advisors and the average expertise distance of students' collaborators in the fixed-effects specification, indicating that advisees of ID advisors engage with more diverse expertise. By contrast, we do not find evidence that these students join more heterogeneous teams, as the standard deviation of collaborator distance is unaffected. Taken together, these within-advisor results suggest that while interdisciplinary advisors may open doors to more diverse expertise, they may also constrain the formation of new coauthor relationships, possibly reflecting a reallocation of research efforts when advisors take on interdisciplinary commitments.

Interdisciplinarity

We now turn to explore whether advisors' engagement with interdisciplinary centers can mitigate collaboration costs. Prior work on interdisciplinarity disentangles three components (Stirling, 2007; Wang et al., 2015): *variety*, *disparity*, and *balance*. *Variety* is defined as the number of distinct disciplines represented in a publication; *Balance* is how evenly different disciplines are referenced in a publication. *Disparity* captures how far apart those disciplines are from each other based on cognitive distance (e.g., how frequently disciplines cite each other in the past).[25] To understand the relationship between advisors' engagement and students research interdisciplinarity, we follow Leahey et al. (2017) and aggregate the publication-level interdisciplinary measures to the researcher,year-level by taking the average across

[25]We provide technical details on the construction of these variables in the Appendix B.3

each researcher set of publications in a given year.

Because the centers in our study are organized around specific research themes, we expect that advisors engaged with the centers promote work anchored in those topics. Rather than pushing students' researchers into more fields, we expect them to conduct research through a broader range of approaches and/or methods. Accordingly, we anticipate that students' research may display greater *disparity*, though not necessarily greater *variety*. By the same logic, we also anticipate a higher *balance*, as centers broaden their knowledge base by fostering cross-fertilization and integrating insights from different traditions focused on similar topics.

We present results on PhD students' research interdisciplinarity in Table 3.12. In Columns (1)–(2), we find a negative association between students' research *variety* and advisors' engagement with interdisciplinary work, which becomes statistically significant only in the within-advisor models. For *disparity* and *balance*, we do not uncover statistically significant relationships. Put differently, we find little evidence that students of ID advisors are better positioned to produce more interdisciplinary research; if anything, their work appears somewhat more specialized. Taken together, these results suggest that while students may gain access to new collaborators through their advisors' involvement in the center, this does not necessarily translate into a broader interdisciplinary approach in their own research.

3.6 Discussion and Conclusion

We examine whether and how advisors who engage in interdisciplinary research influence the research outcomes of their PhD students, a question that remains largely underexplored. This question is important because prior work highlights the central role of advisors and the broader research environment in shaping successful academic training, while interdisciplinary centers are specifically designed to foster new collaborations across otherwise disconnected faculty, potentially altering the mentorship process and the opportunities available to PhD students.

Our analysis focuses on an initiative that established interdisciplinary research centers within a large European university, allowing us to hold constant institutional and environmental factors while isolating the role of mentorship in shaping how the opportunities and challenges of interdisciplinary work affect students. Using detailed publications and longitudinal data linking administrative records of advisor and their PhD students, we exploit variation in advisors' engagement with interdisciplinary centers to estimate their effects on PhD students' research productivity and novelty.

Consistent with prior work on interdisciplinary research (e.g., Leahey et al., 2017), our findings reveal a clear trade-off. Students whose advisors participate in interdisciplinary research centers publish fewer papers during their PhD, indicating a productivity penalty. At the same time, we uncover emerging evidence of gains in scientific novelty, as these students are somewhat more likely to introduce new ideas in their work, suggesting that exposure to interdisciplinarity may stimulate creative scientific contributions. Moreover, we find moderate benefits in other dimensions: joint publications with advisors attract more citations, and students collaborate with coauthors whose expertise is more distant, indicating broader intellectual reach and visibility.

Additional analyses reveal heterogeneity by research orientation. The productivity penalty associated with advisors' interdisciplinary engagement is more pronounced among students in fundamental research environments, where pushing boundaries through interdisciplinarity often requires reconcil-

ing divergent theoretical perspectives and methods, processes that are time-intensive and cognitively demanding. By contrast, we find no systematic differences in translational research settings, where interdisciplinary collaboration may be more structured around targeted problems, clinical trials, and a clearer division of tasks. Extending the observation window to five years after graduation, we observe tentative signs of delayed positive outcomes, as some students appear to realize gains in productivity and novelty over time. However, these longer-term improvements likely reflect selection dynamics, as those who remain actively publishing for a more extended period constitute a non-random group likely better positioned to benefit from interdisciplinary exposure.

Our study extends prior evidence on the trade-offs associated with interdisciplinary research in two key ways. First, while existing work has largely examined established researchers (e.g., Leahey et al., 2017; Biancani et al., 2018), we shift the focus to doctoral students, investigating the ramifications of advisors' interdisciplinary engagement for academic training. This perspective highlights how the challenges and benefits of interdisciplinarity cascade down to the formative stages of research careers, when mentorship by doctoral advisors is particularly influential. Second, we expand the conceptual scope of the trade-off itself. Prior research has emphasized a productivity–visibility tension, whereby interdisciplinary work enhances exposure at the expense of output. We confirm this productivity penalty but also identify potential gains in scientific novelty, as measured by the introduction of new scientific ideas (Arts et al., 2025), suggesting that interdisciplinarity may stimulate creative recombination even as it slows publication.

In doing so, we contribute to the literature on the determinants of PhD student performance, an issue of growing importance as doctoral researchers represent the scientific workforce of the future and face increasingly competitive academic career prospects (Corsini et al., 2022; Daniels, 2015). The productivity penalty we document may therefore have non-trivial consequences for those who aspire to academic careers, particularly if they engage in interdisciplinary work prematurely or without sufficient support to balance its

inherent trade-offs. We also bridge two related strands of research: studies on how micro-level interventions can foster collaboration among established scientists (Boudreau et al., 2017; Carayol et al., 2023), and work examining the organizational role of interdisciplinary centers (Biancani et al., 2018). By focusing on PhD students, we connect these literatures and demonstrate how advisors' interdisciplinary engagement can shape the trajectories of early-stage researchers, by influencing the collaborations they form and the mentoring relationships they experience.

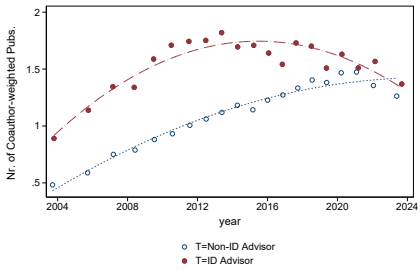
Naturally, our study has limitations. The institutional specificity of our setting may constrain the generalizability of results to other similarly sized, research-intensive European universities. Our data also do not allow for systematic assessment of long-term career consequences. Future research could investigate how specific design features of interdisciplinary centers, such as their internal organization, leadership structure, or the roles researchers assume, mediate the observed trade-offs. Qualitative work could provide further insight into how advisors and students navigate interdisciplinary collaboration in practice, revealing the mechanisms through which such experiences shape doctoral training. Finally, extending the analysis to post-graduation outcomes could clarify whether students of interdisciplinary advisors experience differential career advancement, satisfaction, or scientific impact.

Taken together, this chapter suggests that exposure to interdisciplinarity through advisors can simultaneously broaden doctoral students' intellectual horizons and introduce additional challenges during their training. This tension raises important questions for universities and science policy. Institutions may therefore wish to establish clearer guidelines and provide mentoring support to help advisors and students navigate these trade-offs, ensuring that interdisciplinarity enhances rather than hinders the development of the next generation of scholars.

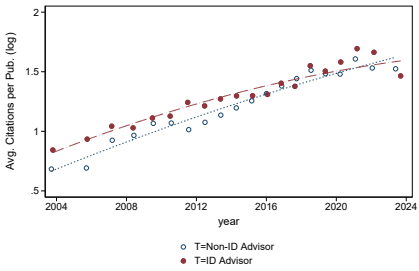
3.7 Tables and Figures

Figure 3.1: Advisor's Research Output by Year with Academic Unit FE

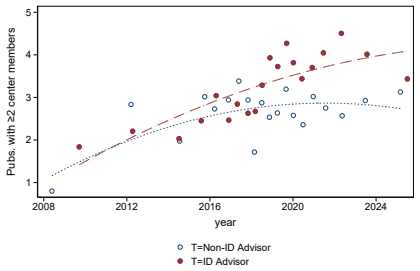
(a) Coauthor-weighted publications



(b) Avg. citations per pub.



(c) Publications with center members



(d) New Coauthors

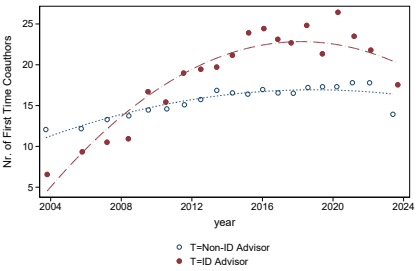


Table 3.1: Share of Students with ID Advisor by Faculty (student-level)

	ID Advisor		
Faculty	mean	(sd)	N
Engineering Science	0.03	(0.17)	799
Medicine	0.34	(0.47)	1,149
Movement and Rehabilitation Sci.	0.25	(0.43)	134
Pharmaceutical Sci.	0.05	(0.21)	120
Psychology and Education	0.18	(0.38)	331
Natural Sciences	0.04	(0.19)	587
Total	0.15	(0.35)	3,658

Table 3.2: Descriptive statistics by Advisor Type (student-level)

	ID Advisor $N = 542$		non-ID Advisor $N = 3,116$		
Variable	mean	(sd)	mean	(sd)	t-val
Student					
PhD Duration	4.24	(0.81)	4.09	(0.50)	5.88
Co-Advisor (0/1)	0.75	(0.43)	0.66	(0.47)	4.27
Pre-doc. Pubs. (0/1)	0.22	(0.42)	0.17	(0.37)	3.03
Advisor					
Nr. Advisees	1.80	(1.07)	2.14	(1.50)	-5.05
Full prof. (0/1)	0.26	(0.44)	0.19	(0.39)	4.28
Advisor Productivity	62.6	(46.2)	47.6	(43.5)	7.36
Advisor Coauthors	30.5	(28.3)	15.5	(14.9)	18.4
Advisor Research Breadth	0.40	(0.12)	0.32	(0.16)	11.0

Table 3.3: Summary statistics (student,year-level)

Panel A: Independent Variables

Variable	Mean	SD	Min	Max	N
d_{ij} (0/1)	0.10	0.31	0.00	1.00	19,750
Pre-doctoral Pubs (0/1)	0.21	0.40	0.00	1.00	19,750
Co-Advisor (0/1)	0.69	0.46	0.00	1.00	19,750
Nr. Advisees (log)	0.67	0.52	0.18	2.42	19,750
Advisor Productivity (log)	3.38	1.50	-1.61	6.33	19,750
Advisor Coauthors (log)	2.33	1.32	-1.61	5.11	19,750
Full Prof. (0/1)	0.22	0.41	0.00	1.00	19,750
Advisor Research Breadth	0.32	0.16	0.00	0.66	19,750
PhD Cohort	2014	4.26	2004	2019	19,750
Year	2018	4.20	2004	2024	19,750
Biology Sci.	0.19	0.40	0.00	1.00	19,750
Biomedical Sci.	0.40	0.49	0.00	1.00	19,750
Chemical Sci.	0.12	0.32	0.00	1.00	19,750
Health Sci.	0.27	0.45	0.00	1.00	19,750
Engineering	0.13	0.34	0.00	1.00	19,750
Computing Sci.	0.10	0.30	0.00	1.00	19,750
Pshychology	0.10	0.31	0.00	1.00	19,750
Other fields	0.34	0.47	0.00	1.00	19,750

Panel B: Outcome Variables

Variable	Mean	SD	Min	Max	N
Publications count	2.78	2.39	1	15	19,750
Coauthor-wt. Pubs.	0.53	0.50	0.01	6	19,750
Pubs. w. Advisor	1.57	1.71	0	15	19,750
Avg. Citations per Pub.	5.76	11.67	0	731	19,750
Avg. Citations per Pub. w. Adivsor	5.53	11.25	0	583	14,373
New Words (0/1)	0.14	0.35	0	1	11,231
New Phrases (0/1)	0.45	0.50	0	1	11,231
Avg. New Words	0.08	0.26	0	6	11,231
Avg. New Phrases	0.36	0.58	0	14	11,231

Table 3.4: Sorting into Advisors' Interdisciplinary Engagement

	DV: 1 if PhD entry year ≥ advisor's first center affiliation year; 0 otherwise ($mean$=0.30; $s.d.$=0.45)							
Panel (A)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pre-doctoral Publication (0/1)	0.107*	0.099*	0.090**	0.012				
	[0.055]	[0.053]	[0.041]	[0.039]				
Pre-doctoral Coauthors (log)					0.036**	0.033**	0.023*	0.001
					[0.017]	[0.016]	[0.012]	[0.012]
Constant	0.277***	0.278***	0.285***	0.310***	0.333***	0.330***	0.325***	0.313***
	[0.042]	[0.043]	[0.017]	[0.008]	[0.039]	[0.040]	[0.021]	[0.011]
N	543	543	526	484	543	543	526	484
R^2	0.009	0.127	0.620	0.778	0.011	0.129	0.619	0.778
Panel (B)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pre-doctoral Research Breadth	0.192	0.207	0.185	-0.136				
	[0.185]	[0.188]	[0.159]	[0.131]				
Pre-doctoral Research Variety					0.003	-0.004	0.006	0.001
					[0.009]	[0.009]	[0.006]	[0.007]
Constant	0.292***	0.292***	0.297***	0.317***	0.298***	0.304***	0.299***	0.311***
	[0.041]	[0.041]	[0.017]	[0.005]	[0.042]	[0.043]	[0.017]	[0.006]
N	543	543	526	484	543	543	526	484
R^2	0.002	0.122	0.617	0.779	0.000	0.120	0.616	0.778
First Publication Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE		Y	Y	Y		Y	Y	Y
Academic Unit FE			Y	Y			Y	Y
Advisor FE				Y				Y

Notes: Student-level fixed effect regressions. Sample includes students of the 164 advisors who eventually become affiliated with an ID center. Robust standard errors clustered at the advisor level in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5: Covariate Balance Before Weighting

	$d_{ij} = 1$			$d_{ij} = 0$		
Variable	Mean	Variance	Skewness	Mean	Variance	Skewness
Pre-doctoral Pubs (0/1)	0.29	0.20	0.95	0.21	0.16	1.45
Co-Advisor (0/1)	0.76	0.18	-1.22	0.67	0.22	-0.74
Nr. Advisees (log)	0.58	0.21	0.67	0.69	0.28	0.66
Advisor Productivity (log)	3.88	0.33	-1.63	3.33	2.24	-2.25
Advisor Coauthors (log)	3.13	1.22	-1.10	2.26	1.69	-1.25
Full Prof. (0/1)	0.23	0.18	1.26	0.20	0.16	1.52
Advisor Research Breadth	0.39	0.01	-0.56	0.32	0.03	-0.45
Year	2019	13.80	-1.23	2018	18.62	-0.88
PhD Entry Year	2016	11.70	-1.56	2015	17.12	-1.00
Biology Sci.	0.24	0.18	1.24	0.18	0.15	1.65
Biomedical Sci.	0.82	0.15	-1.65	0.33	0.22	0.71
Chemical Sci	0.03	0.03	5.82	0.13	0.11	2.20
Health Sci.	0.06	0.05	3.82	0.30	0.21	0.86
Engineering	0.16	0.14	1.82	0.11	0.10	2.45
Computing Sci.	0.05	0.04	4.28	0.11	0.09	2.57
Pshychology	0.24	0.18	1.20	0.08	0.08	3.01
Other fields	0.10	0.09	2.68	0.35	0.23	0.61

Table 3.6: Covariate Balance After Weighting

	$d_{ij} = 1$			$d_{ij} = 0$ (Weighted)		
Variable	Mean	Variance	Skewness	Mean	Variance	Skewness
Pre-doctoral Pubs (0/1)	0.29	0.20	0.95	0.29	0.20	0.95
Co-Advisor (0/1)	0.76	0.18	-1.22	0.76	0.18	-1.22
Nr. Advisees (log)	0.58	0.21	0.67	0.58	0.21	0.67
Advisor Productivity (log)	3.88	0.33	-1.63	3.88	0.33	-1.65
Advisor Coauthors (log)	3.13	1.22	-1.10	3.13	1.22	-1.10
Full Prof. (0/1)	0.23	0.18	1.26	0.23	0.18	1.26
Advisor Research Breadth	0.39	0.01	-0.56	0.39	0.01	-0.56
Year	2019	13.80	-1.23	2019	13.77	-1.08
PhD Entry Year	2016	11.70	-1.56	2016	11.66	-1.40
Biology Sci.	0.24	0.18	1.24	0.24	0.18	1.24
Biomedical Sci.	0.82	0.15	-1.65	0.82	0.15	-1.65
Chemical Sci	0.03	0.03	5.82	0.03	0.03	5.82
Health Sci.	0.06	0.05	3.82	0.06	0.05	3.82
Engineering	0.16	0.14	1.82	0.16	0.14	1.81
Computing Sci.	0.05	0.04	4.28	0.05	0.04	4.28
Pshychology	0.24	0.18	1.20	0.24	0.18	1.20
Other fields	0.10	0.09	2.68	0.10	0.09	2.68

Table 3.7: PhD Students' Research Novelty

	Binary				Average			
	New Words		New Phrases		New Words		New Phrases	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
d_{ij} (0/1)	0.026	0.024	0.098***	0.118*	-0.008	0.062	0.100**	0.444***
	[0.023]	[0.084]	[0.032]	[0.072]	[0.024]	[0.095]	[0.045]	[0.162]
Pre-doc. Pub. (0/1)	0.026	0.001	0.035	0.015	-0.011	-0.020	-0.008	-0.023
	[0.016]	[0.017]	[0.025]	[0.037]	[0.019]	[0.022]	[0.037]	[0.056]
Co-Advisor (0/1)	-0.009	-0.013	0.037	0.023	-0.011	-0.019	0.054	0.037
	[0.020]	[0.022]	[0.030]	[0.030]	[0.027]	[0.034]	[0.045]	[0.050]
Nr. Advisees (log)	-0.027		-0.023		-0.040*		-0.029	
	[0.016]		[0.024]		[0.021]		[0.039]	
Advisor Productivity (log)	0.010		-0.054**		0.019		-0.058*	
	[0.018]		[0.023]		[0.023]		[0.035]	
Advisor Coauthors (log)	-0.005		-0.015		-0.005		-0.041**	
	[0.009]		[0.012]		[0.011]		[0.018]	
Full Prof. (0/1)	0.025		0.030		0.016		0.035	
	[0.024]		[0.026]		[0.027]		[0.040]	
Advisor Research Breadth	-0.103		-0.103		-0.053		-0.018	
	[0.079]		[0.100]		[0.090]		[0.152]	
Constant	0.082	0.080*	0.455***	0.167***	-1.491***	-1.496***	-0.851***	-1.376***
	[0.083]	[0.049]	[0.113]	[0.051]	[0.105]	[0.060]	[0.168]	[0.101]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y		Y
N	9304	9237	9304	9237	9304	9237	9304	9237
adj. R^2	0.120	0.160	0.125	0.179	0.084	0.103	0.102	0.144

Notes: Fixed effect regressions using entropy weights. Robust standard errors clustered at the advisor level.
d_{ij}=1 if advisor demonstrates engagement in interdisciplinary research within a 5-year window around PhD entry
Cols.(5)-(8) dep.vars.in logs$((x + 0.2))$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8: PhD Students' Research Productivity

	Publications						Avg. Citations per Pub.			
	Simple count		Coauthor weighted		with Advisor		All pubs		with Advisor	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
d_{ij} (0/1)	0.114**	-0.285***	0.076*	-0.182***	0.130	-0.042	0.014	0.073	-0.054	0.587***
	[0.056]	[0.054]	[0.040]	[0.050]	[0.099]	[0.109]	[0.055]	[0.193]	[0.072]	[0.222]
Pre-doc. Pub. (0/1)	0.208***	0.101***	0.176***	0.079***	-0.035	-0.091*	-0.022	-0.029	-0.042	-0.089
	[0.031]	[0.026]	[0.027]	[0.023]	[0.066]	[0.055]	[0.054]	[0.060]	[0.061]	[0.068]
Co-Advisor (0/1)	0.052	-0.003	-0.015	-0.049*	0.121**	0.144**	-0.036	0.028	-0.086	-0.041
	[0.032]	[0.029]	[0.024]	[0.026]	[0.060]	[0.062]	[0.053]	[0.063]	[0.061]	[0.069]
Nr. Advisees (log)	-0.003		-0.024		-0.042		0.010		0.025	
	[0.027]		[0.021]		[0.069]		[0.044]		[0.051]	
Advisor Productivity (log)	-0.097***		-0.073***		-0.155***		0.154***		0.207***	
	[0.027]		[0.021]		[0.053]		[0.044]		[0.060]	
Advisor Coauthors (log)	0.053***		0.010		0.102***		0.011		-0.023	
	[0.015]		[0.011]		[0.036]		[0.024]		[0.033]	
Full Prof. (0/1)	0.007		-0.014		-0.065		0.093		0.054	
	[0.039]		[0.030]		[0.073]		[0.060]		[0.070]	
Advisor Research Breadth	-0.516***		-0.367***		-0.324		-0.234		-0.073	
	[0.123]		[0.112]		[0.227]		[0.213]		[0.251]	
Constant	0.528***	0.396***	-0.635***	-0.857***	-0.061	-0.471***	0.661***	1.147***	0.612**	1.054***
	[0.121]	[0.039]	[0.096]	[0.037]	[0.260]	[0.087]	[0.210]	[0.130]	[0.258]	[0.135]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y		Y		Y
N	15748	15740	15748	15740	15748	15740	15748	15740	12111	12073
adj. R^2	0.393	0.504	0.348	0.448	0.205	0.358	0.186	0.246	0.170	0.245

Notes: d_{ij}=1 if advisor demonstrates engagement in interdisciplinary research within a 5-year window around PhD entry

Fixed-effect regressions using entropy weights. Robust standard errors clustered at the advisor level in brackets. Dependent vars. in $\log(x + 0.2)$.

*:p<0.10, **:p<0.05, ***:p<0.01

Table 3.9: PhD Students' Novelty by Research Orientation

	New Words (Binary) (1)	New Phrases (Binary) (2)	New Words (Avg.) (3)	New Phrases (Avg.) (4)
Panel A: Translational / Clinical				
Advisor ID Engagement	0.243	-0.283	0.426*	-0.341
	[0.165]	[0.184]	[0.258]	[0.297]
Pre-doctoral Pubs (0/1)	0.024	-0.030	0.010	-0.086
	[0.023]	[0.041]	[0.032]	[0.061]
Co-Advisor (0/1)	-0.008	-0.043	-0.010	-0.072
	[0.033]	[0.037]	[0.053]	[0.060]
Constant	0.079**	0.250***	-1.473***	-1.112***
	[0.039]	[0.052]	[0.061]	[0.093]
N	3330	3330	3330	3330
adj. R^2	0.149	0.185	0.077	0.129
Panel B: Fundamental / Basic				
Advisor ID Engagement	-0.092	-0.453***	-0.196	-0.883***
	[0.094]	[0.152]	[0.157]	[0.217]
Pre-doctoral Pubs (0/1)	-0.027	0.205	-0.061	0.254
	[0.040]	[0.138]	[0.065]	[0.227]
Co-Advisor (0/1)	0.010	0.077	0.020	0.199
	[0.015]	[0.065]	[0.026]	[0.132]
Constant	0.025	0.176**	-1.546***	-1.213***
	[0.024]	[0.079]	[0.028]	[0.124]
N	3953	3953	3953	3953
adj. R^2	0.210	0.215	0.181	0.202
Panel C: Hybrid				
Advisor ID Engagement	0.085	0.047	0.164	-0.226
	[0.124]	[0.257]	[0.161]	[0.397]
Pre-doctoral Pubs (0/1)	-0.007	0.039	-0.018	0.025
	[0.038]	[0.046]	[0.035]	[0.087]
Co-Advisor (0/1)	0.006	0.147**	-0.007	0.211*
	[0.049]	[0.069]	[0.056]	[0.109]
Constant	0.080	0.223**	-1.532***	-1.108***
	[0.055]	[0.087]	[0.064]	[0.151]
N	1979	1979	1979	1979
adj. R^2	0.166	0.245	0.136	0.219
Year, Cohort, Sci.Field, Unit, Advisor FE	Y	Y	Y	Y

Notes: Classification of Academic Units by Research Orientation reported in Table B.4.2. Fixed-effect regressions using entropy weights. Dependent vars. in $\log((x + 0.2))$. Robust standard errors clustered at the advisor level in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.10: PhD Students' Productivity by Research Orientation

	Publication Count (1)	Coauthor-wt. Pubs. (2)	Pubs. w/Advisor (3)	Avg. Citations (4)	Avg. Cit. w/Advisor (5)
Panel A: Translational / Clinical					
Advisor ID Engagement	-0.072	0.051	0.109	-0.549*	-0.563
	[0.173]	[0.136]	[0.294]	[0.302]	[0.753]
Pre-doctoral Pubs (0/1)	0.104***	0.096***	-0.009	-0.067	-0.177**
	[0.036]	[0.031]	[0.066]	[0.078]	[0.087]
Co-Advisor (0/1)	0.000	-0.054	0.196**	-0.014	-0.100
	[0.046]	[0.038]	[0.085]	[0.090]	[0.103]
Constant	0.258***	-0.943***	-0.510***	1.149***	1.304***
	[0.043]	[0.042]	[0.089]	[0.104]	[0.120]
N	5184	5184	5184	5184	4004
adj. R^2	0.516	0.468	0.331	0.244	0.208
Panel B: Fundamental / Basic					
Advisor ID Engagement	-0.266*	-0.258**	-0.766*	-0.801***	-0.736***
	[0.146]	[0.105]	[0.413]	[0.279]	[0.239]
Pre-doctoral Pubs (0/1)	0.248***	0.163***	0.203	0.256*	0.166**
	[0.052]	[0.037]	[0.129]	[0.137]	[0.081]
Co-Advisor (0/1)	0.043	-0.012	0.125	0.086	-0.038
	[0.063]	[0.045]	[0.121]	[0.109]	[0.166]
Constant	0.074*	-1.032***	-0.595***	1.221***	1.431***
	[0.044]	[0.042]	[0.103]	[0.151]	[0.181]
N	7439	7439	7439	7439	5929
adj. R^2	0.450	0.441	0.358	0.325	0.331
Panel C: Hybrid					
Advisor ID Engagement	-0.445***	-0.275***	-0.279	0.088	1.795*
	[0.144]	[0.101]	[0.328]	[0.370]	[0.977]
Pre-doctoral Pubs (0/1)	0.069	0.035	-0.205*	-0.145	-0.041
	[0.046]	[0.049]	[0.114]	[0.128]	[0.120]
Co-Advisor (0/1)	-0.005	0.008	0.148	0.116	0.179
	[0.045]	[0.038]	[0.119]	[0.107]	[0.147]
Constant	0.303***	-0.978***	-0.564***	1.378***	1.227***
	[0.067]	[0.065]	[0.131]	[0.194]	[0.274]
N	3167	3167	3167	3167	2593
adj. R^2	0.535	0.469	0.378	0.234	0.334
Year, Cohort, Sci.Field, Unit, Advisor FE	Y	Y	Y	Y	Y

Notes: Classification of Academic Units by Research Orientation reported in Table B.4.2. Fixed-effect regressions using entropy weights. Dependent vars. in $\log s((x + 0.2))$. Robust standard errors clustered at the advisor level in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.11: PhD Students' Research Collaboration

	Collaborations				Collaborators' Expertise			
	w. ID members		New Coauthors		Avg. Distance		SD. Distance	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
d_{ij} (0/1)	0.142***	-0.027	0.107	-0.847***	0.005	0.059***	-0.038***	0.036
	[0.024]	[0.034]	[0.072]	[0.200]	[0.013]	[0.019]	[0.011]	[0.026]
Pre-doc. Pub. (0/1)	-0.009	-0.004	-0.084*	-0.175***	-0.041***	-0.024***	0.009	0.007
	[0.016]	[0.019]	[0.047]	[0.050]	[0.010]	[0.009]	[0.010]	[0.012]
Co-Advisor (0/1)	0.033*	0.023	0.147***	0.112**	-0.004	0.000	-0.036***	-0.044***
	[0.018]	[0.015]	[0.053]	[0.057]	[0.011]	[0.014]	[0.012]	[0.013]
Nr. Advisees (log)	0.015		0.117**		-0.002		0.013	
	[0.024]		[0.049]		[0.010]		[0.010]	
Advisor Productivity (log)	-0.006		0.005		-0.003		0.007	
	[0.020]		[0.036]		[0.008]		[0.007]	
Advisor Coauthors (log)	-0.008		0.117***		0.009**		-0.007	
	[0.010]		[0.026]		[0.004]		[0.004]	
Full Prof. (0/1)	0.020		0.010		0.006		-0.013	
	[0.022]		[0.065]		[0.011]		[0.011]	
Advisor Research Breadth	-0.053		-0.475**		0.082*		-0.053	
	[0.076]		[0.205]		[0.045]		[0.043]	
Constant	0.060	0.099***	0.299*	1.049***	-0.270***	-0.266***	-0.697***	-0.726***
	[0.085]	[0.027]	[0.170]	[0.115]	[0.033]	[0.017]	[0.032]	[0.017]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y		Y
N	15748	15740	15236	15226	14665	14651	13856	13834
adj. R^2	0.348	0.464	0.214	0.274	0.096	0.172	0.101	0.210

Notes: Fixed effect regressions using entropy weights. Robust std. errors clustered at the advisor level. Cols.(3)-(8) dep.vars. in $\log(x + 0.2)$.
d_{ij}=1 if advisor demonstrates engagement in interdisciplinary research within a 5-year window around PhD entry
Avg. Distance: Jaccard between each collaborator's prior expertise; SD: Std deviation of Jaccard distances across the student's collaborators
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.12: PhD Students' Interdisciplinarity

	Variety		Disparity		Balance	
	(1)	(2)	(3)	(4)	(5)	(6)
d_{ij} (0/1)	-0.048	-0.072**	-0.025**	0.015	-0.017	-0.043*
	[0.032]	[0.033]	[0.012]	[0.010]	[0.015]	[0.023]
Pre-doc. Pub. (0/1)	-0.098***	-0.073***	-0.026***	-0.016**	-0.040***	-0.036***
	[0.021]	[0.021]	[0.007]	[0.006]	[0.010]	[0.011]
Co-Advisor (0/1)	-0.032	0.016	-0.006	0.002	-0.007	-0.001
	[0.021]	[0.022]	[0.005]	[0.007]	[0.010]	[0.012]
Nr. Advisees (log)	0.070***		0.014**		0.027***	
	[0.019]		[0.006]		[0.010]	
Advisor Productivity (log)	0.083***		0.011**		0.020**	
	[0.017]		[0.006]		[0.008]	
Advisor Coauthors (log)	0.001		-0.004		-0.000	
	[0.010]		[0.003]		[0.005]	
Full Prof. (0/1)	-0.015		0.001		-0.001	
	[0.023]		[0.007]		[0.011]	
Advisor Research Breadth	0.413***		0.133***		0.133***	
	[0.095]		[0.034]		[0.048]	
Constant	1.292***	1.783***	-0.521***	-0.463***	-0.514***	-0.359***
	[0.086]	[0.027]	[0.028]	[0.008]	[0.038]	[0.016]
Year FE	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y
N	14053	14038	14053	14038	14053	14038
adj. R^2	0.369	0.513	0.187	0.323	0.153	0.276

Notes: Fixed effect regressions using entropy weights. Robust std. errors clustered at the advisor level.

$d_{ij} = 1$ if advisor demonstrates engagement in ID research within a 5-year window around PhD entry.

Cols.(1)-(4) dep.vars. in $\log(x + 0.2)$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Variety measures the number of distinct fields of research.

Disparity measures the cognitive distance across those fields.

Balance measures how evenly are different disciplines referenced.

References

Alberts, B., M. W. Kirschner, S. Tilghman, and H. Varmus (2014): "Rescuing US biomedical research from its systemic flaws," *Proceedings of the National Academy of Sciences of the United States of America*, 111, 5773–5777.

Andersen, M. Z., S. Fonnes, and J. Rosenberg (2021): "Time from submission to publication varied widely for biomedical journals: a systematic review," *Current Medical Research and Opinion*, 37, 985–993.

Anderson, E. M. and A. L. Shannon (1988): "Toward a Conceptualization of Mentoring," *Journal of Teacher Education*, 38–42.

Arts, S., N. Melluso, and R. Veugelers (2025): "Beyond Citations: Measuring Novel Scientific Ideas and their Impact in Publication Text," *Review of Economics and Statistics*, 1–33.

Bargar, R. R. and J. Mayo-Chamberlain (1983): "Advisor and Advisee Issues in Doctoral Education," *The Journal of Higher Education*, 54, 407–432.

Bercovitz, J. and M. Feldman (2011): "The mechanisms of collaboration in inventive teams: Composition, social networks, and geography," *Research Policy*, 40, 81–93.

Biancani, S., L. Dahlander, D. A. Mcfarland, and S. Smith (2018): "Superstars in the making? The broad effects of interdisciplinary centers," *Research Policy*, 47, 543–557.

Biancani, S., D. A. McFarland, and L. Dahlander (2014): "The semiformal organization," *Organization Science*, 25, 1306–1324.

Bikard, M., F. Murray, and J. S. Gans (2015): "Exploring trade-offs in the organization of scientific work: Collaboration and scientific reward," *Management Science*, 61, 1473–1495.

Bjork, B. C. and D. Solomon (2013): "The publishing delay in scholarly peer-reviewed journals," *Journal of Informetrics*, 7, 914–923.

Boardman, C. and E. Corley (2008): "University research centers and the composition of research collaborations," *Research Policy*, 37, 900–913.

Boudreau, K. J., T. Brady, I. Ganguli, P. Gaule, E. Guinan, A. Hollenberg, and K. R. Lakhani (2017): "A field experiment on search costs and the formation of scientific collaborations," *The Review of Economics and Statistics*, XCIX, 565–576.

Brostrom, A. (2019): "Academic breeding grounds: Home department conditions and early career performance of academic researchers," *Research Policy*, 48, 1647–1665.

Carayol, N., E. Henry, and M. Lanoe (2023): "Stimulating Collaborations: Evidence from a Research Cluster Policy," *Review of Economics and Statistics*.

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, 1633–1644.

Corsini, A., M. Pezzoni, and F. Visentin (2022): "What makes a productive Ph.D. student?" *Research Policy*, 51.

Cummings, J. N. and S. Kiesler (2005): "Collaborative research across disciplinary and organizational boundaries," .

Daniels, R. J. (2015): "A generation at risk: Young investigators and the future of the biomedical workforce," *Proceedings of the National Academy of Sciences of the United States of America*, 112, 313–318.

D'Este, P. and N. Robinson-Garcia (2023): "Interdisciplinary research and the societal visibility of science: The advantages of spanning multiple and distant scientific fields," *Research Policy*, 52, 104609.

Fleming, L., S. Mingo, and D. Chen (2007): "Collaborative brokerage, generative creativity, and creative success," *Administrative Science Quarterly*, 52, 443–475.

Fontana, M., M. Iori, F. Montobbio, and R. Sinatra (2020): "New and atypical combinations: An assessment of novelty and interdisciplinarity," *Research Policy*, 49.

Fontana, M., M. Iori, V. L. Sciabolazza, and D. Souza (2022): "The interdisciplinarity dilemma: Public versus private interests," *Research Policy*, 51, 104553.

Franzoni, C. and C. Rossi-Lamastra (2017): "Academic tenure, risk-taking and the diversification of scientific research," *Industry and Innovation*, 24, 691–712.

Freeman, R., E. Weinstein, E. Marincola, J. Rosenbaum, and F. Solomon (2001): "Competition and Careers in Biosciences," *Science*, 294, 2293–2294.

Gardner, S. K. (2010): "Contrasting the Socialization Experiences of Doctoral Students in High- and Low-Completing Departments: A Qualitative Analysis of Disciplinary Contexts at One Institution," *The Journal of Higher Education*, 81, 61–81.

Graddy-Reed, A., L. Lanahan, and J. D'Agostino (2021): "Training across the academy: The impact of R&D funding on graduate students," *Research Policy*, 50, 104224.

Guimaraes, P. and P. Portugal (2010): "A simple feasible procedure to fit models with high-dimensional fixed effects," *The Stata Journal*, 10, 628–649.

Haeussler, C. and H. Sauermann (2020): "Division of labor in collaborative knowledge production: The role of team size and interdisciplinarity," *Research Policy*, 49.

Hainmueller, J. (2011): "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies," *Political Analysis*, 16, 25–46.

Hainmueller, J. and Y. Xu (2013): "ebalance: A Stata Package for Entropy Balancing," *Journal of Statistical Software*, 54, 1–18.

Herzog, C., D. Hook, and S. Konkiel (2020): "Dimensions: Bringing down barriers between scientometricians and data," *Quantitative Science Studies*, 1, 387–395.

Hottenrott, H. and C. Lawson (2015): "Flying the nest: how the home department shapes researchers' career paths," *Studies in Higher Education*, 42, 1091–1109.

Hu, Z., J. Li, and H. K. Kwan (2022): "The effects of negative mentoring experiences on mentor creativity: The roles of mentor ego depletion and traditionality," *Human Resource Management*, 61, 39–54.

Jacobs, J. A. (2013): *In defense of disciplines : interdisciplinarity and specialization in the research university*, vol. 1, The University of Chicago Press.

Jacobs, J. A. and S. Frickel (2009): "Interdisciplinarity: A Critical Assessment," *Annual Review of Sociology*, 35, 43–65.

Kaellberg, D. and I. Waernbaum (2023): "Large Sample Properties of Entropy Balancing Estimators of Average Causal Effects," *Econometrics and Statistics*.

Leahey, E. (2018): "The Perks and Perils of Interdisciplinary Research," *European Review*, 26, S55–S67.

Leahey, E. and S. N. Barringer (2020): "Universities' commitment to interdisciplinary research: To what end?" *Research Policy*, 49.

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, 105–139.

Lyall, C. and L. R. Meagher (2012): "A Masterclass in interdisciplinarity: Research into practice in training the next generation of interdisciplinary researchers," *Futures*, 44, 608–617.

Ma, Y., S. Mukherjee, and B. Uzzi (2020): "Mentorship and protégé success in STEM fields," *PNAS*, 117, 14077–14083.

Mansilla, V. B., M. le Lamont, K. Sato, and M. Lamont (2016): "Shared Cognitive-Emotional-Interactional Platforms: Markers and Conditions for Successful Interdisciplinary Collaborations," *Technology, & Human Values*, 41, 571–612.

Nagle, F. and F. Teodoridis (2020): "Jack of all trades and master of knowledge: The role of diversification in new distant knowledge integration," *Strategic Management Journal*, 41, 55–85.

Onal-Vural, M., L. Dahlander, and G. George (2013): "Collaborative Benefits and Coordination Costs: Learning and Capability Development in Science," *Strategic Entrepreneurship Journal*, 7, 122–137.

Paglis, L. L., S. G. Green, and T. N. Bauer (2006): "Does Adviser Mentoring Add Value? A Longitudinal Study of Mentoring and Doctoral Student Outcomes," *Research in Higher Education*, 47.

Pezzoni, M., V. Sterzi, and F. Lissoni (2012): "Career progress in centralized academic systems: Social capital and institutions in France and Italy," *Research Policy*, 41, 704–719.

Reagans, R., E. Miron-Spektor, and L. Argote (2016): "Knowledge Utilization, Coordination, and Team Performance," *Organization Science*, 1108–1124.

Rhoten, D. and A. Parker (2004): "Risks and Rewards of an Interdisciplinary Research Path," *Science*, 306.

Roche, M. P. (2022): "Academic Entrepreneurship: Entrepreneurial Advisors and Their Advisees' Outcomes," *Organization Science*.

Rose, M. E. and S. Shekhar (2023): "Adviser connectedness and placement outcomes in the economics job market," *Labour Economics*, 84, 102397.

Rylance, R. (2015): "Global funders to focus on interdisciplinarity," *Nature*, 525.

Shibayama, S. (2019): "Sustainable development of science and scientists: Academic training in life science labs," *Research Policy*, 48, 676–692.

Simonton, D. K. (1999): *Origins of Genius: Darwinian Perspectives on Creativity*, Oxford University Press.

Starr, E. and B. Goldfarb (2020): "Binned scatterplots: A simple tool to make research easier and better," *Strategic Management Journal*, 41, 2261–2274.

Stephan, P. E. (2015): *How economics shapes science*, Harvard University Press.

Stirling, A. (2007): "A general framework for analysing diversity in science, technology and society," *Journal of The Royal Society Interface*, 4, 707–716.

Su, X. (2014): "Academic scientists' affiliation with university research centers: Selection dynamics," *Research Policy*, 43, 382–390.

Teodoridis, F., M. L. Bikard, and K. Vakili (2019): "Creativity at the Knowledge Frontier: The Impact of Specialization in Fast-and Slow-paced Domains*," *Administrative Science Quarterly*, 64, 894–927.

Trunschke, M., B. Peters, D. Czarnitzki, and C. Rammer (2024): "Pandemic effects: Do innovation activities of firms suffer from Long COVID?" *Research Policy*, 53, 105024.

Tuma, T. T., J. D. Adams, B. C. Hultquist, and E. L. Dolan (2021): "The Dark Side of Development: A Systems Characterization of the Negative Mentoring Experiences of Doctoral Students," *CBE Life Sciences Education*, 20, ar16.

Uzzi, B., S. Mukherjee, M. Stringer, and B. Jones (2013): "Atypical combinations and scientific impact," *Science*, 342, 468–472.

Uzzi, B. and J. Spiro (2005): "Collaboration and creativity: The small world problem," *American Journal of Sociology*, 111, 447–504.

Vakili, K., F. Teodoridis, and M. Bikard (2021): "Detrimental Collaborations in Creative Work: Evidence from Economics," *Organization Science*.

Wang, J. and S. Shibayama (2022): "Mentorship and creativity: Effects of mentor creativity and mentoring style," *Research Policy*, 51.

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

Wuestman, M., I. Wanzenbock, and K. Frenken (2023): "Local peer communities and future academic success of Ph.D. candidates," *Research Policy*, 52, 104844.

Chapter 4

Framing Novelty: How Context Shapes the Evaluation of Entrepreneurial Ideas[1]

[1]Joint work with Stijn Kelchtermans. We thank Alex Povora from the European Innovation Council (EIC) and David Ampudia from the Innovation Growth Lab (IGL) for their support with data access, which was granted through a secure virtual system in the context of the EIC-IGL Datathon initiative. The arguments and views expressed in this paper do not reflect those of the EIC or the IGL. We also thank comments and suggestions by seminar participants at Universitat Pompeu Fabra, the Competition and Innovation Summer School in Montenegro, and the DRUID25 conference.

4.1 Introduction

The success of investors and support programs targeting early-stage ventures largely depends on their ability to identify high-potential ideas (Gompers et al., 2020; Petty et al., 2023). While it is commonly acknowledged that novel ideas are essential for new ventures to create or disrupt markets and build a competitive advantage (Christensen et al., 2018; Henderson and Clark, 1990), external appreciation of novelty, a cognitively demanding interpretive process, often determines whether these ideas are ultimately pursued (Amabile, 1996; Quy and Huy, 2007). In diverse contexts, such as pitches to prospective investors, business plan competitions, and applications to startup accelerators, early-stage ventures seek evaluators' support by articulating their entrepreneurial ideas (Martens et al., 2007; Navis and Glynn, 2011).

Prior research has advanced our understanding of entrepreneurial storytelling, and why some narratives are more persuasive than others (Anglin et al., 2023; Lounsbury and Glynn, 2001). A rich body of work focuses on subtle linguistic choices, often referred to as framing strategies, and their influence on how entrepreneurial ideas are perceived and judged (Moss et al., 2018; Pan et al., 2020; Snihur et al., 2022). Despite extensive research on how entrepreneurs frame their ideas to gain support, less is known about how the evaluation context influences the effectiveness of different framing strategies. This gap in our understanding matters because the evaluation context—through its screening procedures, award criteria, and other rules that determine who evaluates and how they evaluate—has a direct impact on which ideas ultimately receive resources (Boudreau et al., 2016; Criscuolo et al., 2017; Lane et al., 2022). More specifically, we have yet to understand the conditions under which framing strategies may enhance support for novel ideas, particularly in real-world high-stakes competitions aimed at selecting breakthrough ideas (Taeuscher and Rothe, 2024).

To address this gap, we study the association between a venture's framing choices and evaluation outcomes, integrating two key features of the evaluation context. Building on prior work on the selection of novel ideas, we

examine whether and how the effectiveness of framing strategies is moderated by (i) the atypicality of an idea relative to its competitors ("contextual novelty"), which potentially complicates idea interpretation and valuation (Cutolo and Ferriani, 2024); and ii) the extent to which submission volumes exceed levels predicted by industry or time trends ("congestion"), as unusually high evaluation loads may reduce the cognitive bandwidth that evaluators devote to each idea and may shift which framing strategies evaluators pay attention to (Piezunka and Dahlander, 2015).

Empirically, we analyze both funded and rejected proposals submitted to the European Innovation Council (EIC) Accelerator (2014–2020), a prominent large-scale support program that screens thousands of early-stage technology-based ventures and seeks to fund ideas with "the potential to create new markets or disrupt existing ones" (EU, 2024, p. 6). Because the way the venture presents itself plays a crucial role in the assessment of new venture ideas (Berg, 2016; Rindova and Petkova, 2007), we capture the relative emphasis a proposal places on aspects like the innovativeness and feasibility of its value proposition, using a validated vocabulary of "promotional adjectives" that two independent studies have robustly linked to the funding success of novel scientific ideas (Millar et al., 2022; Peng et al., 2024). Moreover, we employ natural language processing techniques to quantify each venture's atypicality (defined as the semantic distance between its value proposition and those of competing applicants), allowing us to disentangle contextually novel ideas from the venture's framing choices. Finally, we leverage the EIC Accelerator's rolling submission process combined with cutoff dates to measure whether evaluators need to assess an abnormally high volume of submissions, interpreting it as the extent of congestion evaluators face in the assessment process (Ferguson and Carnabuci, 2017).

Complementing prior insights on entrepreneurial framing, we find that evaluators reward feasibility-oriented framings. The richer characterization of framing strategies that we use compared to earlier studies offers nuance by showing that ventures benefit from a framing that highlights both the feasibility and the innovativeness of their value proposition, while not adopting

either of these framing emphases results in a higher probability of rejection. We add further insights into the boundary conditions on the theory of successful framing strategies by taking into account the evaluation context in two ways.

First, we differentiate between the novelty of ventures' ideas through the concept of contextual novelty, which captures the extent to which a given venture's idea is atypical relative to other submissions in its cohort. The second contextual factor we consider is the above-expected workload, or congestion, faced by evaluators. Interestingly, we find that the benefit of feasibility framing is not at play for contextually novel ideas. Yet, we find that evaluators are less likely to weed out contextually novel ideas in situations of high congestion, suggesting their role as institutional stalwarts comes to the fore. Finally, analyzing the interaction of the two contextual factors with framing shows that feasibility framing does reduce the odds of rejection for contextually novel ideas, but only in circumstances of high congestion. We find further support for our arguments in placebo tests using alternative framing variables and alternative operationalizations of contextual novelty.

Our study contributes to the literature on the framing strategies of new ventures and the evaluation of novel ideas. Our results are consistent with prior work on strategic framing, showing that support for novel ideas increases when ventures opt for a framing that aligns with evaluators' expectations of the idea's feasibility. Furthermore, we extend theory on the effectiveness of framing strategies by showing the moderating role of the idea evaluation context. In particular, the result on congestion shows that evaluators who are under strain do not necessarily fall victim to a familiarity bias but may rather "carry out the mission" of the organization, rendering support to ideas that stand out from those of competing ventures. We also show the limitations of feasibility framing by showing that, for contextually novel ideas, the benefit only materializes in a high-congestion context. Finally, complementing prior work that relied on experiments (Falchetti et al., 2022, 2024), this is the first study to our knowledge to test these mechanisms in a real-world, observational setting.

4.2 Theory and Hypotheses

Framing and the Evaluation of Novel Ideas

Because of its pivotal role in attracting support for firms, strategic framing — broadly defined as the use of rhetorical devices such as linguistic styles, selected vocabulary, and storytelling, to mobilize external support and resources (Cornelissen and Werner, 2014) — has attracted significant attention in the entrepreneurship literature (Garud et al., 2014; Martens et al., 2007; Snihur et al., 2022). To attract support for their venture ideas, entrepreneurs strategically emphasize specific attributes of their offerings, for instance, highlighting the feasibility of their technological or business solutions or underscoring the venture's innovativeness and potential (Navis and Glynn, 2011; van Werven et al., 2015). Unlike other entrepreneurial positioning efforts that appeal to a broad audience base and remain rather stable over time (e.g., branding), framing is essentially contextual and audience-specific (Navis and Glynn, 2011; Snihur et al., 2022). By emphasizing specific attributes of their value proposition, entrepreneurial ventures purposefully use framing to target an audience in a given context (Taeuscher and Rothe, 2024).

A key insight from this body of work is that concrete-oriented framings, meaning narratives using a more detailed and context-specific vocabulary, constitute effective persuasion strategies (Packard and Berger, 2021; Pan et al., 2018), and can positively influence the evaluation of new venture ideas (Falchetti et al., 2022). Following a cognitive psychology perspective based on Construal Level Theory[2] (CLT; Trope and Liberman, 2010), the chief mechanism evoked to explain such findings is that an individual's reliance on a high-level (abstract) or low-level (concrete) construal shapes her preferences for creative ideas (Mueller et al., 2014). More specifically, while low-level construals focus on the feasibility of achieving a desired end state, such as the practical steps or technical details required to implement an idea, high-

[2]CLT conceptualizes two types of mental representations: concrete language as a low-level construal and abstract language as a high-level construal. Crucially, low-level construals reflect closer psychological distance, shaping how people make decisions, interpret events, and prioritize goals (Trope and Liberman, 2010; Trope et al., 2007).

level construals tend to emphasize the desirability or potential of an idea (Trope et al., 2007). For example, a concrete framing that emphasizes how to use a product or service, details a technological solution, or outlines a clear commercialization strategy, suggests an understanding of user-specific needs that is appreciated by relevant audiences (Packard and Berger, 2021). This rationale explains why experts tend to favor innovative venture ideas framed in concrete terms, emphasizing feasibility.

We argue that existing research on framing strategies provides a clear conceptualization, but it remains highly stylized. Prior work predominantly treats framing as a binary choice of either concrete (i.e., underscoring functions or implementation details) or abstract (i.e., underscoring the potential or desirability), without explicitly allowing for a graded emphasis on each aspect (e.g., Falchetti et al., 2022). We argue that overcoming this dichotomy allows us to reconnect with the original perspective on linguistic framing, following the notion that "*the process of abstraction is not an all-or-nothing phenomenon*" (Trope et al., 2007, p. 84). During the abstraction process of framing an idea, the particular details or "incidental features" that only apply in a specific context can be gradually removed or omitted. What remains is a more simplified, abstract, and schematic representation that captures the essential aspects of the original idea without explicit ties to its original context.

While the "concrete versus abstract" dichotomy adopted in prior work allows for clear hypotheses and experimental tests, the framing of real ventures rarely falls neatly into such clear categories. Consequently, our conceptualization of framing moves beyond this binary distinction and treats framing as a graded construct where both abstract ("innovativeness ") and concrete ("feasibility") elements can be present to varying degrees within the same message (Snihur et al., 2022). Extending prior experimental work (Falchetti et al., 2022), we anticipate that ventures evaluated in a real-world setting will benefit from emphasizing the feasibility of their novel ideas to a greater extent. Hence, we formulate the following baseline hypothesis:

***Hypothesis 1**: Novel ideas that emphasize feasibility in their framing*

are less likely to be rejected

Besides its rather crude conceptualization of framing strategies, we argue that extant work has insufficiently explored the context in which idea selection occurs and that this has hindered our understanding of the effectiveness of ventures' framing strategies. Because the interplay between framing strategies and support for novel ideas may critically depend on the specific conditions under which they are evaluated, empirical evidence beyond experimental studies could enhance our understanding of when framing strategies are most effective. More specifically, contextual factors can shape the evaluation of novel ideas by influencing evaluators' attitudes toward uncertainty, their preferences for specific attributes (e.g., technologies or functionalities), and the criteria they apply when assessing an idea's merits (Boudreau et al., 2016; Criscuolo et al., 2017; Lane et al., 2022).

A first important role for context lies in its implications for the definition of novelty itself. Despite the notion that highly novel entrepreneurial ideas constitute a key source of competitive advantage (Christensen et al., 2018; Henderson and Clark, 1990), the combined effect of different framing strategies and the nature of an idea's novelty remains underexplored. One notable exception is Falchetti et al. (2024), who provide experimental evidence that framing effectiveness depends on the degree of novelty of a new venture idea's, and specifically, that ideas perceived as radical (i.e., highly novel) are more likely to elicit support (from lay evaluators) when framed in concrete terms, while ideas perceived as incremental benefit from abstract framings. In other words, the effectiveness of framing depends on the evaluator's appreciation of novelty, and the perceived novelty of an idea may influence which framing strategy is most effective. Below, we will argue that an idea's novelty is context-dependent, i.e., that it is evaluated relative to the novelty of competing ideas ("contextually novel") rather than in an absolute sense (Deichmann et al., 2024; Mueller et al., 2014, 2018). Second, we know that decision-makers adapt their behavior when faced with information overload (Gifford, 1992; Van Knippenberg et al., 2015), which is common in real-life evaluation contexts. Below, we will hypothesize how a situation

of "congestion" —in terms of evaluators' attention— affects the evaluation of contextually novel ideas and how it interacts with the venture's framing strategy.

Contextual Novelty and Framing

The evaluation context can shape not only how decisions are made but also what counts as "novel". Following Litchfield et al. (2015), ideas may be novel to the extent that they are uncommon in terms of either their envisaged task or social context; and they are best characterized as distant from current practice, knowledge base, or other vantage point (Amabile, 1996). In this view, novelty is a multidimensional construct encompassing both technical and social aspects (Rosenkopf and McGrath, 2011).

Because of the context-dependent and subjective nature of an idea's novelty, its empirical examination is known to be challenging (Deichmann et al., 2024; Mueller et al., 2018, 2014). In competitive selection processes, novelty is often assessed relatively—in comparison to other submissions in the same cohort—rather than in absolute terms, that is, in terms of its atypicality relative to competing ideas (Fontana et al., 2020; Uzzi et al., 2013). Atypicality can be understood as "a deviation from contextualized expectations of representativeness" (Cutolo and Ferriani, 2024, p. 161). This notion is often grounded in a probabilistic perspective, meaning that evaluators have an understanding of what is representative of an idea in a given domain, and what constitutes a non-trivial departure from that reference group: a "contextually novel" idea (Hannan et al., 2019).

Such atypicality presents evaluators with a double-edged sword Cutolo and Ferriani, 2024. On one hand, it signals distinctiveness and potential competitive advantage; on the other, it complicates categorization and increases interpretive uncertainty, which can trigger skepticism (Harvey and Mueller, 2021; Criscuolo et al., 2017). Prior work suggests that more concrete framings—those that focus on feasibility—can help audiences appreciate novel ideas by clarifying their implementation and reducing ambiguity (Falchetti et al., 2024). This logic should apply even more strongly to contextually

novel ideas, which face elevated cognitive hurdles in the evaluation process. In addition, referring again to the role of context, we expect this relation to be particularly salient in screening processes where evaluators are explicitly tasked with selecting ventures that have "breakthrough potential", such as the accelerator process we study (EC, 2024).

> ***Hypothesis 2****: Contextually novel ideas that emphasize feasibility in their framing are less likely to be rejected*

Congestion and the Framing of Contextually Novel Ideas

Organizational theory emphasizes that decision-makers are bounded in their rationality and capacity to process information (Simon, 1957). When evaluation workloads surge—what we term congestion—cognitive impediments become more pronounced (Van Knippenberg et al., 2015). Under such conditions, evaluators have less time and attention to allocate to each proposal, increasing the appeal of decision heuristics that allow for quicker, less effortful judgments (Gifford, 1992; Huang and Pearce, 2015). Consistent with the bounded-rationality perspective, ample evidence indicates that highly novel ideas may not be recognized as such or receive less support when they compete for attention (Ferguson and Carnabuci, 2017; Chai and Menon, 2019; Criscuolo et al., 2017). For instance, Ferguson and Carnabuci (2017) show that when patent examiners face higher workloads, the likelihood of granting a patent to boundary-spanning inventions (i.e., patents combining multiple and distant technological domains) is significantly undermined. Similarly, Piezunka and Dahlander (2015) argue that information overload increases the likelihood that evaluators who seek to simplify their decision-making process and rationalize their restricted cognitive capacity, pay more attention to familiar ideas, and filter out potentially useful but unfamiliar ideas.

When forced to make relatively fast decisions with limited information, decision-makers tend to rely on simplified heuristics (i.e., rules of thumb) or experience-based schemas (i.e., recall of prior cases) (Huang and Pearce, 2015; Kirsch et al., 2009). For instance, experienced angel investors often follow their emotional-laden "gut feel" or "intuition" when making early-stage

investment decisions (Huang and Pearce, 2015). However, because information overload and congestion compress the attention evaluators can devote to each idea, their cognitive capacity to thoroughly scrutinize proposals becomes limited. Consequently, they may shift away from deep, time-intensive judgments and toward quick "minimal criteria" or satisficing heuristics, inducing them to rely on simplified decision-making rules (Van Knippenberg et al., 2015). In particular, given that atypical ideas are harder to understand, as categorization research has shown (e.g., Kovács and Hannan, 2015), we posit that evaluators may fall back on the organization's objectives to guide their decisions. Therefore, if the organization's mission is to select, from the mass of candidates, those ventures with the potential to disrupt entire industries, then evaluators may well be drawn to the atypical ideas that are characteristic of extraordinary innovations (Fleming, 2001). In the context of patent grant decisions, Ferguson and Carnabuci, 2017 show that, conditional on being granted, atypical patents garner more citations. Prior work, such as Ferguson & Carnabuci's (2017) work on patent examiners, has also shown that the cognitive difficulties in assessing atypical ideas can be offset if institutional gatekeepers use the inventiveness of such ideas as a key evaluation criterion. Thus, we hypothesize that evaluators, under situations of high congestion, are more likely to select contextually novel ideas, consistent with the objective they are tasked with.

> ***Hypothesis 3****: Contextually novel ideas are less likely to be rejected when evaluated under conditions of greater congestion than when evaluated under conditions of lower congestion*

Finally, while high congestion limits evaluators' bandwidth to process subtle rhetorical cues, we expect an important exception when evaluators encounter contextually novel ideas. Such ideas deviate more strongly from evaluators' mental templates and are therefore harder to interpret and evaluate (Criscuolo et al., 2017; Cutolo and Ferriani, 2024). Under normal workload conditions, evaluators can invest the cognitive effort required to make sense of atypical offerings, drawing on deliberative processing and deeper scrutiny. Under high congestion, however, evaluators face acute attention scarcity and

must make faster judgments, often resorting to simplifying heuristics or cues that reduce uncertainty (Simon, 1957; Van Knippenberg et al., 2015). In this situation, a feasibility-oriented framing—concrete language that specifies implementation steps, technical robustness, and compatibility—can help evaluators quickly translate an unusual idea into an actionable mental model. By reducing ambiguity, feasibility framing enables evaluators to assess whether an atypical proposal can be executed successfully, even when time for detailed sensemaking is limited. This aligns with Construal Level Theory's prediction that low-level (concrete) construals facilitate judgments about feasibility under cognitive strain (Trope and Liberman, 2010). Thus, we expect that under high congestion, feasibility framing will disproportionately benefit contextually novel ideas, mitigating the interpretive challenges that otherwise risk disadvantaging them.

> ***Hypothesis 4****: Contextually novel ideas that emphasize feasibility in their framing are less likely to be rejected under conditions of greater congestion than when evaluated under conditions of lower congestion*

4.3 Methodology

Research Context and Data

Governments support high technology ventures for a variety of reasons, most notably to reduce information asymmetries that leave new ventures underfunded by traditional financial intermediaries (Lerner, 2002, 2009). While recent evidence argues that government programs are capable of implementing selection processes that prioritize risky technologies (Dutta et al., 2022; Myers et al., 2025), evidence on how framing shapes idea evaluation in public support programs, and particularly accelerators, is scarce (Hallen et al., 2020). We study the European Innovation Council (EIC) Accelerator, a flagship funding program of the European Commission. The EIC targets early-stage, technology-driven start-ups that aim to develop innovative products, services, or business models capable of "*creating new markets or disrupting*

existing ones, both in Europe and globally" (EU, 2024). Prospective applicants typically need substantial resources to develop and scale up their technologies and to demonstrate market potential in a real-world environment.[3] Like most accelerator programs, the EIC offers fixed-term support, combining funding with mentorship and networking services (Cohen et al., 2019; Hochberg, 2016).

Specifically, the EIC finances innovation activities expected to be completed within 24 months, offering grants of up to €2.5 million, a level of support that is generous by international standards.[4] Each year, the EIC issues calls for proposals that remain continuously open but have specific cutoff dates to qualify in a given evaluation round.[5] Unlike smaller-scale processes, such as those performed by venture capital firms that source deals through professional networks and rely on extensive due diligence (Gompers et al., 2020), accelerators such the EIC Accelerator rely on a structured screening process to select the most promising ventures from a very broad applicant pool. Proposals submitted before a given cutoff date form a batch that undergoes evaluation in the corresponding round. The evaluation process is structured in two main steps.[6] The first step is arduous and time-intensive, involving hundreds of proposals remotely assessed by three to four independent evaluators who individually assess a number of proposals, and generate a standardized short report and a score.[7]

To this end, the EIC assembles a pool of external evaluators who are

[3]The program is formally open to individuals, start-ups, SMEs, and small mid-caps (up to 499 employees) from the 27 EU Member States and a list of associated countries. In practice, few SMEs and small mid-caps apply and get support; we provide further details in the data section.

[4]For comparison, the well known Y Combinator Accelerator in the United States provides USD 500,000 in funding and equity. As of 2019, the EIC also offered an equity component that could reach up to €15 million. This equity component is determined on a case-by-case after the formal evaluation process is finalized and the venture has been accepted. We do not analyze equity funding in this paper.

[5]For instance, in 2020 the call featured four cutoff dates for completing applications: 8 January, 18 March, 19 May, and 20 October.

[6]Further details on the evaluation procedure during our study period can be consulted in the following EIC website.

[7]Regrettably, we do not observe individual reports or scores, but rather the aggregated decision: "go" or "no go."

tasked to review each proposal against the publicly disclosed award criteria of "*excellence*," "*impact*," and "*implementation*" (EU, 2024). The evaluators are chosen by the EIC in light of their professional experience and domain expertise (for example, in a relevant technology or industry), and face suitability and conflict of interest checks before being contracted, in a process conducted prior to each call for proposals.[8] Crucially, since the evaluator pool is assembled before submissions open following this bureaucratic process, the number of evaluators remains inelastic relative to the number of proposals submitted at each cutoff date.

Applicants are informed about the assessment criteria evaluators must follow, as this is publicly disclosed on the EIC website, in the call for proposals, and in additional materials such as "Tips for Applicants" that the EIC publishes. Illustratively, the EIC website advised applicants to "*write to convince evaluators [...] take time to think about a catchy title and craft an attractive abstract, [...] bear in mind that form matters, take the time to review and polish the language.*"[9] Anecdotal evidence further suggests that ventures take proposal crafting very seriously, as they often seek guidance from specialized consultants or past successful applicants to tailor their project descriptions and use phrases and keywords presumed to resonate with evaluators.[10]

[8]Illustratively, in November 2013 the European Commission released a call for independent experts to assist in the evaluation of proposals for the period 2014-2020. The pool of experts is said to be updated at the beginning of each year to meet the evaluation requirements of the EIC Accelerator call (Internet Archive, 12 April 2021).

[9]These quotes are taken from a 2021 archived version of the EIC website (Internet Archive, 12 April 2021). More recently, "Tips for Applicants" published by the EIC include additional cues for ventures. For instance, these tips mention: "*Each of the evaluators will look at 3 criteria: excellence, impact and implementation. [...] You need to convince the evaluators that your innovation is excellent, that your product, process or service is disruptive, that it has the potential to change the dynamic of the market and possibly to address a societal challenge. [...] You need to convince the evaluators that you can make it.*" (EIC Work Programme, 2025)

[10]Successful applicants with whom we engaged in informal conversations have raised this possibility. Like the evaluators, we as researchers have no information about whether a proposal was prepared with external help or advice. We assume any such involvement would be reflected in the proposal text and since our focus is on how framing (language choices) influence evaluators' assessments, the origin of specific words or framing choices, whether from the entrepreneur alone, input from consultants, or other type of assistance, is not a concern.

The application itself comprises a concise project description that follows a standardized template and is grounded in a strategic business plan that specifies expected outcomes and shows how the submission meets those criteria. Moreover, applicants also complete administrative forms with structured data on their organization, requested budget, and ethical considerations. If successful at this remote evaluation stage, entrepreneurs will be invited to pitch their projects in front of a panel jury, who will ultimately decide to award the grant funding.[11] As we explain below, our analyses focus on the remote evaluation step and leverage the short project descriptions to capture framing strategies.

We use proprietary EIC Accelerator application records, including information on the venture proposal, evaluation decision, and the timing of submission. We combine these data with the Orbis database, maintained by Bureau van Dijk (BvD), a subsidiary of Moody's and the largest global firm database, to obtain firm-level characteristics. From Orbis, we obtain the year of incorporation, industry classification, whether the firm holds a patent, and yearly performance variables, including turnover and number of employees at the time of application. We link the two data sources via unique identifiers provided by the EIC records and restrict the sample to firms that are eligible for the Accelerator program, have non missing firm-level data at the application year, and project description containing at least 25 words.[12] Our sample consists of 77,717 applicants to the EIC Accelerator, including supported ventures and rejected applicants in the years 2014 to 2020. The two largest applicant categories are ventures in Biotechnology and Life Sciences, Business Services, and Computer Software. Overall, the EIC Accelerator supports a wide array of ventures, as detailed in Appendix C, Table C.1.1.

[11] Rejected applicants may revise and resubmit within 12 months, up to three attempts. In case of re-submission, a new set of evaluators is assigned. We control for this feature, as explained in the Variables section.

[12] We exclude 13 cases reporting more than 500 employees as these firms are ineligible for funding, 380 proposals containing fewer than 25 words (a minimum length commonly adopted in prior work to ensure reliable text-based indicators, e.g., Markowitz, 2019), 85 observations with incomplete data, and 125 duplicated entries.

Variables

Rejected. We define our outcome of interest as a binary variable that equals one if a venture is rejected during the EIC Accelerator remote evaluation process, and zero otherwise. We focus on rejection at the remote assessment stage rather than the jury's final funding decision, since that later stage involves additional, unobserved cues including the oral presentation, extensive Q&A, and supporting materials such as pitch decks. Obviously, all these additional information may play a substantial role in determining whether a venture is ultimately awarded support, beyond the framing aspect. In our sample, 70.8% of applicants are rejected, 22.5% are invited to the interview but receive no funding, and only 6.7% ultimately secure EIC Accelerator support, underscoring the process's highly competitive nature.[13]

Framing. Consistent with the notion that "*to frame is to select some aspects of a perceived reality and make them more salient in a communicating text*" (Entman, 1993), we measure the emphasis that ventures place in their offerings, as reflected by the language choices in their proposals. We use a validated dictionary of "promotional adjectives" that independent studies have robustly linked to funding success in the context of scientific funding (Millar et al., 2022; Peng et al., 2024). Multiple validation tests suggest that words in this dictionary activate attention and cognitive engagement, helping to convey innovative ideas more powerfully than neutral synonyms (Peng et al., 2024, p.2). Following recent work on entrepreneurial framing (Pan et al., 2020; Taeuscher and Rothe, 2024), we calculate two orthogonal variables (*Innovativeness Framing* and *Feasibility Framing*) as the share of words in the proposal's descriptions that match the vocabulary associated with each of these two framing categories, and standardize these shares by cutoff deadline and industry. Examples of words aligned with *Innovativeness Framing* include "groundbreaking," "revolutionary," and "unprecedented." Examples of words aligned with *Feasibility Framing* include "compatible," "rigorous," and "robust."

Put differently, our two framing variables capture the extent to which

[13] We report the funding success rate by industry in the Appendix C, Table C.1.1.

a proposal places emphasis on certain aspects of the value proposition relative to other ventures reviewed at the same time and that belong to the same industry. In Appendix C.4, we list the words associated with each of the framing categories, specific anonymized cases of funded ventures, word clouds of the most frequent keywords in our data (Figures C.4.1a and C.4.1b), and further descriptive statistics across industries (Figure C.4.2). Moreover, we include a binary variable in cases where neither *Feasibility Framing* nor *Innovativeness Framing* is used by the applicant, and we label this variable as *Alternative Framing*. As discussed in the Robustness Checks section, we further validate our approach by running placebo tests with alternative word lists to define framing strategies, and by repeating our analyses using alternative operationalizations of our framing variables (e.g., raw shares without normalization by application deadline or industry).

Congestion Index. We proxy the congestion in the evaluation process by examining the variation in the number of submissions over time and across industries that exceed what would be predicted by time trends or industry-specific opportunities. Building on prior work concerned with the role of examiner's overload in the patent approval process (Ferguson and Carnabuci, 2017; Chaffin et al., 2013), we use an auxiliary regression at the industry-month level to identify abnormal submission volumes. We aggregate applicant-level submissions into a panel of 30 industries across 61 monthly cutoff deadlines, and regress the log-transformed and demeaned submission count on industry and month-year fixed effects, obtaining our *Congestion Index* as the predicted residuals of this auxiliary regression.[14]

Three features of the EIC Accelerator's selection process justify interpreting this *Congestion Index* as the overload faced by evaluators in a given assessment batch. First, the evaluator pool is recruited before the call announcing the cutoff deadlines; second, proposals are assigned to evaluators based on technological/industry expertise (and not workload balancing); and third, evaluators are contractually required to submit a standardized assessment within a fixed time window. Consequently, when faced with an ab-

[14]We use log-transformation on the dependent variable to account for its high skewness and use the demeaned version to account for industry size dependency.

normal number of proposals, a fixed evaluator pool must process these files under congestion.

One potential concern is that ventures could strategically time their submissions if congestion were predictable. We do not believe this concern holds in our setting for three reasons. First, in theory, if applicants could reliably anticipate evaluator load, then congestion would dissipate in equilibrium, as firms would reallocate their submissions toward less crowded deadlines. The persistence of substantial variation in application volumes therefore suggests that such predictability is limited. Second, information frictions restrict applicants' ability to form expectations about congestion. While the set of funded ventures and aggregate success rates are publicly reported after each round, no information is disclosed about rejected applications. Thus, even when applicants observe that success rates are lower in certain rounds, they cannot identify whether this reflects greater participation from relevant competitors in their own industry or technology domain. This lack of visibility makes it difficult for firms to forecast future congestion in a strategically useful way. Third, our descriptive analyses confirm that congestion reflects idiosyncratic rather than predictable patterns. We provide further validation and details on the construction of this variable in the Appendix, C.5, where we show that submission volumes exhibit large swings across industries and deadlines without cyclical regularities, and that observable applicant characteristics do not significantly predict our congestion measure. Consequently, the evidence does not support the concern that certain ventures systematically sort themselves into high- or low-congestion periods.

These points suggest that ventures are unlikely to anticipate congestion strategically, and that our measure can be reasonably interpreted as capturing variation in evaluator load that is not systematically related to the characteristics of individual ventures.[15]

Contextual Novelty. We process the rich narratives in ventures' project

[15]Illustratively, across our sample, we observe that the number of submissions per industry across cutoff deadlines ranges from 1 to 980, with a mean of 65. Moreover, the coefficient of variation in the number of submissions ranges from 0.31 for Industrial, Electric, and Electronic Machinery (on average, 197 applications by call) to 0.63 for Transport, Freight & Storage (on average, 9 applications by call).

descriptions using natural language processing techniques to measure what we term "contextual novelty." Building on the concept that "*individuals make evaluations of atypicality through relative judgments, comparing each object under consideration against a set of others*" (Cutolo and Ferriani, 2024, p.163), we define contextual novelty as the degree to which a venture's offering differs from others within the same EIC Accelerator call. To operationalize this measure, we generate vector representations of all proposals using the BERT language model (Devlin et al., 2019) and calculate the maximum cosine distance between each venture and all others in the same application round. We define *Contextually Novel* as a binary variable indicating whether a venture falls in the top quartile (above the 75th percentile) of this distance distribution for its evaluation round, and zero otherwise. This identifies ventures that stand out most prominently across all proposals evaluated concurrently by the same evaluator pool.[16]

Control variables. Throughout our analyses, we account for the following characteristics measured at the year of application. We control for the (log-transformed) *Firm age*, as more mature ventures likely have experience in seeking funding and have had more time to develop their technologies and value proposition. Moreover, because the program is open for both start-ups and SMEs, we also account for the (log-transformed) number of *Employees*, as larger firms likely accumulate, and have access to a larger set of resources. Moreover, given prior research suggesting that evaluators are sensitive to a range of cues, including venture characteristics that may imperfectly signal status or quality (Hoenig and Henkel, 2015; Quy and Huy, 2007), we introduce a binary variable that equals one for cases in which the applicant has reported *Revenue* during or before the application year, and zero otherwise. We also introduce a binary variable for cases in which the venture reports patents (*Patent*), and zero otherwise. We also control for the overall *Proposal Word Count* (in hundreds of words), the key source describing the venture

[16]In supplementary analysis, we also use *Contextually Typical* as a binary variable equal to one for ventures in the bottom quartile (below percentile 25) of the atypicality distribution, and zero otherwise. This measure shows a coherent pattern of results, showing coefficients in the opposite direction. We provide further details in the Robustness Checks section.

offerings, where shorter (longer) proposals may describe more narrowly (extensively) describe the project. To account for the technological scope and ambition, we further control for the (log-transformed) *Requested Grant*, and we add an indicator variable that equals one for those that exceed €2.5 million, and zero otherwise *Exceptional Grant Requests*. To account for the competitive pressure in an evaluation round, we measure the (log-transformed) *Number of Applicants* at the cutoff level. In addition, because the application process allows ventures to re-submit in case of a rejection, we include an indicator variable to account for applicants who attempted *Multiple submissions*.[17] Finally, throughout all our analyses we include *Industry*, *Country*, and *Call* fixed effects.[18]

4.4 Results

Descriptive Analyses

We report summary statistics for the analytic sample in Table 4.1. The average applicant has 10 employees and is 7.5 years old at the time of application.[19] About 40% have ever obtained a patent and 60% report revenue, while more than 80% of applicants attempt multiple submissions. The project description, the core piece of information for remote evaluators, contains on average 260 words. Overall, 1,683 applicants enter the evaluation round of a given call cutoff, albeit with substantial variation (SD = 951). Regarding framing, it is worth noting that while *Innovativeness* and *Feasibility Framing* are orthogonal, implying that a venture can in principle emphasize both aspects, in practice, the correlation is negligible (0.04).

Table 4.2 reports further descriptive statistics and t-tests contrasting proposals above and below the median on *Innovativeness* and *Feasibility Framing*. On one hand, above median *Innovativeness Framing* is on average as-

[17] We only observe the project characteristics of the last submission of a given applicant, and thus we are unable to study the same applicant at different points in time.

[18] Our results are robust to replacing the *Call* fixed effects by *Application Year* fixed effects. These supplementary results are not reported but available on request.

[19] Roughly half of the ventures in our sample have five or fewer years since incorporation at the time of application, and over a quarter have two or fewer years.

sociated with younger ventures, which submit shorter proposals and request less funding. On the other hand, above median *Feasibility Framing* is on average associated with older and larger ventures, which are more likely to hold patents, report prior revenue, and submit longer proposals. Moreover, we find that above median *Feasibility Framing* is associated with a lower proportion of firms being Rejected, whereas no such relationship is found for *Innovativeness Framing*.

We also observe commonalities across framing strategies. First, ventures showing above the median on either *Innovativeness* or *Feasibility Framing* are disproportionately more likely to qualify as contextually novel (*High Atypicality*), and are more likely to submit multiple applications. Crucially, neither framing emphasis shows a statistically significant difference in the *Congestion Index*. These nuanced patterns further motivate the multivariate regression analyses in the next section, where we jointly estimate the effects of framing for contextually novel ideas on their evaluation outcomes, while controlling for a range of confounding factors.

Evaluation Outcome

We estimate the probability of rejection in EIC Accelerator selection process estimating the relationship between the evaluation outcome and three focal factors: (i) the venture's framing strategy (*Innovativeness Framing* and *Feasibility Framing*), (ii) a venture's contextual novelty relative to competing proposals (*High Atypicality*), and (iii) unusually high volume of applications that evaluators must process in a fixed time (Congestion Index). In Table 4.3 and Table 4.4 we report estimates from linear probability models employing robust standard errors clustered by application cutoff, and including fixed effects for *Call*, *Country*, and *Industry*, as well as the control variables previously described. Table 4.3 presents full-sample results and tests hypotheses H1-H3. In Table 4.4 we present the split-sample analysis contrasting evaluation outcomes of venture proposals reviewed under high- and low- congestion, as defined by the percentile 75 and 25 of the *Congestion Index*, respectively. The interaction terms of Table 4.4 allow for testing hypotheses H3 and H4.

Our baseline hypothesis (H1) proposes that the novel ideas presented by innovative ventures benefit from a framing that emphasizes feasibility, in line with prior research. In Table 4.3, Column 7, we observe a statistically significant and negative effect for *Feasibility Framing*, suggesting that evaluators are less likely to reject proposals from ventures that emphasize the feasibility of their idea. Our approach, which allows capturing alternative framing strategies, provides insight into alternative framing strategies. More specifically, also *Innovativeness Framing* (Column 6) is associated with a lower chance of rejection, while ventures whose framing is not focused on either Innovativeness or Feasibility are, on average, more likely to be rejected. A formal test of the difference between the coefficients of *Feasibility Framing* and *Innovativeness Framing* suggests these coefficients are statistically equivalent.[20] These results bring nuance to existing knowledge on framing strategies, showing that not only stressing feasibility but also underscoring the innovativeness of the value proposition may be effective. The results also indicate that a lack of clear framing hurts ventures' odds of making it through the initial screening stage. Finally, note that ideas that stand out more vis-à-vis the competition in the submission cohort, i.e., ideas that are contextually novel, are consistently associated with a lower rejection probability (Columns 2-8).

We subsequently hypothesized that contextually novel ideas would benefit from a framing that emphasizes feasibility (H2), based on the logic that the benefits of *Feasibility Framing* apply a fortiori when setting a higher bar for novelty. However, the interaction term is not statistically significant (Column 7), and neither is the interaction with *Innovativeness Framing* (column 6). Combined with the main effects for *High Atypicality*, *Innovativeness Framing*, and *Feasibility Framing*, which remain significant, the interaction terms indicate that neither framing provides an additional evaluative premium for contextually novel ideas. One possible explanation is that in the EIC Accelerator's setting —where novelty is an explicit selection criterion— evaluators may already expect atypical proposals to demonstrate feasibility

[20] We test that the difference between the coefficient of *Feasibility Framing* and *Innovativeness Framing* is zero, failing to reject the null hypothesis ($F(1, 40) = 0.27$).

as a baseline, making additional emphasis less differentiating. Another possibility is that, under normal workload conditions, evaluators have sufficient cognitive bandwidth to assess the feasibility of novel ideas based on substantive content rather than relying on linguistic cues, reducing the marginal benefit of *Feasibility Framing*. We further address the role of workload in the test of Hypothesis 4 below.

Our third hypothesis (H3) proposed that congestion in the evaluation process lowers the rejection probability of contextually novel ideas. As reported in Table 4.3 (Column 8), we find a statistically significant negative interaction between the Congestion Index and *High Atypicality*, suggesting that congested evaluators are less likely to reject contextually novel ideas. This finding aligns with our theorizing that, under time constraints, EIC evaluators do not apply a heuristic that favors familiar, lower-risk proposals. Rather, they appear to adopt a mission-oriented rationale, prioritizing ventures that better fit the program's strategic emphasis on breakthrough potential. In line with our earlier theorizing, when pressed for time, evaluators may rely more on category-level signals, such as the distinctiveness of a proposal relative to its peers, to identify candidates worth advancing. In this setting, contextual novelty may function as a salient shortcut for "high-impact potential", offsetting the cognitive burden it typically imposes.

This result also reflects evaluators' professional identity and mandate: as external experts tasked with identifying transformative ideas, they may consciously lean toward atypical proposals when forced to prioritize. Finally, in terms of our control variables, we find that larger firms, longer project proposals, and larger grant requests, in addition to reporting patents or revenue, lowers the probability of rejection.[21]

Turning to the split-sample analyses, the interaction between *Feasibility Framing* and contextual novelty under high congestion (Table 4.4, Column 7) is negative and statistically significant, supporting H4. This result suggests

[21] Our results are consistent when estimated with logistic regressions (reported in the Appendix, Table C.3.1 and Table C.3.2), and also when using either clustered errors by industry or call, and heteroskedasticity-consistent standard errors without clustering (these additional results are not reported but available from authors).

that *Feasibility Framing* helps contextually novel ideas overcome evaluation hurdles when cognitive resources are most constrained. As we argued in the theory, concrete feasibility cues provide evaluators with a rapid, low-effort way to translate an atypical proposal into a credible implementation pathway, reducing the interpretive uncertainty that novelty creates. Under normal workload conditions, evaluators can devote more attention to unpacking the substance of an unusual idea, making linguistic cues less critical. Under congestion, however, such cues may become decisive by clarifying "how" an unconventional solution could work in practice. In other words, that *Feasibility Framing* reassures evaluators about execution risk, allowing them to justify advancing a high-novelty proposal even when less time is available to reflect on technical and market details.

Taken together, our results suggest that both framing strategies and contextual factors shape the evaluation of novel ideas, but their effects are not uniform. We find that feasibility and *Innovativeness Framing* each independently reduce the likelihood of rejection, yet contrary to H2, *Feasibility Framing* does not provide an additional advantage for contextually novel ideas under normal workload conditions. This may reflect evaluators' baseline expectations for feasibility in atypical proposals, or the ability to assess feasibility from substantive content when cognitive resources are sufficient. In line with H3, congestion reduces the rejection rate of contextually novel ideas, consistent with the notion that evaluators under time pressure may lean on novelty as a shortcut for identifying high-impact ventures.

Finally, supporting H4, *Feasibility Framing* becomes particularly effective for contextually novel ideas in high-congestion contexts, likely because concrete cues help evaluators quickly make sense of unconventional offerings and assess their implementation potential. These patterns highlight that the benefits of framing depend not only on what is emphasized but also on when and under what conditions it is used. The interplay between message content and evaluation context offers a richer, more contingent view of how novel ideas can be advanced in competitive selection processes.

Robustness Checks

To verify that our *Innovativeness* and *Feasibility Framing* variables capture substantive message cues rather than generic promotional language, we repeat the analysis using two unrelated adjective lists that may also resonate in the evaluation process.

First, from the same curated vocabulary in (Peng et al., 2024), we measure the emphasis on usability (e.g., key words such as "accessible," "seamless," and "user-friendly"), as this may signal market potential and/or costumer-orientation in the value proposition, which arguably also aligns with concrete-oriented framings that have been shown to elicit support from investors (Falchetti et al., 2022).

Second, using a large language model we generate a list of words that aim to capture the emphasis a venture places on *team capabilities*.[22] Examples of such keywords frequently found across proposals include "training", "qualified," and "partners." Highlighting team capabilities aligns with some of the guidelines given by the EIC to prospective applicants in recent years (EU, 2024), and also with prior literature suggesting that funders react to information on founding team characteristics (Bernstein et al., 2017).

In these supplementary regressions, we apply similar cutoff- and industry-standardization, and repeat both full- and split-sample analyses. We find that neither placebo framings have a strong and consistent statistical significance in the probability of rejection (Appendix C, Tables C.2.1-C.2.2). That said, we do observe a negative albeit weakly significant negative coefficient associated with *Team Capability Framing*, which aligns with prior work. Overall, these supplementary analyses add confidence to the framing measures adopted in our main analyses and our argument that they capture linguistic cues that evaluators genuinely respond to.

[22]Specifically, we ask ChatGPT-4o to "generate a list of keywords designed to identify references to human capital and/or team capabilities in a proposal text," which we manually curate.

4.5 Conclusion

We study evaluation outcomes of real venture proposals submitted to a high-profile accelerator, analyzing how strategic emphasis in venture descriptions influences funding success. Building on the entrepreneurial framing literature, we investigate how context shapes both the appreciation of novelty and the effectiveness of linguistic framing strategies, two issues that prior research has largely overlooked in observational settings. Our analysis focuses on two critical contextual factors: (i) an idea's novelty relative to competing peers, and (ii) evaluators' workload during the assessment process. Overall, these factors prove consequential in our setting, where evaluators must review numerous applications of technology-based ventures under strict time constraints.

Specifically, our results provide a nuanced understanding of how ventures can strategically frame novel ideas to enhance their chances in competitive selection processes. While feasibility and innovativeness framing both reduce rejection risk, their effectiveness depends on the evaluation context. Contrary to expectations, feasibility framing does not further advantage contextually novel ideas under normal workloads, suggesting that evaluators may already expect feasibility to be demonstrated, or can discern it without relying on promotional language, when cognitive resources are sufficient. The lack of a significant association between contextual novelty and feasibility framing under these conditions further reinforces the context-dependent nature of the effectiveness of feasibility framing. It also suggests that specific institutional missions or evaluation procedures may reduce evaluators' reliance on such cues, particularly under routine workload pressures.

By contrast, high congestion appears to shift evaluators' decision logic: rather than defaulting to familiar, lower-risk proposals, they give greater consideration to atypical ideas, consistent with the EIC Accelerator's mission to select breakthrough innovations. In this high-pressure environment, feasibility framing becomes a valuable tool for helping evaluators make quick, confident judgments about unconventional proposals. These insights extend existing research on entrepreneurial framing and novelty evaluation by show-

ing that message effects are contingent not only on audience characteristics but also on situational constraints, such as workload.

Our analysis makes several theoretical contributions. First, our study advances the entrepreneurial framing literature by moving beyond the traditional "abstract versus concrete" dichotomy. We operationalize framing as a graded construct in which both feasibility-oriented (concrete) and innovativeness -oriented (abstract) elements can be present and influential. Our results show that these two emphases are not mutually exclusive and can independently enhance the likelihood of support, challenging assumptions that concrete framing is categorically superior for novel ideas. Second, we extend theory on the evaluation of novelty by introducing contextual novelty—the relative atypicality of a proposal within its competitive cohort—as a moderator of framing effectiveness. This approach captures the inherently comparative nature of novelty assessments.

Third, we contribute to the literature on decision-making under constraint by showing that evaluator congestion can increase support for contextually novel ideas when organizational objectives prioritize breakthrough potential. This finding runs counter to prevailing assumptions that cognitive overload leads to conservative, familiarity-biased decisions. Our results suggest that under mission-oriented mandates, congestion may activate heuristics that favor distinctiveness rather than penalize it. Finally, by documenting that feasibility framing particularly benefits contextually novel ideas under high congestion, we identify a boundary condition for the effectiveness of message cues. This highlights the contingent nature of framing strategies and integrates insights from Construal Level Theory into real-world, high-stakes evaluation settings.

Our findings offer tractable implications for both entrepreneurs and accelerators. For entrepreneurs and early-stage ventures, our results underscore the importance of tailoring framing strategies to the evaluation environment. While emphasizing either feasibility or innovativeness can help, ventures should recognize that evaluators may respond differently under varying workload conditions. In particular, for unconventional proposals com-

peting in crowded evaluation rounds, explicitly signaling feasibility through concrete, implementation-focused language can make a decisive difference. For funding bodies and accelerator program designers, the findings highlight the role of evaluation context in shaping selection outcomes. Monitoring workload levels and ensuring evaluators have sufficient cognitive bandwidth remains important for quality decision-making. However, in contexts where congestion is unavoidable, clear communication of selection priorities (e.g., breakthrough potential) and encouraging applicants to clarify feasibility may help evaluators identify and support high-novelty ventures.

Our study has several limitations that open avenues for future research. First, the proprietary dataset we analyze lacks evaluator-level information, preventing us from analyzing how individual expertise, cognitive distance, or personal workload shape responsiveness to framing. Future work could link proposal content to individual evaluator characteristics to explore these mechanisms more directly. Along the same lines, information on proposal revisions between selection rounds or upon resubmission after rejection would allow identifying learning effects and controlling for venture-level fixed effects. Second, our setting is a large, mission-driven public accelerator. While ideal for studying novelty evaluation under explicit breakthrough mandates, results may differ in private-sector or sector-specific selection contexts where incentives and evaluative norms vary (Petty et al., 2023; Scott et al., 2020). Comparative studies across organizational types could test the generalizability of our findings.

Third, our measure of framing captures the relative use of promotional adjectives linked to feasibility or innovativeness, which offers precision but may miss more subtle rhetorical strategies. Advances in large language model (LLM)–based text analysis (Csaszar et al., 2024) could capture richer linguistic features and their interaction with contextual novelty. LLMs offer promising tools to refine and extend our measure of contextual novelty. While our current approach captures linguistic distinctiveness among proposals within the same evaluation round, future research could incorporate the technological attributes of a broader set of relevant ventures, providing a more accurate

reference point. For instance, LLMs could be fine-tuned on domain-specific cases using a more complete description of prior funded and unfunded technologies to embed proposals within a fine-grained domain. Such an approach would capture not only how proposals differ linguistically from their direct competitors but also how they differ in their technical attributes, providing a more comprehensive understanding of how evaluators may value them. Validating such AI-based measures of novelty remains a critical direction for future research, for instance, through expert assessments or crowd-based evaluations that benchmark automated measures against human judgments. In parallel, further research could examine how evaluators respond to higher congestion and uncover the cognitive mechanisms underlying these effects. For example, experimental studies could investigate how congestion affects evaluators' attention allocation, the use of heuristics, or their sensitivity to framing cues, thereby providing deeper insight into how workload pressures influence evaluation behavior.

Fourth, because the ventures in our sample are early stage and the quality of a new venture idea is highly uncertain at this stage (Bergemann and Hege, 1998, 2005; Scott et al., 2020), additional information cues beyond those included in our analyses—such as entrepreneurs' past experience or networks—could further explain how remote evaluators value applicants. We therefore acknowledge that unobserved factors, including founders' identity, could bias our results if they influence both framing choices and evaluation outcomes. Although these factors are not necessarily available to evaluators at the time of assessment, we remain cautious and interpret the estimated relationships as associations rather than causal effects. Finally, future research could examine whether the framing strategies that help secure selection also predict post-award outcomes, such as venture growth, survival, or downstream innovation performance. This would clarify whether framing simply aids in "getting through the gate" or has lasting implications for venture success.

4.6 Tables and Figures

Table 4.1: Summary Statistics

Variable	Description	Mean	SD	Min	Max
Rejected	Binary equal to 1 if the venture was rejected	0.71	0.46	0.00	1
Firm age	Years elapsed between application year and year of incorporation	7.54	8.04	0.00	44
Employees	Number of employees at year of application	10.10	23.40	0.00	441
Revenue	Binary equal to 1 if the venture reported revenue before the application year	0.61	0.49	0.00	1
Patent	Binary equal to 1 if the venture has ever obtained a patent	0.41	0.49	0.00	1
Multiple Submissions	Binary equal to 1 if the venture submitted multiple applications to the accelerator	0.83	0.38	0.00	1
Proposal Word Count	Word count of the submitted proposal (in 100s)	2.62	0.58	0.25	4
Requested Grant	Grant amount requested by the applicant (in 1000s EUR)	783	1022	0.05	44,488
Exceptional Grant	Binary equal to 1 if the requested grant exceeds 2.5 million EUR	0.02	0.14	0.00	1
Congestion Index	Residualized number of applicants exceeding what trend and industry factors predict	0.03	0.21	-1.94	0.96
Applications per Call	Number of applicants per call cutoff deadline	1683	951	4.00	3,576
Innovativeness Framing	Share of words in the proposal that emphasize innovativeness, standardized by industry and cutoff	0.35	0.71	-1.40	15
Feasibility Framing	Share of words in the proposal that emphasize feasibility, standardized by industry and cutoff	0.41	0.72	-1.10	10
Alternative Framing	Binary equal to 1 if none of the words in the proposal matches the Innovativeness or Feasibility Framing	0.23	0.42	0	1
Contextually Novel	Binary equal to 1 for ventures in the top quartile of the distribution of atypicality	0.25	0.43	0	1

Note: $N = 77,717$.

Table 4.2: Differences in Means, Above- and Below-Median Framing

Panel A: Innovativeness Framing

Variable	> Median		< Median		
	mean	(sd)	mean	(sd)	t-value
Rejected (0/1)	0.71	(0.46)	0.71	(0.45)	-0.65
Firm Age	7.44	(7.92)	7.61	(8.13)	-2.98
Employees	10.00	(23.50)	10.20	(23.30)	-1.12
Revenue (0/1)	0.61	(0.49)	0.61	(0.49)	1.39
Patent (0/1)	0.41	(0.49)	0.41	(0.49)	-1.65
Proposal Word Count	2.59	(0.53)	2.64	(0.62)	-12.50
Requested Grant	752	(1031)	810	(1030)	-7.88
Exceptional Grant (0/1)	0.02	(0.15)	0.02	(0.14)	0.36
Multiple Submissions (0/1)	0.84	(0.37)	0.82	(0.38)	4.93
Applicants per Cutoff	1668	(960)	1696	(943)	-4.45
Congestion Index	0.030	(0.21)	0.033	(0.21)	-1.90
Contextually Novel (0/1)	0.26	(0.44)	0.24	(0.43)	4.28

Panel B: Feasibility Framing

Variable	> Median		< Median		
	mean	(sd)	mean	(sd)	t-value
Rejected (0/1)	0.68	(0.47)	0.73	(0.45)	-12.80
Firm Age	7.65	(8.09)	7.46	(8.00)	3.18
Employees	10.50	(23.50)	9.91	(23.40)	3.37
Revenue (0/1)	0.62	(0.49)	0.60	(0.49)	5.08
Patent (0/1)	0.40	(0.49)	0.41	(0.49)	8.75
Proposal Word Count	2.72	(0.45)	2.55	(0.65)	41.30
Requested Grant	795	(1014)	775	(1027)	2.69
Exceptional Grant (0/1)	0.02	(0.15)	0.02	(0.14)	3.04
Multiple Submissions (0/1)	0.84	(0.37)	0.80	(0.39)	11.00
Applicants per Call	1666	(949)	1695	(952)	-4.17
Congestion Index	0.031	(0.20)	0.032	(0.21)	-0.76
Contextually Novel (0/1)	0.26	(0.44)	0.24	(0.43)	4.22

Table 4.3: Analysis of Rejection Probability

DV = Rejected (0/1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(1+Firm age)	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
ln(1+Employees)	-0.015***	-0.014***	-0.014***	-0.014***	-0.014***	-0.014***	-0.014***	-0.014***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Revenue (0/1)	-0.020***	-0.020***	-0.021***	-0.020***	-0.020***	-0.020***	-0.020***	-0.020***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Patent (0/1)	-0.097***	-0.097***	-0.096***	-0.096***	-0.096***	-0.096***	-0.096***	-0.096***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Multiple Submissions (0/1)	-0.153***	-0.152***	-0.151***	-0.151***	-0.151***	-0.151***	-0.151***	-0.150***
	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Proposal Word Count	-0.055***	-0.052***	-0.050***	-0.050***	-0.052***	-0.051***	-0.052***	-0.052***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
ln(Requested Grant)	-0.047***	-0.047***	-0.047***	-0.047***	-0.047***	-0.047***	-0.047***	-0.047***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Exceptional Grant (0/1)	0.061***	0.061***	0.062***	0.061***	0.061***	0.061***	0.061***	0.062***
	[0.016]	[0.016]	[0.016]	[0.015]	[0.015]	[0.015]	[0.016]	[0.015]
Congestion Index	-0.005	-0.006	-0.007	-0.006	-0.006	-0.006	-0.006	0.005
	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.014]
ln(Applicants per Call)	-0.011	-0.011	-0.010	-0.010	-0.010	-0.010	-0.010	-0.010
	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]
Contextually Novel (0/1)		-0.026***	-0.026***	-0.026***	-0.026***	-0.025***	-0.024***	-0.025***
		[0.005]	[0.005]	[0.005]	[0.005]	[0.006]	[0.005]	[0.005]
Alternative Framing (0/1)			0.026***	0.025***	0.021***	0.021***	0.021***	0.021***
			[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Innovativeness Framing			-0.004**		-0.005**	-0.004*	-0.005**	-0.005**
			[0.002]		[0.002]	[0.002]	[0.002]	[0.002]
Feasibility Framing (*H1*)				-0.006***	-0.007***	-0.007***	-0.006***	-0.007***
				[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Contextually Novel × Innovativeness Framing						-0.005		
						[0.006]		
Contextually Novel × Feasibility Framing (*H2*)							-0.005	
							[0.006]	
Contextually Novel × Congestion Index (*H3*)								-0.048**
								[0.021]
Constant	1.521***	1.520***	1.510***	1.509***	1.518***	1.516***	1.517***	1.520***
	[0.122]	[0.120]	[0.118]	[0.119]	[0.119]	[0.118]	[0.119]	[0.118]
Call FE	y	y	y	y	y	y	y	y
Country FE	y	y	y	y	y	y	y	y
Industry FE	y	y	y	y	y	y	y	y
N	77717	77717	77717	77717	77717	77717	77717	77717
adj. R^2	0.138	0.139	0.140	0.140	0.140	0.140	0.140	0.140

Notes: Linear probability estimates. Robust standard errors clustered at the application cutoff level in brackets. *p<0.10, **p<0.05, ***p<0.01.
H1, *H2*, and *H3* refer to Hypothesis 1, Hypothesis 2, and Hypothesis 3, respectively.

Table 4.4: Analysis of Rejection Probability, Split by Evaluator Congestion

	Low Congestion			High Congestion		
DV=Rejected(0/1)	(1)	(2)	(3)	(4)	(5)	(6)
ln(1+Firm age)	-0.003	-0.003	-0.003	0.004	0.004	0.004
	[0.004]	[0.004]	[0.004]	[0.005]	[0.005]	[0.005]
ln(1+Employees)	-0.011***	-0.011***	-0.011***	-0.017***	-0.017***	-0.017***
	[0.003]	[0.003]	[0.003]	[0.004]	[0.004]	[0.004]
Revenue (0/1)	-0.032***	-0.031***	-0.032***	-0.020**	-0.020**	-0.019**
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Patent (0/1)	-0.093***	-0.093***	-0.093***	-0.082***	-0.082***	-0.082***
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Multiple Submissions (0/1)	-0.143***	-0.143***	-0.143***	-0.163***	-0.163***	-0.163***
	[0.013]	[0.013]	[0.013]	[0.011]	[0.011]	[0.011]
Proposal Word Count	-0.054***	-0.053***	-0.054***	-0.062***	-0.062***	-0.062***
	[0.006]	[0.006]	[0.006]	[0.007]	[0.007]	[0.007]
ln(Requested Grant)	-0.056***	-0.056***	-0.056***	-0.032***	-0.032***	-0.032***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Exceptional Grant Request (0/1)	0.061*	0.061*	0.061*	0.062**	0.062**	0.062**
	[0.033]	[0.033]	[0.033]	[0.025]	[0.025]	[0.025]
Congestion Index	-0.028	-0.028	-0.028	0.030	0.030	0.030
	[0.035]	[0.035]	[0.035]	[0.027]	[0.027]	[0.027]
ln(Applicants per Call)	0.012	0.012	0.012	-0.021	-0.021	-0.021
	[0.022]	[0.022]	[0.022]	[0.019]	[0.019]	[0.019]
Contextually Novel (0/1)	-0.014	-0.010	-0.015*	-0.033***	-0.036***	-0.025**
	[0.008]	[0.007]	[0.008]	[0.011]	[0.011]	[0.010]
Alternative Framing (0/1)	0.022**	0.022***	0.022***	0.008	0.008	0.008
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Innovativeness Framing	-0.013***	-0.011**	-0.013***	-0.011***	-0.013***	-0.011***
	[0.005]	[0.005]	[0.005]	[0.003]	[0.004]	[0.003]
Feasibility Framing	-0.005	-0.006	-0.006	-0.009***	-0.009***	-0.005
	[0.004]	[0.004]	[0.004]	[0.003]	[0.003]	[0.003]
Contextually Novel × Innovativeness Framing		-0.011			0.011	
		[0.013]			[0.011]	
Contextually Novel × Feasibility Framing *(H4)*			0.003			-0.019**
			[0.010]			[0.009]
Constant	1.551***	1.547***	1.553***	1.316***	1.320***	1.313***
	[0.094]	[0.094]	[0.095]	[0.071]	[0.071]	[0.070]
Call FE	y	y	y	y	y	y
Country FE	y	y	y	y	y	y
Industry FE	y	y	y	y	y	y
N	19384	19384	19384	19436	19436	19436
adj. R^2	0.123	0.123	0.123	0.165	0.165	0.165

Notes: Linear probability model estimates. Robust standard errors clustered at the application cutoff level in brackets.
Low (High) Congestion is defined as *Congestion Index* values below (above) the 25 (75) percentile, equivalent to -0.08 (0.14).
H4 refers to Hypothesis 4. *p<0.10, **p<0.05, ***p<0.01.

References

Amabile, T. M. (1996): *Creativity in Context: Update to the Social Psychology of Creativity*, Taylor and Francis, 1 ed.

Anglin, A. H., S. W. Reid, and J. C. Short (2023): "More Than One Way to Tell a Story: A Configurational Approach to Storytelling in Crowdfunding," *Entrepreneurship: Theory and Practice*, 47, 461–494.

Berg, J. M. (2016): "Balancing on the Creative Highwire: Forecasting the Success of Novel Ideas in Organizations," *Administrative Science Quarterly*, 61, 433–468.

Bergemann, D. and U. Hege (1998): "Venture capital financing, moral hazard, and learning," *Journal of Banking & Finance*, 22, 703–735.

——— (2005): "The Financing of Innovation: Learning and Stopping," *The RAND Journal of Economics*, 36, 719–752.

Bernstein, S., A. Korteweg, and K. Laws (2017): "Attracting Early-Stage Investors: Evidence from a Randomized Field Experiment," *Journal of Finance*, 72, 509–538.

Boudreau, K. J., E. C. Guinan, K. R. Lakhani, and C. Riedl (2016): "Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science," *Management Science*, 62, 2765–2783.

Chaffin, T. D., R. Heidl, M. Chari, and R. Calantone (2013): "The Role of Examiner Workload and Applicant Reputation in Intellectual Property Protection," in *Academy of Management Proceedings*, vol. 2013, 1442–1447.

Chai, S. and A. Menon (2019): "Breakthrough recognition: Bias against novelty and competition for attention," *Research Policy*, 48, 733–747.

Christensen, C. M., R. McDonald, E. J. Altman, and J. E. Palmer (2018): "Disruptive Innovation: An Intellectual History and Directions for Future Research," *Journal of Management Studies*, 55, 1043–1078.

Cohen, S., D. C. Fehder, Y. V. Hochberg, and F. Murray (2019): "The design of startup accelerators," *Research Policy*, 48, 1781–1797.

Cornelissen, J. P. and M. D. Werner (2014): "Putting Framing in Perspective: A Review of Framing and Frame Analysis across the Management and Organizational Literature," *Academy of Management Annals*, 8, 181–235.

Criscuolo, P., L. Dahlander, T. Grohsjean, and A. Salter (2017): "Evaluating novelty: The role of panels in the selection of R&D projects," *Academy of Management Journal*, 60, 433–460.

Csaszar, F. A., H. Ketkar, and H. Kim (2024): "Artificial Intelligence and Strategic Decision-Making: Evidence from Entrepreneurs and Investors," *Strategy Science*.

Cutolo, D. and S. Ferriani (2024): "Atypicality: Toward an Integrative Framework in Organizational and Market Settings," *Academy of Management Annals*, 18, 157–209.

Deichmann, D., G. Cattani, and S. Ferriani (2024): "Recognizing, evaluating, and selecting new ideas: the problematic journey of novelty," *Innovation: Organization and Management*, 26, 349–358.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019): "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint arXiv:1810.04805*.

Dutta, S., T. B. Folta, and J. Rodrigues (2022): "Do Governments Fund the Best Entrepreneurial Ventures? The Case of the Small Business Innovation Research Program," *Academy of Management Discoveries*, 8, 103–138.

Entman, R. M. (1993): "Framing: Toward Clarification of a Fractured Paradigm," *Journal of Communication*, 43, 51–58.

European Union (2024): "European Innovation Council: guide for applicants," .

Falchetti, D., G. Cattani, and S. Ferriani (2022): "Start with "Why," but only if you have to: The strategic framing of novel ideas across different audiences," *Strategic Management Journal*, 43, 130–159.

——— (2024): "Radically concrete or incrementally abstract? The contingent role of abstract and concrete framing in pitching novel ideas," *Innovation: Organization and Management*, 26, 473–493.

Ferguson, J.-P. and G. Carnabuci (2017): "Risky recombinations: Institutional gatekeeping in the innovation process," *Organization Science*, 28, 133–151.

Fleming, L. (2001): "Recombinant uncertainty in technological search," *Management Science*, 47, 117–132.

Fontana, M., M. Iori, F. Montobbio, and R. Sinatra (2020): "New and atypical combinations: An assessment of novelty and interdisciplinarity," *Research Policy*, 49.

Garud, R., H. A. Schildt, and T. K. Lant (2014): "Entrepreneurial storytelling, future expectations, and the paradox of legitimacy," *Organization Science*, 25, 1479–1492.

Gifford, S. (1992): "Allocation of entrepreneurial attention," *Journal of Economic Behavior & Organization*, 19, 265–284.

Gompers, P. A., W. Gornall, S. N. Kaplan, and I. A. Strebulaev (2020): "How do venture capitalists make decisions?" *Journal of Financial Economics*, 135, 169–190.

Hainmueller, J. (2011): "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies," *Political Analysis*, 16, 25–46.

Hallen, B. L., S. L. Cohen, and C. B. Bingham (2020): "Do accelerators work? If so, how?" *Organization Science*, 31, 378–414.

Hannan, M. T., G. Le Mens, G. Hsu, B. Kovács, G. Negro, L. Pólos, E. G. Pontikes, and A. J. Sharkey (2019): *Concepts and Categories*, Columbia University Press.

Harvey, S. and J. S. Mueller (2021): "Staying alive: Toward a diverging consensus model of overcoming a bias against novelty in groups," *Organization Science*, 32, 293–314.

Henderson, R. M. and K. B. Clark (1990): "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms," *Administrative Science Quarterly*, 35, 9–30.

Hochberg, Y. V. (2016): "Accelerating entrepreneurs and ecosystems: The seed accelerator model," *Innovation Policy and the Economy*, 16, 25–51.

Hoenig, D. and J. Henkel (2015): "Quality signals? The role of patents, alliances, and team experience in venture capital financing," *Research Policy*, 44, 1049–1064.

Huang, L. and J. L. Pearce (2015): "Managing the Unknowable: The Effectiveness of Early-stage Investor Gut Feel in Entrepreneurial Investment Decisions," *Administrative Science Quarterly*, 60, 634–670.

Kirsch, D., B. Goldfarb, and A. Gera (2009): "Form or substance: The role of business plans in venture capital decision making," *Strategic Management Journal*, 30, 487–515.

Kovács, B. and M. T. Hannan (2015): "Conceptual Spaces and the Consequences of Category Spanning," *Sociological Science*, 2, 252–286.

Lane, J. N., M. Teplitskiy, G. Gray, H. Ranu, M. Menietti, E. C. Guinan, and K. R. Lakhani (2022): "Conservatism Gets Funded? A Field Experiment on the Role of Negative Information in Novel Project Evaluation," *Management Science*, 68, 4478–4495.

Lerner, J. (2002): "When Bureaucrats Meet Entrepreneurs: The Design of Effective "Public Venture Capital" Programmes," *The Economic Journal*, 112, 73–84.

——— (2009): *Boulevard of Broken Dreams: Why Public Efforts to Boost Entrepreneurship and Venture Capital Have Failed - and What to Do about It*, Princeton University Press.

Litchfield, R. C., L. L. Gilson, and P. W. Gilson (2015): "Defining Creative Ideas: Toward a More Nuanced Approach," *Group and Organization Management*, 40, 238–265.

Lounsbury, M. and M. A. Glynn (2001): "Cultural entrepreneurship: Stories, legitimacy, and the acquisition of resources," *Strategic Management Journal*, 22, 545–564.

Markowitz, D. M. (2019): "What Words Are Worth: National Science Foundation Grant Abstracts Indicate Award Funding," *Journal of Language and Social Psychology*, 38, 264–282.

Martens, M. L., J. E. Jennings, and P. Devereaux Jennings (2007): "Do the Stories They Tell Get Them the Money They Need? The Role of Entrepreneurial Narratives in Resource Acquisition," *Academy of Management Journal*, 50, 1107–1132.

Millar, N., B. Batalo, and B. Budgell (2022): "Trends in the Use of Promotional Language (Hype) in Abstracts of Successful National Institutes of Health Grant Applications, 1985-2020," *JAMA Network Open*, 5, e2228676.

Moss, T. W., M. Renko, E. Block, and M. Meyskens (2018): "Funding the story of hybrid ventures: Crowdfunder lending preferences and linguistic hybridity," *Journal of Business Venturing*, 33, 643–659.

Mount, M. P., M. Baer, and M. J. Lupoli (2021): "Quantum leaps or baby steps? Expertise distance, construal level, and the propensity to invest in novel technological ideas," *Strategic Management Journal*, 42, 1490–1515.

Mueller, J., S. Melwani, J. Loewenstein, and J. Deal (2018): "Reframing the decision-makers' dilemma: Towards a social context model of creative idea recognition," *Academy of Management Journal*, 61, 94–110.

Mueller, J. S., C. J. Wakslak, and V. Krishnan (2014): "Construing creativity: The how and why of recognizing creative ideas," *Journal of Experimental Social Psychology*, 51, 81–87.

Myers, K. R., L. Lanahan, and E. Johnson (2025): "Small Business Innovation Applied to National Needs," Tech. rep., NBER Working Papers.

Navis, C. and M. A. Glynn (2011): "Legitimate distinctiveness and the entrepreneurial identity: Influence on investor judgments of new venture plausibility," *Academy of Management Review*, 36.

Packard, G. and J. Berger (2021): "How Concrete Language Shapes Customer Satisfaction," *Journal of Consumer Research*, 47, 787–806.

Pan, L., X. Li, J. Chen, and T. Chen (2020): "Sounds novel or familiar? Entrepreneurs' framing strategy in the venture capital market," *Journal of Business Venturing*, 35.

Pan, L., G. McNamara, J. J. Lee, J. Haleblian, and C. E. Devers (2018): "Give it to us straight (most of the time): Top managers' use of concrete language and its effect on investor reactions," *Strategic Management Journal*, 39, 2204–2225.

Peng, H., H. S. Qiu, H. B. Fosse, and B. Uzzi (2024): "Promotional language and the adoption of innovative ideas in science," *Proceedings of the National Academy of Sciences of the United States of America*, 121, e2320066121.

Petty, J. S., M. Gruber, and D. Harhoff (2023): "Maneuvering the odds: The dynamics of venture capital decision-making," *Strategic Entrepreneurship Journal*, 17, 239–265.

Piezunka, H. and L. Dahlander (2015): "Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing," *Academy of Management Journal*, 58, 856–880.

Quy, C. Z. and Q. N. Huy (2007): "How Entrepreneurs Use Symbolic Management to Acquire Resources," *Administrative Science Quarterly*, 52, 70–105.

Rindova, V. P. and A. P. Petkova (2007): "When is a new thing a good thing? Technological change, product form design, and perceptions of value for product innovations," *Organization Science*, 18, 217–232.

Rosenkopf, L. and P. McGrath (2011): "Advancing the conceptualization and operationalization of novelty in organizational research," *Organization Science*, 22, 1297–1311.

Scott, E. L., P. Shu, and R. M. Lubynsky (2020): "Entrepreneurial uncertainty and expert evaluation: An empirical analysis," *Management Science*, 66, 1278–1299.

Simon, H. A. (1957): *Models of man: Social and rational*, Wiley.

Snihur, Y., L. D. W. Thomas, R. Garud, and N. Phillips (2022): "Entrepreneurial Framing: A Literature Review and Future Research Directions," *Entrepreneurship Theory and Practice*, 46, 578–606.

Taeuscher, K. and H. Rothe (2024): "Entrepreneurial framing: How category dynamics shape the effectiveness of linguistic frames," *Strategic Management Journal*, 45, 362–395.

Trope, Y. and N. Liberman (2010): "Construal-Level Theory of Psychological Distance," *Psychological Review*, 117, 440–463.

Trope, Y., N. Liberman, and C. J. Wakslak (2007): "Construal Levels and Psychological Distance: Effects on Representation, Prediction, Evaluation, and Behavior," *Journal of Consumer Psychology*, 17, 83–95.

Tushman, M. L. and P. Anderson (1986): "Technological Discontinuities and Organizational Environments," *Administrative Science Quarterly*, 31, 439–465.

Uzzi, B., S. Mukherjee, M. Stringer, and B. Jones (2013): "Atypical combinations and scientific impact," *Science*, 342, 468–472.

Van Knippenberg, D., L. Dahlander, M. R. Haas, and G. George (2015): "From the editors: Information, attention, and decision making," *Academy of Management Journal*, 58, 649–657.

van Werven, R., O. Bouwmeester, and J. P. Cornelissen (2015): "The power of arguments: How entrepreneurs convince stakeholders of the legitimate distinctiveness of their ventures," *Journal of Business Venturing*, 30, 616–631.

General Conclusions

Motivated by the growing body of evidence suggesting that breakthrough ideas are becoming increasingly difficult to achieve (Bloom et al., 2020; Park et al., 2023), this dissertation examines the emergence of novel ideas in scientific research, and the support for novel ideas in technology-based entrepreneurship. More specifically, in Chapter 2, we examine the link between researchers' deviation from grant proposals and scientific novelty. In Chapter 3, we explore whether and how interdisciplinary advisors may influence PhD students' research novelty and productivity. Finally, in Chapter 4, we study how contextual novelty and congestion influence the evaluation of entrepreneurial ideas.

Taken together, these three chapters contribute to our understanding of how micro-level processes, operating within organizations and shaped by policies, both enable and constrain the emergence and support of novelty. With respect to the generation of novel ideas, the dissertation highlights the tension inherent in knowledge recombination as a 'double-edged sword' (Kaplan and Vakili, 2015): it creates opportunities for generating valuable, innovative ideas while simultaneously introducing coordination challenges for those who pursue them. Drawing on the context of grant-funded research, we demonstrate that the research process frequently involves adapting and reshaping ideas initially proposed in grant applications, highlighting the challenges of integrating diverse expertise while mitigating coordination frictions. Moreover, we document how interdisciplinary engagement, though potentially enriching, can also produce adverse effects at the individual level, particularly for PhD students who are still developing their skills during academic train-

ing. Finally, since the generation of ideas does not guarantee their uptake, we also examine the contextual factors that influence the mobilization of resources for novel ideas. Here, we build on existing notions of the subjective recognition of ideas and argue that their appreciation hinges not only on their intrinsic value but also on how they are perceived by evaluators. We further show that specific contextual factors—such as congestion in the assessment process—shape how linguistic framing influences the support for contextually novel ideas, that is, ideas that are distinct from their peers.

Because each chapter concludes with its own discussion of contributions and limitations, the remainder of this section briefly revisits their key takeaways and implications, highlights overarching limitations, and outlines illustrative avenues for future research.

Chapter 2 showed that funded projects often deviate from their original research plans, as measured by the grant proposal, and that new team members play a central role in catalyzing that process. Using language processing techniques to measure the semantic distance between grant proposals and publications that acknowledge the grant, we documented a positive association between deviation and the presence of members who join after the award. The association is stronger when newcomers bring outside affiliations or prior work that is distant from the proposal. First-time collaborators, however, are associated with less deviation, which is consistent with coordination hurdles that dampen the pursuit of new directions. In a second step, we related the deviation to scientific impact and novelty with a control function approach that exploits linguistic features of proposals. Deviation is positively associated with citations and with the introduction of new concepts, which is consistent with a recombination mechanism. These findings suggest that strict monitoring of adherence to initial proposals may not be beneficial under certain conditions, such as when research direction follows curiosity-driven changes. For research teams, adding members from distant domains or outside affiliations likely works best when a deliberate integration process, clear task organization, and consistent communication routines are in place, enabling gains from recombination while keeping coordination

costs in check. Future research could further explore how opportunity-driven versus necessity-driven changes differentially shape team dynamics and research direction. In particular, it would be valuable to examine whether strategic adjustments initiated by principal investigators, such as recruiting new members with unique skills, complementary expertise, or valuable networks, reflect efforts to expand research scope based on new insights or methods, or adaptive responses to external pressures, such as competing research teams 'scooping' their ideas. Additionally, examining the potential synergies and tensions arising from multiple sources of funding at the laboratory level—and how these interactions shape research agendas—represents a promising avenue for extending this line of inquiry.

Chapter 3 focuses on academic training and interdisciplinary research. We study the establishment of interdisciplinary research centers at a large European university, which allow us to hold the broader research environment constant and focus on mentorship as the channel through which the opportunities and frictions of interdisciplinarity reach students. Combining records on PhD students and advisors with administrative and publication data, and leveraging variation in advisors' center engagement, we examine the relationship between PhD students' productivity and novelty, with advisors' engagement in interdisciplinary research. Our findings suggest that while can foster students' collaboration with distant researchers and may help their visibility, but it does not raise students' research novelty or interdisciplinarity, and it is associated with lower research productivity. Our study emphasizes that mentoring can be critical to navigating interdisciplinary work, as the insights we uncover rely on within-advisor effects. Importantly, our findings suggest that mere access to distant collaborators is not sufficient to foster interdisciplinarity. By documenting the trade-offs that interdisciplinary work entails for doctoral students, our findings offer insights for universities and researchers participating in such initiatives. In particular, universities and researchers need to consider how these programs alter the research environment and mitigate potentially negative consequences for doctoral training at the individual level. Future research should investigate the longer-term

effects of exposure to interdisciplinarity for doctoral students, recognizing that interdisciplinary work—and particularly the development of impactful, novel research ideas—often takes time to mature. In particular, examining how such exposure influences students' subsequent career trajectories, including their placement in academia or industry, would provide a more complete understanding of the influence of interdisciplinary research on academic training.

Chapter 4 turned to entrepreneurship and the evaluation of novel ideas in a large-scale European accelerator program. We introduced contextual novelty and congestion as key moderators of framing effectiveness, which has been previously unexplored. We treat framing as a graded construct in which feasibility-oriented and innovativeness-oriented elements can both matter. Two contextual features were central. Contextual novelty captured how atypical a proposal was relative to competing ideas. Congestion captured how far submission volumes exceeded expected levels, with implications for evaluator cognitive bandwidth. We find that while evaluators reward feasibility framing, the benefit of feasibility framing does not extend to contextually novel ideas on average. Under high congestion, however, evaluators are less likely to reject contextually novel ideas, and feasibility framing further helps contextually novel ideas to avoid rejection. Our results suggest that when support program mandates emphasize 'breakthrough market potential,' cognitive load can activate heuristics that favor novelty rather than penalize it. We offer practical implications for entrepreneurs and accelerators. For entrepreneurs, tailor framing to the evaluation context: both feasibility and innovativeness matter, but under heavy evaluator workload, and especially in crowded rounds with unconventional proposals, signals of feasibility can be decisive. For start-up funders and other support mechanisms, selection outcomes depend on context and it is therefore advisable to monitor workload and protect evaluators' cognitive bandwidth. When congestion is unavoidable, state priorities clearly (e.g., breakthrough potential) and prompt applicants to document feasibility so evaluators can identify and support high-novelty ventures. Future research should further examine how

evaluators process and respond to instances of heightened congestion, and what institutional mechanisms might better accommodate such dynamics so that novel ideas are not systematically selected against. Moreover, refining the metrics used to capture both framing and (contextual) novelty in entrepreneurial ideas could strengthen confidence in this line of inquiry, providing funders and policymakers with reliable tools to design and implement effective mitigating measures.

While these three contextually grounded studies, based on originally assembled data combining proprietary and public sources, allow for underexplored and context-rich analyses, two limitations recur across the three chapters. First, external validity is constrained by the specific settings we examine. Chapters 2 and 3 rely on a single research university, which enables precise tracking of projects, teams, and outputs and clear identification of advisor–student relationships and the research environment shaped by interdisciplinary center membership. Although many research universities share features with our setting, the extent to which our results reflect the idiosyncrasies of this institution remains an empirical question. Similarly, Chapter 4 centers on a large, publicly supported European accelerator with a mandate to fund ideas with "breakthrough potential." We recognize that the effectiveness of specific framings can vary in other settings, where mandates may center on e.g., short-term profitability or job creation, and therefore emphasize the importance of context in the evaluation of entrepreneurial ideas. Taking a broader perspective, the processes we study remain underexplored in innovation systems outside Europe and other high-income countries, where funding systems and institutions likely differ in consequential ways. New, valuable research and insights may emerge from a deep dive into these underexplored contexts.

Second, our empirical designs are primarily observational and correlational, and we therefore refrain from claiming causality. This choice is well-suited to mapping patterns and opening new lines of inquiry around compelling phenomena and questions. Still, to place the relationships we document on a firmer footing, future work could exploit quasi-experimental

variation or field interventions. Institutional thresholds, staggered rollouts, eligibility rules, or workload assignments can generate plausibly exogenous shocks while keeping organizations and real-world contexts at the center, thereby extending the validity and scope of our findings. A related issue is to disentangle selection from treatment. For example, linking ventures' application content and evaluation context to post-award outcomes (such as survival or sales) would show whether the framings that help under congestion also predict subsequent performance.

Finally, aligned with the dissertation's spirit of understanding the pathways to novelty across science and entrepreneurship, I would like to briefly conclude by reflecting on the possibilities and challenges that recent breakthroughs in the development of generative artificial intelligence (AI) pose. Although GPT-3 was not released to the public until 2022, and the research leading to this dissertation was already underway, GPT-like and subsequent variants, competitors, and enhancements have made clear its transformative potential for scientific research and innovation. The possibility to integrate AI both as a cutting-edge research tool and also as a new method of invention has already attracted scholars' attention (Bianchini et al., 2022; Tranchero et al., 2024; Cockburn et al., 2018).

From a methodological standpoint, generative language models offer unprecedented opportunities to enhance the precision of our theoretical constructs, enabling more nuanced analyses of text to capture meanings and semantic relationships more accurately. Indeed, we make use of some of these tools in Chapter 2 where we measure *deviation* as the semantic similarity between proposals and publications, and in Chapter 4 where we also calculate the atypicality of competing venture proposals, both our measures relying on BERT-type models. Validating such AI-based measures of novelty remains a critical direction for future research, for instance, through expert assessments or crowd-based evaluations that benchmark automated measures against human judgments. Still, we are far from exploring all the opportunities that this tool can offer. For example, how can we better measure the knowledge recombination processes that we aim to uncover? Can

we better capture whether and how PhD students' research becomes more interdisciplinary? Looking ahead, can we learn more about these processes by integrating AI models directly into e.g., research equipment to understand interactions among scientists and the generation of creative ideas?

More profoundly, however, AI may reshape the very processes this dissertation has studied. Each of the phenomena we examined — the role of grant proposals in delineating research direction, the use of language to articulate and attract support for entrepreneurial ideas, and the mechanisms that enable productive cross-field interdisciplinary collaborations — can be fundamentally altered by the emerging capabilities of AI. These developments present important challenges for future research: to develop a deeper understanding of what novelty is, while simultaneously investigating whether and how these technological capabilities alter (some of) the fundamental dynamics of innovation processes. I hope this dissertation inspires future work on additional pathways to novelty in science and entrepreneurship, including (but not limited to) those enabled by artificial intelligence. Advancing this agenda will likely require refining existing theories and methods. I look forward to building on this work.

References

Bianchini, S., M. Müller, and P. Pelletier (2022): "Artificial intelligence in science: An emerging general method of invention," *Research Policy*, 51, 104604.

Bloom, N., C. I. Jones, J. V. Reenen, and M. Webb (2020): "Are Ideas Getting Harder to Find? †," *American Economic Review*, 110, 1104–1144.

Cockburn, I. M., R. Henderson, and S. Stern (2018): "The Impact of Artificial Intelligence on Innovation," *NBER Working Papers*, 24449.

Kaplan, S. and K. Vakili (2015): "The double-edged sword of recombination in breakthrough innovation," *Strategic Management Journal*, 36, 1435–1457.

Park, M., E. Leahey, and R. J. Funk (2023): "Papers and patents are becoming less disruptive over time," *Nature*, 613, 138–144.

Tranchero, M., C.-F. Brenninkmeijer, A. Murugan, and A. Nagaraj (2024): "Theorizing with Large Language Models," *NBER Working Papers*.

Appendices

A Appendix Chapter 2

A.1 Supplementary Data Description

The Dimensions database is a comprehensive bibliometric record of more than 140 million publications and 7 million grants with disambiguated data on individual researchers as well as their affiliations (Herzog et al., 2020). Information on grants is provided directly by partner funding bodies and complemented by public sources, covering more than 600 funders. Dimensions further complements these data by processing grant codes mentioned in the acknowledgement section of publications, and then matched to the Dimensions grant database (Grassano et al., 2017).

Using researchers' full names and affiliations, we curated a list of researchers to match personnel data to researcher IDs in the Dimensions database. For 95% of the matched researchers we find a unique ID (6,550 individuals). For the remaining 5% we perform manual searches and retain multiple IDs to ensure data completeness. We collect a publication corpus of 202,354 publications linked to these researchers, where 36,994 publications (18%) cite a supporting grant. For each grant we obtain information such as the grant title, abstract (i.e., a summary of the grant proposal), funder details, grant start date and end date, researchers that were awarded the grant, and their affiliation(s) from the Dimensions database. From Dimensions we also capture all grants awarded to our curated list of researchers and their co-awardees.

We match these records with data from the FRIS database based on grant title, funder details, and researcher names. Of the grant-supported publications, 9,183 publications acknowledge grants awarded to (at least one) KUL researcher(s) in the years 2000–2018. The remaining publication-grant pairs are out of scope since they have been awarded in a date before our observation period (2000–2018) or are linked to a group of scientists that is entirely external to KU Leuven. We are able to collect project descriptions (abstracts) for 1,107 grants linked to 6,196 publications, representing 67% of the grant-supported publications in our study period. Below, we provide further details about the criteria used to construct the estimation sample.

Publications supported by a single grant. From our corpus of publication-grant pairs, approximately 46% of the grants are linked to publications that mention multiple sources of funding. After dropping these publications linked to multiple grants, we retain 2,318 publications and 600 grants. Given the substantial number of exclusions, we address concerns about potential sample selection bias. Descriptive analyses in Table A.1.2 below indicate that publications supported by a single grant tend to show a lower degree of *de-*

viation and receive fewer citations on average, suggesting that our analytic sample takes a conservative approach, consistent with our arguments. We repeat our analyses including publications supported by multiple grants and report it together with other supplementary analyses of the antecedents of deviation (Appendix C).

Publication date one year after grant award. In total, 230 observations—equivalent to 10% of the remaining publications—fail to meet the criterion of publication date at least 1 year after the grant start date. In setting this threshold, we consider two factors: first, the time required to conduct research (e.g., perform experiments, collect data, conduct analyses, write up results), which varies across fields and projects and can easily take months; and second, the time it takes for journals to make a publication publicly available (e.g., initial decision, peer review, revisions, final decision). Recent estimates indicate that the total review time, taking into consideration multiple evaluation rounds but not the authors' time spent in revising and resubmitting, ranges from an average of 12–14 weeks in Biomedical and Natural sciences, to an average of 25 weeks in Economics and Business (Huisman and Smits, 2017). Second, previous studies suggest that to maximize the likelihood of grant award, grant applications are often based on advanced but unpublished research showing promising results (Li, 2017).

Outliers. To limit the influence of outliers in our results, we also exclude 26 grants with more than 75 resulting publications, 21 publications resulting more than ten years after the grant start, and 22 articles with more than 200 references (analogous to trimming the top 1% in each case). We show summary statistics of the researchers (Table A.1.1) and grants (Table A.1.2) included in our estimation sample.

Table A.1.1: Summary statistics, researchers (N=4,861)

	New team members N = 3,885		Grant holders N = 976		Diff. of means > 0
Variables	Mean	(SD)	Mean	(SD)	t-stat
Academic age	11.30	(10.59)	18.10	(9.76)	-18.22***
Sum of publications	7.36	(17.81)	12.72	(21.70)	-8.03***
Fractionalized sum of publications	0.47	(1.16)	0.94	(1.48)	-10.49***
Avg. coauthors per publication	2.75	(3.95)	3.15	(3.75)	-2.85***
Unique coauthors	8.89	(26.95)	17.52	(37.25)	-8.23***
Sum of grants	0.49	(1.81)	1.31	(3.17)	-10.63***
Avg. yearly citations per publication	7.04	(23.53)	7.60	(22.85)	-0.66
Frac. avg. yearly citations per publication	0.099	(0.38)	0.094	(0.26)	0.33

Notes: Except for academic age, all variables are calculated in the 3 years prior to the grant start. Fractionalized implies weighting by the number of coauthors. Significance level: *** p<0.01

Table A.1.2: Summary statistics, publications supported by single-grant vs multiple-grants

Variables	Publications supported by a single grant N = 1,341		Publications supported by multiple grants N = 2,618		Diff. of means >0
	Mean	(SD)	Mean	(SD)	t-stat
Citations count	41.46	(54.37)	58.01	(117.4)	-4.89***
Field citation ratio	7.86	(10.44)	10.08	(21.16)	-3.56***
Deviation (SciBERT)	0.287	(0.07)	0.292	(0.07)	-2.07**
Deviation (SciSpacy)	0.17	(0.09)	0.19	(0.09)	-6.45***

Notes: Significance levels: *** p<0.01; ** p<.05

Table A.1.3: Pairwise correlations

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)
(1) Deviation	1																											
(2) Citations	-0.05	1																										
(3) Field Citation Ratio	-0.07	0.75*	1																									
(4) Novelty (0/1)	0.04	0.02	-0.04	1																								
(5) Top 10 (0/1)	-0.02	0.59*	0.59*	-0.01	1																							
(6) Bottom 10 (0/1)	0.02	-0.27*	-0.24*	-0.04	-0.07*	1																						
(7) Authors count	0.08*	0.06	-0.07*	0.10*	-0.08*	-0.09*	1																					
(8) Grant holders (GH)	0.01	-0.01	0.01	-0.03	0.02	-0.02	0.12*	1																				
(9) New team members (NTM)	0.09*	0.06	-0.07	0.11*	-0.09*	-0.09*	0.95*	-0.15*	1																			
(10) Share of new team members	0.13*	-0.05	-0.04	0.05	-0.06	-0.02	0.22*	-0.02	0.22*	1																		
(11) Internal (0/1)	-0.10*	0.03	-0.01	-0.04	-0.03	-0.03	-0.22*	-0.20*	-0.16*	-0.09*	1																	
(12) External (0/1)	0.11*	-0.05	-0.01	-0.02	0.00	0.04	-0.03	0.24*	-0.09*	0.15*	-0.51*	1																
(13) Mixed (0/1)	0.01	0.02	0.00	0.08*	0.00	-0.04	0.36*	-0.12*	0.39*	0.11*	-0.4*	-0.52*	1															
(14) Distant new team members (0/1)	0.25*	-0.01	-0.10*	0.04	-0.05	-0.02	0.16*	0.05	0.14*	0.05	-0.08*	0.04	0.06	1														
(15) First collaboration (0/1)	-0.07	0.03	-0.03	0.04	-0.05	-0.04	0.38*	0.14*	0.34*	0.07*	-0.13*	0.04	0.19*	-0.05	1													
(16) GH publication rate (log)	-0.04	-0.06	0.02	-0.07	-0.02	-0.05	-0.02	0.19*	-0.07*	-0.06	0.01	0.01	-0.04	-0.18*	0.17*	1												
(17) GH citation rate (log)	-0.09*	0.08*	0.03	0.04	0.01	-0.04	-0.01	0.12*	-0.05	0.04	-0.06	0.08*	-0.03	-0.08*	0.09*	-0.13*	1											
(18) GH multiple funding (0/1)	-0.10*	0.06	0.03	-0.01	0.03	-0.06	0.07*	0.09*	0.05	-0.02	-0.02	0.03	-0.02	-0.06	0.15*	0.25*	0.08*	1										
(19) GH absence (0/1)	0.13*	0.01	-0.04	0.03	0.02	0.04	0.00	-0.46*	0.13*	0.08*	0.12*	-0.14*	0.07	0.20*	-0.39*	-0.52*	-0.19*	-0.27*	1									
(20) NTM publication rate (log)	0.01	0.01	-0.03	0.01	-0.03	-0.05	0.03	0.08*	0.01	0.21*	0.00	0.00	0.05	-0.03	0.02	0.11*	0.19*	0.07	-0.07*	1								
(21) NTM citation rate (log)	0.06	-0.02	-0.02	0.03	-0.06	-0.05	0.14*	-0.07*	0.17*	0.36*	0.12*	-0.14*	0.12*	0.00	-0.04	0.01	-0.02	0.04	0.10*	0.15*	1							
(22) NTM multiple funding (0/1)	0.02	0.03	-0.01	0.08*	-0.03	-0.06	0.26*	-0.13*	0.29*	0.16*	0.03	-0.11*	0.15*	0.06	0.04	-0.02	-0.08*	0.44*	0.19*	0.10*	0.37*	1						
(23) Embeddedness in prior research	-0.21*	0.22*	0.21*	-0.05	0.17*	-0.17*	-0.12*	-0.05	-0.10*	0.05	0.04	-0.02	0.02	-0.08*	0.01	0.01	0.09*	0.06	-0.04	0.04	-0.04	0.00	1					
(24) Time to publication (field-norm.)	-0.01	-0.06	-0.02	0.01	0.00	0.04	-0.03	-0.02	-0.02	0.09*	-0.03	-0.02	0.05	0.06	-0.06	-0.14*	-0.04	-0.39*	0.12*	-0.08*	-0.12*	-0.27*	0.05	1				
(25) Year	-0.15*	-0.29*	-0.10*	-0.11*	-0.10*	0.05	-0.03	0.07	-0.06	0.03	-0.01	0.00	0.01	-0.21*	0.04	0.26*	0.00	0.06	-0.14*	0.06	0.06	0.02	0.07*	0.05	1			
(26) Grant productivity (field-norm.)	-0.02	0.03	-0.01	0.02	0.03	0.01	0.09*	-0.18*	0.14*	0.02	0.06	-0.08*	0.05	0.17*	0.00	0.17*	-0.01	0.11*	0.11*	0.06	-0.01	0.17*	0.06	0.05	0.02	1		
(27) Curiosity	0.05	0.00	0.16*	0.01	0.11*	0.01	-0.14*	-0.05	-0.12*	0.04	0.05	-0.04	0.00	-0.02	-0.07*	0.00	0.03	0.07*	-0.02	-0.09*	-0.05	-0.01	0.00	-0.02	0.06	-0.02	1	
(28) Grant abstract word count	0.18*	-0.01	-0.07	0.04	-0.02	-0.03	0.07	-0.09*	0.09*	0.02	-0.02	-0.02	0.06	0.49*	-0.09*	-0.20*	-0.09*	-0.11*	0.29*	-0.04	-0.01	0.07*	-0.03	0.09*	-0.11*	0.31*	-0.03	1

Notes: N=1,341; Significance level: * p<0.01

A.2 Supplementary information on deviation from grant proposal and publications

Promotional language in grant proposals

To investigate whether the vocabulary used in grant writing may focus on securing funding rather than solely describing the research project, we measured the use of adjectives with a "promotional sense" in the grant abstracts (Millar et al., 2022; Peng et al., 2024). Specifically, we looked for words in the grant abstracts in our sample that match one of the 139 promotional adjectives collected by Millar et al. (2022) from successful grant proposals of the NIH during 1982–2020, and we find 125 of these adjectives used in our sample. Common exemplars of these adjectives with a promotional sense lacking a direct scientific meaning include "unprecedented", "groundbreaking", "renowned", "rigorous", etc.

Following Peng et al. (2024), we calculated the percentage of adjectives as the ratio of the number of adjectives to the total number of words in an abstract, and found that they constitute a small share of the grant abstracts in our dataset. Specifically, the average proportion of promotional adjectives is 1.4%, with a maximum observed value of 6.4% and a minimum of 0. Furthermore, our text preprocessing procedure removes stop words and frequently used terms, filtering out 25% of the identified adjectives. Many of the remaining adjectives may add scientific meaning to the project description (e.g., "transdisciplinary," "interdisciplinary," "systematic," etc.). We provide three examples of such cases in Table A.2.1. Consequently, we posit that the role of language used solely for promotional purposes in the grant abstracts, with little relevance to the research content, is minimal for the purpose of our analysis.

Figure A.2.1: Distribution of *deviation* metric

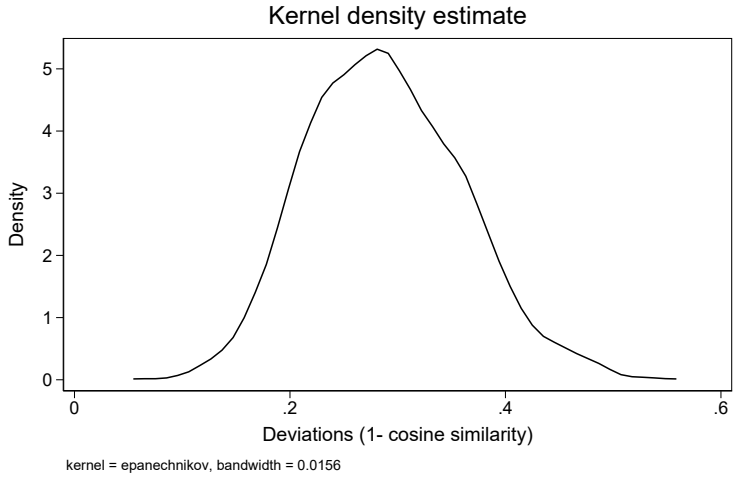


Table A.2.1: Examples of grant abstracts and promotional adjectives

Grant abstract	Promotional adj.
TransMID focuses on the development of **novel** methods to estimate **key** epidemiological parameters from both serological and social contact data, with the aim to significantly expand the range of public health questions that can be adequately addressed using such data. Using new statistical and mathematical theory and newly collected as well as readily available serological and social contact data (mainly from Europe), fundamental mathematical and epidemiological challenges as outlined in the following work packages [. . .]. TransMID is **transdisciplinary** in nature with applications on diseases of **major** public health interest, such as pertussis, cytomegalovirus and measles.	key; major; novel; transdisciplinary;
Extracted evolutionary and epidemiological information from pathogen genomes has grown into an **important** instrument across infectious disease research [. . .]. However, despite the revolution in genome sequencing technologies and advances in statistical methodology, **key** questions about pathogen emergence and establishment in human populations remain unresolved for **major** viral epidemics. When confronted with new viral outbreaks, such as the recent **devastating** Ebola virus epidemic, we also struggle to deploy these technologies in a **systematic** and concerted way despite a **critical** need to support public health interventions. In this project, we propose to unravel **crucial** steps in the emergence and establishment of **key** viral pathogens. We will scrutinise the reservoir dynamics of HCV by sequencing complete hepacivirus genomes from infected samples **emerging** from a large-scale screening of African rodents, and analyze the cross-species transmission history using **novel** evolutionary methods that accommodate spatial and temporal variability in selective pressures.	critical; crucial; important; key; major; emerging; novel; systematic; devastating
Based on an **interdisciplinary** approach that combines expertise in **international** penal law and theory and methodology from social psychology, this research aims to contribute to the evaluation of the **international** criminal justice case law and its legitimacy with regard to the conventional objectives of criminal law (special deterrence, rehabilitation and retribution). Semi-directed interviews with 20 individuals who where accused and judged (convicted or acquitted) by the International Criminal Tribunal for the former Yugoslavia (ICTY) will be carried out. In-depth qualitative analysis, using thematic and typological methods, of the interviewees' discourses concerning their perceptions of procedures and sanctions stemming from the ICTY, as well as of their criminal responsibility will be carried out. We will then evaluate the case law of international criminal jurisdictions by confronting it to the perceptions of convicted individuals. Demonstrating the feasibility of this research, a pilot interview with a person convicted by the ICTY has already been conducted [...].	interdisciplinary; international

Table A.2.2: Deviation by Field of Research (N=1,341)

Field of Research	Mean	(SD)	Min	Max
Agricultural, vet. & food sci.	0.26	(0.02)	0.24	0.28
Biological sciences	0.28	(0.07)	0.12	0.47
Biomedical & clinical sciences	0.29	(0.08)	0.07	0.54
Built environment & design	0.24	(0.05)	0.17	0.32
Chemical sciences	0.28	(0.06)	0.19	0.40
Commerce, management, tourism	0.27	(0.07)	0.17	0.42
Earth sciences	0.28	(0.06)	0.19	0.38
Economics	0.32	(0.08)	0.13	0.45
Education	0.31	(0.06)	0.21	0.40
Engineering	0.28	(0.05)	0.16	0.46
Environmental sciences	0.29	(0.06)	0.24	0.37
Health sciences	0.32	(0.09)	0.11	0.51
History, heritage & archaeology	0.27	(0.07)	0.18	0.38
Human society	0.26	(0.05)	0.15	0.41
Informatics & computer sciences	0.28	(0.07)	0.12	0.44
Language & communication	0.23	(0.07)	0.14	0.34
Law & legal studies	0.22	(0.00)	0.21	0.22
Mathematical sciences	0.32	(0.06)	0.21	0.47
Multiple fields	0.29	(0.07)	0.11	0.48
Philosophy & religion	0.26	(0.04)	0.19	0.29
Physical sci.	0.28	(0.09)	0.12	0.48
Psychology	0.30	(0.07)	0.18	0.51

Table A.2.3: Examples of deviation between grant proposal and publication

Example 1. ***Deviation*** **= 0.07**	
Grant title: The influence of formulation and process parameters on the formation of solid dispersion nano and microparticles via one-step co-axial electrospray	*Publication title*: Electrospraying of polymer solutions: Study of formulation and process parameters
PI: G. Van den Mooter	*Authors*: A. Smeets ; C. Clasen; G. Van den Mooter
Abstract: The number of new **drugs with low aqueous solubility** and hence low therapeutic efficiency is very high. One formulation strategy to increase the **solubility** and the uptake in the body is the use of **amorphous solid dispersions** (this is a solid product that is made up of a **non-crystalline polymer** in which the drug is chaotically distributed but that releases the drug upon contact with water). Often such an instantaneous release needs to be delayed to avoid fast precipitation of the poorly soluble drug in the **gastrointestinal tract**, in case of formulation of powders for reconstitution or when one wants to target the drug to specific sites in the gastrointestinal tract. Delayed release can be achieved by bringing the **dispersions** into the form of **coated tablets**, but those are difficult to administer to very young or elderly patients. [...]. We recently found that co-axial electrospraying has the potential to prepare coated solid dispersion submicron particles in one single step. However, the observed variability in **coating efficiency points to the lack of fundamental understanding of the process**. Hence the general objective of this project is to identify and quantify the **process and formulation parameters as well as their interplay** in order to design future processes based on scientific rationale instead of trial-and-error.	*Abstract*: Over the past decade, **electrospraying** has proven to be a promising method for the preparation of **amorphous solid dispersions**, an established formulation strategy to improve the **oral bioavailability** of poorly **soluble drug compounds**. Due to the lack of fundamental knowledge concerning adequate single nozzle electrospraying conditions, a trial-and-error approach is currently the only option. The objective of this paper is to study/investigate the influence of the **different formulation and process parameters**, **as well as their interplay**, on the formation of **a stable cone-jet mode** as a prerequisite for a reproducible production of **monodisperse micro- and nanoparticles**. To this purpose, different **polymers** commonly used in the formulation of **solid dispersions** were electrosprayed to map out the workable parameter ranges of the process. The experiments evaluate the importance of the **experimental parameters as flow rate, electric potential difference** and the distance between the tip of the nozzle and collector. Based on this, the type of solvent and the concentration of the polymer solutions, along with their viscosity and conductivity, were identified as determinative **formulation parameters**. This information is of utmost importance to rationally design further electrospraying methods for the preparation of amorphous solid dispersions.
Example 2. ***Deviation*** **= 0.35**	
Grant title: Nanoparticles in suspension: study of thermophysical behavior by means of photothermal and related techniques	*Publication title*: Adiabatic scanning calorimetry study of ionic liquid crystals with highly ordered crystal smectic phases
PI: W. Thielemans	*Authors*: S. Lombardo; S. Eyley; C. Schütz; H. van Gorp -; S. Rosenfeldt; G. Van den Mooter; W. Thielemans
Abstract: This project aims to **create one, two and three dimensional materials** through **self-assembly of rod-like cellulose and platelet -like starch** with a variety of surface functionalities. These surface functionalities will be introduced both as structure directing groups and to influence the chemistry and applications of the final materials. The project builds on the previous experience of the applicant in the field of surface modification of these nanocrystals with successful incorporation of a variety of charged functionalities, fluorescence and redox active species. By controlling the density and location of various functionalities to specific faces of the modified nanocrystals this project aims to control the **self-assembly of these modified nanocrystals** both with each other and with other species (such as metal nanoparticles). The fundamental **interactions** involved in the assembly of these species will be studied to develop better understanding of the processes involved. The innovative materials designed during this project along with the fundamental understanding developed from studies of the self-assembly process will spur advances in the field of **intelligent materials design**.	*Abstract*: The **interaction of bovine serum albumin (BSA) with** sulfated, carboxylated, and pyridinium-grafted **cellulose nanocrystals (CNCs)** was studied as a function of the degree of substitution by determining the adsorption isotherm and by directly measuring the thermodynamics of interaction. The adsorption of BSA onto positively charged pyridinium-grafted cellulose nanocrystals followed Langmuirian adsorption with the maximum amount of adsorbed protein increasing linearly with increasing degree of substitution. The **binding mechanism between** the positively charged pyridinium-grafted **cellulose nanocrystals and BSA** was found to be endothermic and based on charge neutralization. [...] . Negatively charged cellulose nanocrystals bearing sulfate and/or carboxylic functionalities were found to not interact significantly with the BSA protein. To investigate in more detail the role of single amino acids in the adsorption of proteins onto cellulose nanocrystals, we also studied the **interaction of different types of amino acids with CNCs**, i.e., charged (lysine, aspartic acid), aromatic (tryptophan, tyrosine), and polar (serine) amino acids. We found that none of the single amino acids bound with CNCs irrespective of surface charge and that therefore the binding of proteins with CNCs appears to require larger amino acid sequences that induce a greater entropic contribution to stabilize binding. Single amino acids are thus not adsorbed onto cellulose nanocrystals.

Table A.2.4: Deviation, teams with and without new members

	0 new team members N=54		>0 new team members N=1,287		Diff. of means > 0
Variables	Mean	(SD)	Mean	(SD)	t-stat
Deviation (SciBERT)	0.26	(0.06)	0.28	(0.07)	-2.63***
Deviation (Scispacy)	0.11	(0.04)	0.17	(0.09)	-5.22***

Note: Significance levels: *** p<0.01; ** p<.05

Table A.2.5: Deviation, teams without new members by (co)author status

	Single author N=32		Multiple coauthors N=22		Diff. of means > 0
Variables	Mean	(SD)	Mean	(SD)	t-stat
Deviation (SciBERT)	0.25	(0.07)	0.27	(0.06)	-2.62***
Deviation (Scispacy)	0.10	(0.04)	0.12	(0.05)	-3.16***

Note: Significance levels: *** p<0.01; ** p<.05

A.3 Supplementary analysis, antecedents of deviation

Table A.3.1: The antecedents of deviation (OLS)

DV: *Deviation*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Authors count (ln)		0.008*	0.005	0.004	0.004	0.007	0.006	0.007
		[0.004]	[0.004]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Share of new team members			0.018***	0.016**	0.016**	0.016**	0.015**	0.016**
			[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
Affiliations of new team members (NTM) (base: Internal)								
Mixed (0/1)				0.002	0.003	0.003	-0.003	0.003
				[0.005]	[0.005]	[0.005]	[0.008]	[0.005]
External (0/1)				0.013**	0.013**	0.013**	0.018**	0.013**
				[0.006]	[0.006]	[0.006]	[0.008]	[0.006]
Distant new team members (0/1)					0.017**	0.016**	0.016**	0.020**
					[0.008]	[0.008]	[0.008]	[0.010]
First collaboration (0/1)						-0.008*	-0.007	-0.007
						[0.004]	[0.007]	[0.005]
First collab. x Affiliations of NTM (baseline: Internal)								
First collab. x Mixed								0.009
								[0.010]
First collab. x External								-0.008
								[0.009]
First collab. x Distant new team members								-0.007
								[0.010]
GH citation rate (ln; 3yrs pre-grant)	-0.013	-0.012	-0.013	-0.014	-0.011	-0.011	-0.012	-0.011
	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.015]	[0.015]	[0.015]
GH publication rate (ln; 3yrs pre-grant)	0.017**	0.017**	0.017**	0.017**	0.018**	0.017**	0.017**	0.017**
	[0.008]	[0.008]	[0.008]	[0.007]	[0.008]	[0.008]	[0.008]	[0.008]
GH multiple funding (0/1)	-0.008	-0.009	-0.009	-0.009	-0.010*	-0.010*	-0.010*	-0.010
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
GH absence (0/1)	0.016	0.017	0.015	0.016	0.016	0.010	0.013	0.009
	[0.012]	[0.013]	[0.013]	[0.012]	[0.012]	[0.013]	[0.013]	[0.013]
Time to publication (field-norm.)	-0.003	-0.003	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Embeddedness in prior research	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Grant productivity (field-norm.)	-0.007	-0.008	-0.008	-0.008	-0.011	-0.011	-0.010	-0.011
	[0.009]	[0.009]	[0.009]	[0.009]	[0.010]	[0.010]	[0.010]	[0.010]
Funding source (*Baseline= FWO*)								
Belspo (0/1)	0.069***	0.067***	0.067***	0.066***	0.056***	0.056***	0.055***	0.055***
	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
ERC (0/1)	0.004	0.004	0.004	0.002	0.004	0.004	0.004	0.004
	[0.011]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Other funder within the EU (0/1)	0.035***	0.035***	0.033***	0.027***	0.023**	0.022**	0.024**	0.022**
	[0.008]	[0.008]	[0.008]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]
Other funder outside the EU (0/1)	0.017	0.017	0.016	0.010	0.004	0.004	0.005	0.004
	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.011]	[0.012]	[0.011]
Constant	0.282***	0.267***	0.263***	0.264***	0.263***	0.262***	0.263***	0.260***
	[0.034]	[0.035]	[0.034]	[0.035]	[0.034]	[0.033]	[0.034]	[0.033]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y
N	1341	1341	1341	1341	1341	1341	1341	1341
adj. R-sq	0.181	0.183	0.189	0.192	0.198	0.2	0.201	0.199

Notes: Deviation is measured as the inverse cosine similarity between the publication and grant abstract. GH denotes grant holders. Robust standard errors, clustered at the grant level, are shown in brackets. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A.3.2: The antecedents of deviation (Frac. Reg. Logit)

DV: *Deviation*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Authors count (ln)		1.040*	1.024	1.023	1.020	1.036	1.032	1.036
		[0.022]	[0.022]	[0.023]	[0.023]	[0.026]	[0.026]	[0.026]
Share of new team members			1.093***	1.082**	1.086**	1.085**	1.078**	1.085**
			[0.036]	[0.035]	[0.036]	[0.035]	[0.035]	[0.035]
Affiliations of new team members (NTM) (base: Internal)								
Mixed (0/1)				1.013	1.015	1.016	0.988	1.016
				[0.025]	[0.025]	[0.025]	[0.037]	[0.025]
External (0/1)				1.065**	1.065**	1.067**	1.089**	1.067**
				[0.032]	[0.031]	[0.031]	[0.041]	[0.031]
Distant new team members (0/1)					1.082**	1.080**	1.079**	1.098**
					[0.041]	[0.041]	[0.041]	[0.051]
First collaboration (0/1)						0.959*	0.966	0.965
						[0.021]	[0.034]	[0.024]
First collab. x Affiliations of NTM (baseline: Internal)								
First collab. x Mixed								1.044
								[0.050]
First collab. x External								0.963
								[0.044]
First collab. x Distant new team members								0.973
								[0.043]
GH citation rate (ln; 3yrs pre-grant)	0.939	0.942	0.936	0.934	0.946	0.946	0.943	0.946
	[0.073]	[0.074]	[0.074]	[0.075]	[0.073]	[0.072]	[0.072]	[0.073]
GH publication rate (ln; 3yrs pre-grant)	1.086**	1.089**	1.086**	1.085**	1.091**	1.089**	1.087**	1.089**
	[0.039]	[0.039]	[0.039]	[0.039]	[0.039]	[0.039]	[0.039]	[0.039]
GH other grants (0/1)	0.960	0.958	0.957	0.955	0.952*	0.952*	0.953*	0.953*
	[0.027]	[0.027]	[0.027]	[0.027]	[0.027]	[0.027]	[0.027]	[0.027]
GH absence (0/1)	1.082	1.089	1.074	1.079	1.080	1.048	1.063	1.044
	[0.063]	[0.065]	[0.063]	[0.063]	[0.063]	[0.064]	[0.066]	[0.064]
Time to publication (field-norm.)	0.984	0.984	0.981	0.981	0.980	0.980	0.979	0.981
	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]
Embeddedness in prior research	0.907***	0.911***	0.907***	0.908***	0.909***	0.910***	0.910***	0.910***
	[0.014]	[0.014]	[0.014]	[0.014]	[0.015]	[0.015]	[0.015]	[0.015]
Grant productivity (field-norm.)	0.965	0.961	0.963	0.961	0.946	0.946	0.950	0.947
	[0.043]	[0.043]	[0.042]	[0.042]	[0.043]	[0.043]	[0.043]	[0.043]
Funding source (*Baseline= FWO*)								
Belspo (0/1)	1.392***	1.382***	1.379***	1.371***	1.308***	1.307***	1.304***	1.305***
	[0.087]	[0.087]	[0.085]	[0.083]	[0.080]	[0.080]	[0.080]	[0.080]
ERC (0/1)	1.018	1.021	1.020	1.011	1.020	1.020	1.019	1.019
	[0.053]	[0.052]	[0.052]	[0.051]	[0.052]	[0.051]	[0.052]	[0.051]
Other funder within the EU (0/1)	1.190***	1.190***	1.178***	1.141***	1.120**	1.117**	1.125***	1.116**
	[0.048]	[0.049]	[0.048]	[0.048]	[0.050]	[0.049]	[0.050]	[0.049]
Other funder outside the EU (0/1)	1.090	1.087	1.083	1.051	1.024	1.022	1.027	1.021
	[0.062]	[0.062]	[0.062]	[0.059]	[0.058]	[0.057]	[0.057]	[0.056]
Constant	0.389***	0.360***	0.354***	0.354***	0.352***	0.350***	0.352***	0.347***
	[0.060]	[0.057]	[0.055]	[0.055]	[0.053]	[0.052]	[0.055]	[0.051]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y
N	1341	1341	1341	1341	1341	1341	1341	1341
adj. R-sq	-800.624	-800.570	-800.473	-800.398	-800.301	-800.264	-800.231	-800.259

Notes: Deviation is measured as the inverse cosine similarity between the publication and grant abstract. GH denotes grant holders. Notes: We report exponentiated coefficients from fractional regression logit models. Robust standard errors in brackets are clustered at the grant level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Table A.3.3: The antecedents of deviation (incl. publications supported by multiple grants)

DV: Deviation	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Authors count (ln)		0.009***	0.007**	0.006*	0.006*	0.007**	0.006*	0.007*
		[0.003]	[0.003]	[0.003]	[0.003]	[0.004]	[0.004]	[0.004]
Share of new team members			0.012**	0.011**	0.011**	0.011**	0.010**	0.011**
			[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Affiliations of new team members (NTM) (base: Internal)								
Mixed (0/1)				0.005	0.005	0.005	-0.001	0.005
				[0.004]	[0.004]	[0.004]	[0.006]	[0.004]
External (0/1)				0.004	0.004	0.004	0.013**	0.004
				[0.005]	[0.005]	[0.005]	[0.006]	[0.005]
Distant new team members (0/1)					0.011	0.011	0.010	0.016*
					[0.007]	[0.007]	[0.007]	[0.008]
First collaboration (0/1)						-0.003	0.002	-0.002
						[0.003]	[0.005]	[0.003]
First collab. × Affiliations of NTM (baseline: Internal)								
First collab. × Mixed								0.008
								[0.006]
First collab. × External								-0.014**
								[0.007]
First collab. × Distant new team members								-0.010
								[0.009]
GH citation rate (ln; 3yrs pre-grant)	-0.003	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
GH publication rate (ln; 3yrs pre-grant)	0.014**	0.015***	0.015***	0.015***	0.015***	0.015***	0.015***	0.015***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
GH multiple funding (0/1)	-0.003	-0.004	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
GH absence (0/1)	0.018*	0.020**	0.019**	0.020**	0.019*	0.017*	0.021**	0.016
	[0.009]	[0.009]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Time to publication (field-norm.)	-0.005***	-0.005***	-0.005***	-0.005***	-0.005***	-0.005***	-0.005***	-0.005***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Embeddedness in prior research	-0.019***	-0.018***	-0.018***	-0.018***	-0.018***	-0.018***	-0.018***	-0.018***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Grant productivity (field-norm.)	-0.003	-0.003	-0.003	-0.003	-0.004	-0.004	-0.004	-0.004
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
Funding source (Baseline= FWO)								
Belspo (0/1)	0.065***	0.064***	0.064***	0.063***	0.061***	0.061***	0.061***	0.061***
	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]
ERC (0/1)	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Other funder within the EU (0/1)	0.031***	0.032***	0.031***	0.030***	0.029***	0.028***	0.030***	0.028***
	[0.007]	[0.007]	[0.007]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Other funder outside the EU (0/1)	0.013*	0.013*	0.013*	0.012	0.011	0.010	0.012	0.010
	[0.007]	[0.007]	[0.007]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Nr of supporting grants per publication	0.004***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Constant	0.245***	0.229***	0.225***	0.224***	0.223***	0.223***	0.223***	0.222***
	[0.018]	[0.018]	[0.018]	[0.018]	[0.017]	[0.017]	[0.018]	[0.017]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y
N	3,959	3,959	3,959	3,959	3,959	3,959	3,959	3,959
adj. R-sq	0.161	0.165	0.167	0.167	0.168	0.168	0.171	0.169

Notes: In contrast to our main results, in this analysis we also include publications supported by multiple grants (840 grants in total). The average number of grants mentioned in a publications is 2.7 (S.D.=2.1). *Deviation* is measured as the inverse cosine similarity between the publication and grant abstract. GH denotes grant holders. Robust standard errors in brackets are clustered at the grant level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Table A.3.4: The antecedents of deviation (alt. share of new team members)

DV: *Deviation*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Authors count (ln)	0.007	0.007	0.007	0.008	0.007	0.008	0.006	0.005	0.006
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.006]	[0.006]	[0.006]
Share new team members (academic age ≥2)	0.013**	0.011*	0.013*						
	[0.007]	[0.007]	[0.007]						
Share new team members (academic age≥4)				0.013**	0.012*	0.013**			
				[0.006]	[0.006]	[0.006]			
Share new team members (alternative)							0.021**	0.020**	0.021**
							[0.010]	[0.010]	[0.010]
Affiliations of new team members (NTM) (base: Internal)									
Mixed (0/1)	0.003	-0.003	0.003	0.003	-0.004	0.003	0.002	-0.004	0.002
	[0.005]	[0.008]	[0.005]	[0.005]	[0.008]	[0.005]	[0.005]	[0.008]	[0.005]
External (0/1)	0.013**	0.018**	0.013**	0.013**	0.018**	0.013**	0.014**	0.019**	0.014**
	[0.006]	[0.008]	[0.006]	[0.006]	[0.008]	[0.006]	[0.006]	[0.008]	[0.006]
Distant new team members (0/1)	0.016*	0.016*	0.020**	0.016**	0.016**	0.020**	0.017**	0.016**	0.021**
	[0.008]	[0.008]	[0.010]	[0.008]	[0.008]	[0.010]	[0.008]	[0.008]	[0.010]
First collaboration (0/1)	-0.008*	-0.007	-0.007	-0.008*	-0.007	-0.007	-0.009**	-0.008	-0.008
	[0.004]	[0.007]	[0.005]	[0.004]	[0.007]	[0.005]	[0.004]	[0.007]	[0.005]
First collab. × Affiliations of NTM (baseline: Internal)									
First collab. × Mixed		0.010			0.010			0.009	
		[0.010]			[0.010]			[0.010]	
First collab. × External		-0.008			-0.008			-0.009	
		[0.009]			[0.010]			[0.010]	
First collab. × Distant new team members			-0.007			-0.007			-0.006
			[0.010]			[0.010]			[0.009]
GH citation rate (ln; 3yrs pre-grant)	-0.011	-0.011	-0.011	-0.012	-0.012	-0.012	-0.010	-0.011	-0.010
	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]	[0.016]	[0.016]	[0.016]
GH publication rate (ln; 3yrs pre-grant)	0.017**	0.017**	0.017**	0.017**	0.017**	0.017**	0.017**	0.017**	0.017**
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.007]	[0.007]	[0.007]
GH multiple funding sources (0/1)	-0.010*	-0.009	-0.009	-0.010*	-0.010	-0.010	-0.010*	-0.010*	-0.010*
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
GH absence (0/1)	0.010	0.013	0.010	0.010	0.013	0.009	0.005	0.008	0.004
	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
Time to publication (field-norm.)	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Embeddedness in prior research	-0.018***	-0.019***	-0.018***	-0.018***	-0.019***	-0.018***	-0.018***	-0.018***	-0.018***
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Grant productivity (field-norm.)	-0.011	-0.010	-0.011	-0.011	-0.010	-0.011	-0.012	-0.011	-0.012
	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Funding source (Baseline: FWO)									
Belspo (0/1)	0.056***	0.055***	0.056***	0.056***	0.055***	0.055***	0.056***	0.056***	0.056***
	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
ERC (0/1)	0.004	0.004	0.004	0.004	0.004	0.004	0.005	0.004	0.004
	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Other funder within the EU (0/1)	0.023**	0.024***	0.023**	0.022**	0.024**	0.022**	0.023**	0.025***	0.023**
	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]
Other funder outside the EU (0/1)	0.004	0.005	0.004	0.004	0.005	0.004	0.005	0.006	0.005
	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]
Constant	0.260***	0.262***	0.258***	0.263***	0.264***	0.261***	0.264***	0.265***	0.262***
	[0.032]	[0.034]	[0.033]	[0.033]	[0.034]	[0.033]	[0.032]	[0.034]	[0.033]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	1341	1341	1341	1341	1341	1341	1341	1341	1341
adj. R-sq	0.198	0.199	0.197	0.198	0.199	0.198	0.199	0.200	0.199

Notes: Deviation is measured as the inverse cosine similarity between the publication and grant abstract. GH denotes grant holders. Share new team members (alternative) uses the total count of coauthors in the denominator, while the numerator represents the number of new team members with an academic age of 3 years or more. Robust standard errors in brackets are clustered at the grant level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Table A.3.5: The antecedents of deviation (incl. junior new team members)

DV: *Deviation*	(1)	(2)	(3)	(4)	(5)	(6)
Authors count (ln)	0.004	0.004	0.004	0.007	0.006	0.007
	[0.004]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Share of new team members	0.019***	0.017**	0.017**	0.016**	0.015**	0.016**
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
Share of junior new team members	0.011	0.007	0.004	0.002	-0.001	0.002
	[0.019]	[0.019]	[0.019]	[0.019]	[0.019]	[0.019]
Affiliations of new team members (NTM) (base: Internal)						
Mixed (0/1)		0.003	0.003	0.003	-0.003	0.003
		[0.005]	[0.005]	[0.005]	[0.008]	[0.005]
External (0/1)		0.013**	0.013**	0.013**	0.018**	0.013**
		[0.006]	[0.006]	[0.006]	[0.008]	[0.006]
Distant new team members (0/1)			0.017**	0.016**	0.016**	0.020**
			[0.008]	[0.008]	[0.008]	[0.010]
First collaboration (0/1)				-0.008*	-0.007	-0.007
				[0.004]	[0.007]	[0.005]
First collab. × Affiliations of NTM (baseline: Internal)						
First collab. × Mixed					0.009	
					[0.010]	
First collab. × External					-0.008	
					[0.010]	
First collab. × Distant new team members						-0.007
						[0.010]
GH citation rate (ln; 3yrs pre-grant)	-0.013	-0.014	-0.011	-0.011	-0.012	-0.011
	[0.016]	[0.016]	[0.016]	[0.015]	[0.015]	[0.015]
GH publication rate (ln; 3yrs pre-grant)	0.017**	0.016**	0.018**	0.017**	0.017**	0.017**
	[0.008]	[0.007]	[0.008]	[0.008]	[0.008]	[0.008]
GH multiple funding sources (0/1)	-0.009	-0.009	-0.010*	-0.010*	-0.010	-0.010
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
GH absence (0/1)	0.014	0.015	0.016	0.010	0.013	0.009
	[0.013]	[0.012]	[0.012]	[0.013]	[0.013]	[0.013]
Time to publication (field-norm.)	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Embeddedness in prior research	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Grant productivity (field-norm.)	-0.008	-0.008	-0.011	-0.011	-0.010	-0.011
	[0.009]	[0.009]	[0.010]	[0.010]	[0.010]	[0.010]
Funding source (*Baseline = FWO*)						
Belspo (0/1)	0.067***	0.066***	0.056***	0.056***	0.055***	0.055***
	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
ERC (0/1)	0.004	0.002	0.004	0.004	0.004	0.004
	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Other funder within the EU (0/1)	0.033***	0.027***	0.023**	0.022**	0.024**	0.022**
	[0.008]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]
Other funder outside the EU (0/1)	0.016	0.010	0.004	0.004	0.005	0.004
	[0.012]	[0.012]	[0.012]	[0.011]	[0.012]	[0.011]
Constant	0.262***	0.263***	0.262***	0.261***	0.263***	0.259***
	[0.034]	[0.035]	[0.034]	[0.033]	[0.035]	[0.033]
Year FE	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y
N	1341	1341	1341	1341	1341	1341
adj. R-sq	0.188	0.192	0.197	0.199	0.200	0.199

Notes: Deviation is measured as the inverse cosine similarity between the publication and grant abstract. GH denotes grant holders. Share of junior team members represents the number of new team members with an academic age of less than 3 years over the total number of new team members. Robust standard errors in brackets are clustered at the grant level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Table A.3.6: The antecedents of deviation (alt. first collaboration variables)

DV: *Deviation*	(1)	(2)	(3)	(4)	(5)	(6)
Authors count (ln)	0.006	0.006	0.006	0.004	0.004	0.004
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Share of new team members	0.016**	0.015**	0.016**	0.018**	0.017**	0.018**
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
Affiliations of new team members (NTM) (base: Internal)						
Mixed (0/1)	0.003	0.001	0.003	0.003	-0.003	0.003
	[0.005]	[0.007]	[0.005]	[0.005]	[0.006]	[0.005]
External (0/1)	0.013**	0.020**	0.013**	0.013**	0.011	0.013**
	[0.006]	[0.008]	[0.006]	[0.006]	[0.007]	[0.006]
Distant new team members (0/1)	0.017**	0.016**	0.023**	0.016**	0.017**	0.016*
	[0.008]	[0.008]	[0.010]	[0.008]	[0.008]	[0.009]
First collaboration ALL (0/1)	-0.007*	-0.002	-0.004			
	[0.004]	[0.007]	[0.005]			
First collab. ALL × Affiliation of NTM (baseline: Internal)						
First collab. ALL × Mixed		0.003				
		[0.009]				
First collab. ALL × External		-0.014				
		[0.009]				
First collab. ALL × Distant new team members			-0.013			
			[0.009]			
First collaboration NO FTA (0/1)				-0.004	-0.011	-0.004
				[0.004]	[0.007]	[0.005]
First collab. NO FTA × Affiliation of NTM (baseline: Internal)						
First collab. NO FTA × Mixed					0.016	
					[0.010]	
First collab. NO FTA × External					0.007	
					[0.009]	
First collab. NO FTA × Distant new team members						0.001
						[0.009]
GH citation rate (ln; 3yrs pre-grant)	-0.011	-0.012	-0.011	-0.011	-0.012	-0.011
	[0.016]	[0.015]	[0.016]	[0.016]	[0.016]	[0.016]
GH publication rate (ln; 3yrs pre-grant)	0.017**	0.017**	0.017**	0.018**	0.018**	0.018**
	[0.007]	[0.007]	[0.007]	[0.008]	[0.008]	[0.008]
GH multiple funding sources (0/1)	-0.009	-0.009	-0.009	-0.010*	-0.010*	-0.010*
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
GH absence (0/1)	0.012	0.014	0.010	0.014	0.015	0.014
	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
Time to publication (field-norm.)	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Embeddedness in prior research	-0.019***	-0.018***	-0.019***	-0.019***	-0.019***	-0.019***
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Grant productivity (field-norm.)	-0.011	-0.011	-0.011	-0.011	-0.011	-0.011
	[0.010]	[0.009]	[0.009]	[0.010]	[0.010]	[0.010]
Funding source (Baseline: FWO)						
Belspo (0/1)	0.056***	0.056***	0.055***	0.056***	0.055***	0.056***
	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
ERC (0/1)	0.004	0.004	0.004	0.004	0.004	0.004
	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Other funder within the EU (0/1)	0.022**	0.024***	0.022**	0.023**	0.023**	0.023**
	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]
Other funder outside the EU (0/1)	0.004	0.005	0.004	0.004	0.005	0.004
	[0.011]	[0.012]	[0.011]	[0.012]	[0.012]	[0.012]
Constant	0.261***	0.260***	0.259***	0.262***	0.268***	0.262***
	[0.034]	[0.034]	[0.034]	[0.033]	[0.034]	[0.033]
Year FE	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y
N	1341	1341	1341	1341	1341	1341
adj. R-sq	0.199	0.201	0.200	0.198	0.198	0.197

Notes: First collaboration ALL: 1 if all new team members are first-time collaborators with the grant holder(s) on the publication. *First collaboration* (NO FTA): 1 if any new team member is a first-time collaborator with the grant holder(s) on the publication, net of first time authors. Robust standard errors in brackets are clustered at the grant level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Table A.3.7: The antecedents of deviation (alternative DV: Deviation (SciSpacy))

DV: *Deviation* (SciSpacy)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Authors count (ln)		0.016	0.015	0.009	0.008	0.015	0.013	0.014
		[0.011]	[0.011]	[0.011]	[0.011]	[0.013]	[0.013]	[0.013]
Share of new team members			0.013	0.009	0.011	0.012	0.004	0.009
			[0.018]	[0.019]	[0.019]	[0.019]	[0.021]	[0.019]
Affiliations of new team members (NTM) (base: Internal)								
Mixed (0/1)				0.026**	0.026**	0.026**	0.013	0.027**
				[0.013]	[0.013]	[0.013]	[0.022]	[0.013]
External (0/1)				0.060***	0.059***	0.060***	0.091***	0.060***
				[0.018]	[0.018]	[0.018]	[0.028]	[0.018]
Distant new team members (0/1)					0.018	0.016	0.013	0.040
					[0.016]	[0.016]	[0.016]	[0.027]
First collaboration (0/1)						-0.021	-0.003	-0.009
						[0.015]	[0.020]	[0.019]
First collab. × Affiliations of NTM (baseline: Internal)								
First collab. × Mixed								0.016
								[0.029]
First collab. × External								-0.045
								[0.029]
First collab. × Distant new team members								-0.035
								[0.026]
GH citation rate (ln; 3yrs pre-grant)	-0.018	-0.017	-0.017	-0.022	-0.014	-0.015	-0.015	-0.012
	[0.027]	[0.028]	[0.028]	[0.026]	[0.027]	[0.027]	[0.027]	[0.027]
GH publication rate (ln; 3yrs pre-grant)	0.003	0.003	0.003	0.003	0.003	0.002	0.002	0.003
	[0.018]	[0.018]	[0.018]	[0.017]	[0.017]	[0.017]	[0.017]	[0.018]
GH other grants (0/1)	0.020	0.019	0.019	0.015	0.013	0.014	0.015	0.014
	[0.014]	[0.015]	[0.015]	[0.014]	[0.014]	[0.015]	[0.015]	[0.015]
GH absence (0/1)	0.067**	0.069**	0.067**	0.076**	0.077**	0.059*	0.076**	0.059*
	[0.032]	[0.032]	[0.032]	[0.031]	[0.031]	[0.033]	[0.035]	[0.034]
Time to publication (field-norm.)	-0.009	-0.009	-0.009	-0.008	-0.008	-0.008	-0.008	-0.007
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
Embeddedness in prior research	-0.042***	-0.037***	-0.037***	-0.036***	-0.036***	-0.035***	-0.036***	-0.037***
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.008]	[0.007]
Grant productivity (field-norm.)	0.004	0.004	0.004	-0.002	-0.004	-0.002	-0.002	-0.001
	[0.015]	[0.015]	[0.015]	[0.016]	[0.016]	[0.016]	[0.015]	[0.016]
Funding source (Baseline: FWO)								
Belspo (0/1)	0.103***	0.098***	0.099***	0.095***	0.083***	0.085***	0.087***	0.079***
	[0.025]	[0.025]	[0.025]	[0.024]	[0.025]	[0.025]	[0.025]	[0.026]
ERC (0/1)	0.001	0.001	0.000	0.001	0.005	0.003	0.004	0.003
	[0.027]	[0.025]	[0.025]	[0.024]	[0.024]	[0.024]	[0.023]	[0.023]
Other funder within the EU (0/1)	-0.011	-0.009	-0.010	-0.033*	-0.038**	-0.040**	-0.034*	-0.041**
	[0.018]	[0.017]	[0.017]	[0.019]	[0.019]	[0.019]	[0.019]	[0.019]
Other funder outside the EU (0/1)	-0.006	-0.007	-0.008	-0.034*	-0.038*	-0.039**	-0.034*	-0.039**
	[0.018]	[0.018]	[0.018]	[0.020]	[0.019]	[0.019]	[0.019]	[0.019]
Constant	0.170***	0.136***	0.131***	0.118***	0.117***	0.115***	0.121**	0.102**
	[0.040]	[0.042]	[0.044]	[0.043]	[0.043]	[0.044]	[0.048]	[0.043]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	NO	NO	NO	NO	NO	NO	NO	NO
N	323	323	323	323	323	323	323	323
adj. R-sq	0.263	0.267	0.265	0.295	0.296	0.300	0.307	0.302

Notes: Deviation (SciSpacy) is measured as the inverse cosine similarity between the publication and grant abstract. The vector representations of the abstracts are generated with an alternative language model who is proficient within the Biomedical domain (SciSpacy). The estimation sample includes publications in the Biomedical and Health sciences only. Robust standard errors in brackets are clustered at the grant level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Table A.3.8: The antecedents of deviation (alternative DV: Breakaway)

DV: *Breakaway* (0/1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Authors count (ln)		-0.001	-0.005	-0.006	-0.005	-0.008	-0.010	-0.008
		[0.005]	[0.006]	[0.006]	[0.006]	[0.008]	[0.008]	[0.008]
Share of new team members			0.022**	0.023**	0.022**	0.023**	0.020**	0.023**
			[0.009]	[0.010]	[0.010]	[0.010]	[0.009]	[0.010]
Affiliations of new team members (NTM) (base: Internal)								
Mixed (0/1)				0.005	0.005	0.005	-0.012	0.005
				[0.009]	[0.009]	[0.009]	[0.011]	[0.009]
External (0/1)				-0.001	-0.001	-0.001	-0.009	-0.001
				[0.009]	[0.009]	[0.009]	[0.013]	[0.009]
Distant new team members (0/1)					-0.011	-0.010	-0.011	-0.010*
					[0.007]	[0.007]	[0.007]	[0.005]
First collaboration (0/1)						0.008	-0.006	0.009
						[0.007]	[0.011]	[0.007]
First collab. × Affiliations of NTM (baseline: Internal)								
First collab. × Mixed								0.030***
								[0.011]
First collab. × External								0.016
								[0.013]
First collab. × Distant new team members								-0.001
								[0.008]
GH citation rate (ln; 3yrs pre-grant)	-0.012	-0.012	-0.014	-0.014	-0.016	-0.016	-0.016	-0.016
	[0.010]	[0.010]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]
GH publication rate (ln; 3yrs pre-grant)	-0.011	-0.011	-0.011*	-0.011	-0.012*	-0.011*	-0.011*	-0.011*
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
GH other grants (0/1)	-0.007	-0.007	-0.008	-0.008	-0.007	-0.007	-0.007	-0.007
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
GH absence (0/1)	-0.020*	-0.020*	-0.024**	-0.023**	-0.023**	-0.018	-0.016	-0.018*
	[0.010]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.010]	[0.011]
Time to publication (field-norm.)	-0.002	-0.002	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Embeddedness in prior research	-0.011**	-0.011**	-0.012**	-0.012***	-0.012***	-0.013***	-0.013***	-0.013***
	[0.004]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Grant productivity (field-norm.)	0.016*	0.016*	0.017*	0.017**	0.019**	0.019**	0.020**	0.019**
	[0.009]	[0.009]	[0.009]	[0.009]	[0.010]	[0.010]	[0.010]	[0.010]
Funding source (Baseline: FWO)								
Belspo (0/1)	-0.009	-0.008	-0.009	-0.010	-0.003	-0.003	-0.005	-0.003
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
ERC (0/1)	0.008	0.008	0.008	0.009	0.007	0.007	0.006	0.007
	[0.012]	[0.012]	[0.012]	[0.013]	[0.013]	[0.012]	[0.012]	[0.012]
Other funder within the EU (0/1)	0.002	0.002	-0.001	0.001	0.003	0.004	0.004	0.004
	[0.008]	[0.008]	[0.008]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]
Other funder outside the EU (0/1)	0.000	0.000	-0.001	0.000	0.004	0.004	0.004	0.004
	[0.010]	[0.010]	[0.010]	[0.011]	[0.011]	[0.012]	[0.011]	[0.012]
Constant	0.011	0.013	0.009	0.008	0.009	0.010	0.026*	0.010
	[0.010]	[0.017]	[0.016]	[0.016]	[0.016]	[0.016]	[0.015]	[0.016]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y
N	1341	1341	1341	1341	1341	1341	1341	1341
adj. R-sq	-0.003	-0.004	0.000	-0.001	-0.001	0.000	0.001	-0.001

Notes: Breakaway equals 1 when there are no concepts in common in the publication- and grant-abstract. Robust standard errors in brackets are clustered at the grant level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Table A.3.9: The antecedents deviation and grant openness (Frac. Reg. Logit)

DV: Deviation	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grant openness								
Curiosity	1.032**	1.034***	1.032**	1.034***	1.033***	1.032**	1.032**	1.033**
	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
Grant abstract wordcount (in 100s)	0.961**	0.960**	0.963**	0.962**	0.948***	0.946***	0.947***	0.946***
	[0.018]	[0.018]	[0.018]	[0.018]	[0.018]	[0.018]	[0.018]	[0.018]
Grant abstract wordcount sq. (in 100s)	1.002*	1.002**	1.002*	1.002*	1.003***	1.003***	1.003***	1.003***
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Authors count (ln)		1.044**	1.030	1.030	1.026	1.044*	1.040	1.044*
		[0.023]	[0.023]	[0.023]	[0.023]	[0.026]	[0.026]	[0.026]
Share of new team members			1.078**	1.065*	1.067*	1.065*	1.059*	1.065*
			[0.035]	[0.035]	[0.036]	[0.035]	[0.036]	[0.035]
Affiliations of new team members (NTM) (base: Internal)								
Mixed (0/1)				1.013	1.015	1.015	0.991	1.016
				[0.025]	[0.025]	[0.025]	[0.037]	[0.025]
External (0/1)				1.072**	1.072**	1.074**	1.095**	1.074**
				[0.032]	[0.031]	[0.031]	[0.041]	[0.031]
Distant new team members (0/1)					1.115***	1.114***	1.113***	1.140***
					[0.044]	[0.044]	[0.043]	[0.056]
First collaboration (0/1)						0.955**	0.963	0.963
						[0.020]	[0.034]	[0.024]
First collab. × Affiliations of NTM (baseline: Internal)								
First collab. × Mixed								1.038
								[0.049]
First collab. × External								0.964
								[0.044]
First collab. × Distant new team members								0.963
								[0.042]
GH citation rate (ln; 3yrs pre-grant)	0.931	0.935	0.930	0.928	0.944	0.945	0.942	0.944
	[0.071]	[0.071]	[0.072]	[0.073]	[0.071]	[0.070]	[0.070]	[0.070]
GH publication rate (ln; 3yrs pre-grant)	1.082**	1.086**	1.084**	1.082**	1.089**	1.087**	1.086**	1.088**
	[0.038]	[0.038]	[0.038]	[0.037]	[0.038]	[0.038]	[0.038]	[0.038]
GH multiple funding sources (0/1)	0.954*	0.952*	0.951*	0.949*	0.943**	0.944**	0.945**	0.945**
	[0.026]	[0.026]	[0.026]	[0.026]	[0.026]	[0.026]	[0.026]	[0.026]
GH absence (0/1)	1.079	1.086	1.074	1.079	1.081	1.047	1.061	1.041
	[0.062]	[0.063]	[0.063]	[0.062]	[0.062]	[0.062]	[0.065]	[0.062]
Time to publication (field-norm.)	0.983	0.983	0.981	0.981	0.980	0.980	0.979	0.981
	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]
Embeddedness in prior research	0.910***	0.913***	0.910***	0.911***	0.914***	0.915***	0.914***	0.914***
	[0.014]	[0.014]	[0.014]	[0.014]	[0.015]	[0.015]	[0.015]	[0.014]
Grant productivity (field-norm.)	0.974	0.969	0.970	0.968	0.950	0.949	0.953	0.950
	[0.043]	[0.043]	[0.043]	[0.042]	[0.043]	[0.043]	[0.043]	[0.043]
Funding source (Baseline= FWO)								
Belspo (0/1)	1.475***	1.457***	1.448***	1.437***	1.385***	1.385***	1.384***	1.382***
	[0.115]	[0.115]	[0.113]	[0.110]	[0.105]	[0.106]	[0.107]	[0.105]
ERC (0/1)	1.039	1.042	1.040	1.030	1.051	1.053	1.052	1.052
	[0.054]	[0.054]	[0.053]	[0.052]	[0.054]	[0.054]	[0.054]	[0.054]
Other funder within the EU (0/1)	1.238***	1.237***	1.222***	1.181***	1.169***	1.167***	1.175***	1.166***
	[0.055]	[0.056]	[0.055]	[0.055]	[0.055]	[0.055]	[0.056]	[0.055]
Other funder outside the EU (0/1)	1.109*	1.106*	1.101*	1.065	1.036	1.034	1.039	1.034
	[0.063]	[0.062]	[0.062]	[0.059]	[0.058]	[0.056]	[0.057]	[0.056]
Constant	0.425***	0.391***	0.383***	0.384***	0.394***	0.394***	0.395***	0.390***
	[0.068]	[0.064]	[0.062]	[0.063]	[0.063]	[0.062]	[0.065]	[0.061]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y
N	1341	1341	1341	1341	1341	1341	1341	1341
adj. R-sq	-800.426	-800.362	-800.294	-800.202	-800.033	-799.988	-799.961	-799.980

Notes: Deviation is measured as the inverse cosine similarity between the publication and grant abstract. GH denotes grant holders. We report exponentiated coefficients from fractional regression logit models. Robust standard errors in brackets are clustered at the grant level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Table A.3.10: The antecedents of deviation and grant openness (alternative DV: inverse Jaccard index)

DV: *Inverse Jaccard index*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grant openness								
Curiosity	0.008***	0.008***	0.008***	0.008***	0.008***	0.008***	0.008***	0.008***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Grant abstract wordcount (in 100s)	-0.010***	-0.010***	-0.010***	-0.010***	-0.012***	-0.012***	-0.012***	-0.012***
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Grant abstract wordcount sq. (in 100s)	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Authors count (ln)		-0.003	-0.003	-0.003	-0.004	-0.004	-0.005	-0.004
		[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Share of new team members			0.002	-0.001	-0.001	-0.001	-0.002	-0.001
			[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
Affiliations of new team members (NTM) (base: Internal)								
Mixed (0/1)				0.001	0.001	0.001	-0.005	0.001
				[0.005]	[0.005]	[0.005]	[0.008]	[0.005]
External (0/1)				0.014**	0.014**	0.014**	0.017**	0.014**
				[0.006]	[0.006]	[0.006]	[0.007]	[0.006]
Distant new team members (0/1)					0.014**	0.014**	0.014**	0.018**
					[0.007]	[0.007]	[0.007]	[0.008]
First collaboration (0/1)						0.001	0.001	0.003
						[0.004]	[0.008]	[0.005]
First collab. × Affiliations of NTM (baseline: Internal)								
First collab. × Mixed								0.010
								[0.010]
First collab. × External								-0.005
								[0.009]
First collab. × Distant new team members								-0.006
								[0.008]
GH citation rate (ln; 3yrs pre-grant)	0.002	0.002	0.002	0.002	0.004	0.004	0.003	0.004
	[0.012]	[0.012]	[0.012]	[0.013]	[0.012]	[0.012]	[0.012]	[0.012]
GH publication rate (ln; 3yrs pre-grant)	0.008	0.008	0.008	0.008	0.009	0.009	0.009	0.009
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
GH multiple funding sources (0/1)	0.004	0.004	0.004	0.004	0.003	0.003	0.004	0.004
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
GH absence (0/1)	0.029**	0.028**	0.028**	0.029**	0.029***	0.030**	0.032***	0.029**
	[0.012]	[0.012]	[0.012]	[0.011]	[0.011]	[0.012]	[0.012]	[0.012]
Time to publication (field-norm.)	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Embeddedness in prior research	-0.012***	-0.012***	-0.012***	-0.012***	-0.011***	-0.011***	-0.012***	-0.011***
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Grant productivity (field-norm.)	0.020***	0.021***	0.021***	0.020***	0.018***	0.018***	0.019***	0.018***
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
Funding source (Baseline= FWO)								
Belspo (0/1)	0.036***	0.037***	0.037***	0.036***	0.031**	0.031**	0.030**	0.030**
	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
ERC (0/1)	0.006	0.005	0.005	0.003	0.006	0.006	0.005	0.006
	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]
Other funder within the EU (0/1)	0.022***	0.022***	0.022***	0.015	0.013	0.013	0.015	0.013
	[0.008]	[0.008]	[0.008]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]
Other funder outside the EU (0/1)	-0.007	-0.007	-0.007	-0.014	-0.017*	-0.017*	-0.016*	-0.017*
	[0.009]	[0.009]	[0.009]	[0.009]	[0.010]	[0.010]	[0.010]	[0.010]
Constant	0.914***	0.919***	0.919***	0.920***	0.924***	0.923***	0.926***	0.922***
	[0.033]	[0.033]	[0.033]	[0.031]	[0.029]	[0.029]	[0.030]	[0.029]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y
N	1341	1341	1341	1341	1341	1341	1341	1341
adj. R-sq	0.117	0.117	0.116	0.122	0.126	0.126	0.126	0.125

Notes: Inverse Jaccard index is defined as one minus the ratio of the size of the intersection of concepts between the publication and grant abstracts to the size of their union, providing a measure of dissimilarity between the two abstracts. GH denotes grant holders. Robust standard errors in brackets are clustered at the grant level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

A.4 Supplementary analysis: funding source heterogeneity

In this section, we discuss and provide further evidence of the influence of funding heterogeneity on the extent of *Deviation*. In our main results reported in Table 3, we find that teams funded by the Belgian federal government (Belspo) or other national agencies within the EU, have a higher tendency to deviate, relative to the baseline category of the main funding source in our setting (i.e., the Research Foundation Flanders – FWO). This outcome is consistent with the unique networking objectives of Belspo and other funding sources that are broader in geographic scope. For example, funding by Belspo via the Interuniversity Attraction Poles (IAP) mandates academic collaboration between academics in the two language communities in Belgium (Flanders and Wallonia) and encourages building networks with international partners. The IAP represents 78% of the grant-supported publications in our sample for which the funding source is Belspo.

We further examine funding policy heterogeneity, focusing on FWO Research Projects and ERC's Starting, Consolidator, and Advanced Grants. Based on our own knowledge of the nature of these funding programs and interviews with 8 scientists at KU Leuven (conducted in July–August 2024), these funding sources are similar in their focus on high-quality, investigator-initiated research, with little pressure to include outputs that are only tenuously linked to the grant. We revisit our analysis of the antecedents of deviation, and estimate our models on a restricted sample focusing on the FWO and ERC grants only (Table D1). We find further support for the mechanisms behind deviations and highlight the role of funding heterogeneity. Specifically, we find that in FWO Research Projects, the knowledge recombination mechanism, as captured by the external collaborators and distant new team members are less consistent (Table D1, Models 3–4) while the coordination challenges are more pronounced, as indicated by the strong negative coefficient associated with *first collaboration* (Table D1, Models 3–5). Conversely, in ERC funding, we find strong support for the knowledge recombination mechanism (Table D1, Models 7–10), as indicated by the positive and statistically significant coefficient associated with the *share of new team members*, and particularly by new team members whose prior publications are distant from the grant proposal (*distant new team members*). In contrast, we find no evidence in support for the coordination challenges.

To further assess whether deviations are driven by unique characteristics of each grant, we conduct an analysis of antecedents at the grant level (Table D2) and we estimate grant fixed-effects models (Table D3). To conduct the grant-level analysis, we aggregate variables across all publications linked to

the grant and use the average *Deviation* across all publications linked to the grant as the dependent variable. To estimate the grant fixed-effects models, we exclude grant-level covariates such as the funding source dummies, *grant productivity*, and key explanatory variables that do not variate at the grant-level, including *grant openness* calculated from the linguistic characteristics of the grant proposal and *distant new team members*, which is calculated as the distance of new team members' prior research from the grant proposal. Additionally, we exclude 206 grants that are linked to only a single publication (i.e., singletons).

Our analysis at the grant level shows similar results to the main findings, with the exception of the affiliations of new team members, which become less consistent. This suggests that the direction of the project as a whole is less influenced by the role of external new team members. Nevertheless, we still find a consistently positive and statistically significant role for both the *share of new team members* and the *distant new team members* variables (Table D2, Models 2–8). Finally, focusing on the grant fixed effect models, we find that the *share of new team members* is positively and significantly associated with deviations (Table D3, Models 2–6). This indicates that a higher proportion of new team members contributes to greater deviations from the original research plan, also consistent with our main findings. Examining the affiliations of new team members, we observe that mixed affiliations —where at least one team member is external— are positively and significantly associated with deviations (Table D3, Models 4–6). In line with the results at the grant level, this suggests that the influence of external new team members, established at the publication-grant pair level in our main analysis, is no longer consistent. Instead, when considering the research direction of the grant as a whole (i.e., across its multiple publications), a mix of internal and external affiliations appears to foster deviations. This is consistent with the redefinition of the new team member affiliation variables —aggregated from multiple publications to the grant level— required for these additional analyses.

Regarding first-time collaborations, these do not exhibit a significant association with deviations (Table D3, Models 5–6). Similarly, the interaction terms between *first collaboration* and team member affiliations (mixed or external) are negative but not significant, suggesting that potential coordination challenges associated with first-time collaborations do not meaningfully amplify deviations. As in our main results, *embeddedness in prior research* is a strong and consistent negative predictor of deviations across all models. Overall, these additional analyses provide consistent support for the knowl-

edge recombination mechanism, with a more limited role for coordination challenges in driving deviations.

Table A.4.1: The antecedents deviation (subsample: ERC FWO only)

	FWO – Research Projects				ERC Grants			
	(2)	(3)	(4)	(5)	(7)	(8)	(9)	(10)
Authors count (ln)	0.023**	0.036***	0.036***	0.036***	0.008	0.002	0.002	-0.001
	[0.011]	[0.012]	[0.012]	[0.012]	[0.009]	[0.010]	[0.010]	[0.010]
Share of new team members	0.005	0.007	0.008	0.009	0.026**	0.028**	0.028**	0.026**
	[0.016]	[0.015]	[0.015]	[0.016]	[0.012]	[0.011]	[0.011]	[0.011]
Affiliations of new team members (NTM)								
Mixed (0/1)	0.012	0.010	0.010	0.018	-0.016*	-0.012	-0.012	-0.020
	[0.009]	[0.010]	[0.010]	[0.015]	[0.008]	[0.008]	[0.008]	[0.015]
External (0/1)	0.035*	0.031*	0.031*	0.022	-0.013	-0.008	-0.008	0.001
	[0.018]	[0.018]	[0.018]	[0.021]	[0.008]	[0.009]	[0.009]	[0.012]
Distant new team members (0/1)		0.021	0.055***	0.021		0.059***	0.063***	0.058***
		[0.033]	[0.019]	[0.032]		[0.015]	[0.015]	[0.016]
First collaboration (0/1)		-0.031***	-0.029***	-0.030**		0.009	0.009	0.012
		[0.010]	[0.010]	[0.013]		[0.009]	[0.009]	[0.013]
First collab. x Affiliations of NTM (baseline: Internal)								
First collab. x Mixed				-0.011				0.012
				[0.020]				[0.020]
First collab. x External				0.016				-0.015
				[0.025]				[0.015]
First collab. x Distant new team members		-0.047				-0.005		
		[0.035]				[0.021]		
GH citation rate (ln; 3yrs pre-grant)	-0.012	-0.009	-0.009	-0.007	0.041*	0.043**	0.044**	0.040*
	[0.030]	[0.031]	[0.031]	[0.031]	[0.021]	[0.020]	[0.021]	[0.020]
GH publication rate (ln; 3yrs pre-grant)	-0.008	-0.009	-0.009	-0.009	0.013	0.012	0.013	0.012
	[0.014]	[0.015]	[0.015]	[0.015]	[0.011]	[0.010]	[0.010]	[0.010]
GH other grants (0/1)	-0.002	-0.002	-0.001	-0.002	-0.017*	-0.012	-0.012	-0.011
	[0.012]	[0.011]	[0.011]	[0.011]	[0.010]	[0.009]	[0.009]	[0.009]
GH absence (0/1)	-0.007	-0.025	-0.024	-0.025	-0.019	-0.007	-0.006	-0.006
	[0.024]	[0.025]	[0.025]	[0.025]	[0.029]	[0.030]	[0.031]	[0.030]
Time to publication (field-norm.)	0.005	0.005	0.005	0.005	-0.002	0.000	0.000	0.000
	[0.006]	[0.006]	[0.006]	[0.006]	[0.004]	[0.003]	[0.003]	[0.003]
Embeddedness in prior research	-0.011	-0.010	-0.011	-0.010	-0.018***	-0.020***	-0.020***	-0.020***
	[0.008]	[0.008]	[0.008]	[0.008]	[0.006]	[0.006]	[0.006]	[0.006]
Grant productivity (field-norm.)	0.004	0.002	0.002	0.002	0.013	0.015	0.015	0.018*
	[0.018]	[0.018]	[0.018]	[0.018]	[0.010]	[0.010]	[0.010]	[0.009]
Constant	0.279***	0.278***	0.277***	0.277***	0.212***	0.208***	0.208***	0.208***
	[0.054]	[0.050]	[0.051]	[0.049]	[0.028]	[0.027]	[0.027]	[0.029]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y
N	300	300	300	300	287	287	287	287
adj. R-sq	0.080	0.108	0.108	0.105	0.143	0.170	0.167	0.171

Notes: Deviation is measured as the inverse cosine similarity between the publication and grant abstract. The estimation sample focuses on FWO Research Projects (Models 2-5) and ERC Grants (Models 6-10), which consists ERC's Starting, Consolidator, and Advanced Grants. Robust standard errors in brackets are clustered at the grant level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Table A.4.2: The antecedents of deviation (Grant-level)

DV: Average *Deviation* across publications	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grant-average Authors count (ln)		0.014**	0.007	0.005	0.005	0.007	0.007	0.007
		[0.006]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
Grant-average Share of new team members			0.036***	0.036***	0.036***	0.036***	0.036***	0.036***
			[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
Affiliations of new team members (NTM)								
Mixed (0/1)				0.016*	0.015*	0.015*	0.018*	0.014*
				[0.009]	[0.009]	[0.009]	[0.010]	[0.009]
External (0/1)				0.016	0.013	0.014	0.017	0.014
				[0.011]	[0.011]	[0.011]	[0.013]	[0.011]
Distant new team members (0/1)					0.044**	0.045**	0.045**	0.049
					[0.019]	[0.020]	[0.020]	[0.031]
First collaboration (0/1)						-0.009	0.001	-0.008
						[0.007]	[0.019]	[0.007]
First collab. x Affiliations of NTM (baseline: Internal)								
First collab. x Mixed							-0.011	
							[0.020]	
First collab. x External							-0.012	
							[0.023]	
First collab. x Distant new team members								-0.011
								[0.037]
GH citation rate (ln; 3yrs pre-grant)	-0.006	-0.003	-0.005	-0.005	-0.003	-0.003	-0.004	-0.003
	[0.011]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]
GH publication rate (ln; 3yrs pre-grant)	0.017***	0.017***	0.014**	0.015**	0.016**	0.015**	0.015**	0.015**
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
GH multiple funding (0/1)	0.000	-0.002	-0.002	-0.001	-0.003	-0.003	-0.003	-0.002
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
GH absence (0/1)	0.030**	0.031**	0.025**	0.027**	0.024**	0.021*	0.021*	0.021*
	[0.013]	[0.013]	[0.013]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]
Grant-average Time to publication (field-norm.)	-0.007	-0.006	-0.007	-0.007	-0.006	-0.006	-0.006	-0.006
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Grant-average Embeddedness in prior research	-0.008	-0.008	-0.011*	-0.012*	-0.010	-0.010	-0.010	-0.010
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
Grant productivity (field-norm.)	0.003	0.001	0.002	0.001	0.002	0.000	0.000	0.000
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Funding source (Baseline = FWO)								
Belspo (0/1)	0.049***	0.046***	0.046***	0.045***	0.041***	0.041***	0.042***	0.041***
	[0.013]	[0.013]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]
ERC (0/1)	-0.003	-0.001	-0.001	-0.003	-0.003	-0.004	-0.003	-0.004
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Other funder within the EU (0/1)	0.037***	0.038***	0.035***	0.032***	0.033***	0.031***	0.032***	0.031***
	[0.007]	[0.007]	[0.007]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Other funder outside the EU (0/1)	0.024**	0.022**	0.020*	0.018*	0.017	0.016	0.016	0.016
	[0.010]	[0.010]	[0.010]	[0.011]	[0.010]	[0.010]	[0.010]	[0.010]
Constant	0.212***	0.196***	0.188***	0.176***	0.177***	0.176***	0.172***	0.176***
	[0.014]	[0.015]	[0.016]	[0.018]	[0.018]	[0.018]	[0.020]	[0.018]
Grant start year	Y	Y	Y	Y	Y	Y	Y	Y
Grant Scientific Group	Y	Y	Y	Y	Y	Y	Y	Y
N Grants	460	460	460	460	460	460	460	460
adj. R-sq	0.139	0.149	0.162	0.164	0.179	0.180	0.177	0.179

Notes: All variables are aggregated at the grant-level by taking the average across publications (unless stated otherwise). For the DV, we take the mean of deviations across all publications. We re-define the affiliation dummies as *Internal* (*External*) for cases in which all publications are linked to only internal (external) non grant-holders, while the rest of cases we label as *Mixed*. We also redefine the *First collaboration* dummy for cases in which all publications involve (at least) one first-time collaboration. *Grant Scientific Group* classifies grants into 4 main scientific groups: Biomedical & Health Sci.; Natural Sci., Engineering & Technology; Social Sciences & Humanities; and Multiple Groups. Robust standard errors in brackets. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Table A.4.3: The antecedents of deviation (Grant Fixed-Effects)

DV: *Deviation*	(1)	(2)	(3)	(4)	(5)	(6)
Authors count (ln)		0.007	0.005	0.004	0.005	0.005
		[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Share of new team members			0.011*	0.012**	0.012*	0.013**
			[0.006]	[0.006]	[0.006]	[0.006]
Affiliations of new team members (NTM)						
Mixed (0/1)				0.008*	0.008*	0.014**
				[0.005]	[0.005]	[0.007]
External (0/1)				0.002	0.002	0.006
				[0.005]	[0.005]	[0.007]
First collaboration = 1					-0.003	0.004
					[0.004]	[0.006]
First collab. x Affiliations of NTM (baseline: Internal)						
First collab. x Mixed						-0.011
						[0.008]
First collab. x External						-0.009
						[0.007]
Time to publication (field-norm.)	0.000	0.001	0.001	0.000	0.000	0.000
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Embeddedness in prior research	-0.017***	-0.016***	-0.016***	-0.017***	-0.017***	-0.016***
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Constant	0.289***	0.280***	0.274***	0.273***	0.272***	0.268***
	[0.012]	[0.014]	[0.014]	[0.014]	[0.015]	[0.015]
Grant FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y
N	1135	1135	1135	1135	1135	1135
adj. R-sq	0.577	0.578	0.580	0.580	0.580	0.580

Notes: Deviation is measured as the inverse cosine similarity between the publication and grant abstract. Grant-level covariates, such as funding source dummies, grant productivity, and non-varying explanatory variables (e.g., grant openness and distant new team members), are excluded. The analysis also excludes 206 single-publication grants (singletons). Robust standard errors in brackets are clustered at the grant level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

A.5 Supplementary analyses, consequences of deviation

Table A.5.1: The consequences of deviation (OLS)

DV: *Field Citation Ratio* (ln)	Naïve OLS			
	(1)	(2)	(3)	(4)
Deviation (ln)	-0.029	-0.026	-0.029	-0.025
	[0.111]	[0.111]	[0.111]	[0.111]
Authors count (ln)	0.047	0.033	0.032	0.033
	[0.052]	[0.060]	[0.060]	[0.060]
Share of new team members	-0.104	-0.106	-0.098	-0.106
	[0.088]	[0.088]	[0.089]	[0.088]
Affiliations of new team members (NTM) (baseline: Internal)				
Mixed (0/1)	0.026	0.025	0.072	0.025
	[0.068]	[0.068]	[0.105]	[0.068]
External (0/1)	-0.025	-0.026	0.063	-0.025
	[0.073]	[0.073]	[0.096]	[0.073]
Distant new team members (0/1)	-0.075	-0.074	-0.077	-0.083
	[0.074]	[0.074]	[0.074]	[0.105]
First collaboration (0/1)		0.037	0.140	0.034
		[0.060]	[0.094]	[0.068]
First collab. x Affiliations of NTM (baseline: Internal)				
First collab. x Mixed			-0.098	
			[0.119]	
First collab. x External			-0.173*	
			[0.101]	
First collab. x Distant new team members				0.016
				[0.104]
GH citation rate (ln; 3yrs pre-grant)	0.147	0.146	0.144	0.146
	[0.122]	[0.123]	[0.122]	[0.123]
GH publication rate (ln; 3yrs pre-grant)	0.105	0.107	0.104	0.106
	[0.071]	[0.071]	[0.071]	[0.071]
GH multiple funding (0/1)	0.013	0.013	0.013	0.012
	[0.068]	[0.068]	[0.068]	[0.069]
GH absence (0/1)	0.076	0.101	0.117	0.104
	[0.124]	[0.133]	[0.135]	[0.135]
NTM publication rate (ln; 3yrs pre-grant)	-0.042	-0.038	-0.054	-0.038
	[0.127]	[0.128]	[0.127]	[0.128]
NTM citation rate (ln; 3yrs pre-grant)	0.014	0.020	0.018	0.020
	[0.084]	[0.084]	[0.085]	[0.084]
NTM multiple funding (0/1)	0.056	0.055	0.055	0.056
	[0.068]	[0.068]	[0.068]	[0.068]
Time to publication (field-norm.)	-0.026	-0.026	-0.029	-0.026
	[0.026]	[0.026]	[0.026]	[0.026]
Embeddedness in prior research	0.372***	0.372***	0.373***	0.372***
	[0.044]	[0.044]	[0.044]	[0.044]
Grant productivity (field-norm.)	-0.057	-0.057	-0.053	-0.057
	[0.085]	[0.085]	[0.085]	[0.085]
Funding source (Baseline: FWO)				
Belspo (0/1)	-0.120	-0.121	-0.116	-0.120
	[0.117]	[0.117]	[0.116]	[0.116]
ERC (0/1)	0.224**	0.225**	0.227**	0.225**
	[0.106]	[0.106]	[0.106]	[0.106]
Other funder within the EU (0/1)	0.049	0.051	0.061	0.051
	[0.088]	[0.088]	[0.089]	[0.088]
Other funder outside the EU (0/1)	0.047	0.049	0.058	0.049
	[0.101]	[0.101]	[0.101]	[0.101]
Constant	1.532***	1.540***	1.457***	1.545***
	[0.358]	[0.358]	[0.361]	[0.360]
Year FE	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y
N	1296	1296	1296	1296
adj. R-sq	0.170	0.170	0.170	0.169

Notes: Deviation is measured as the inverse cosine similarity between the publication and grant abstract. GH denotes grant holders. NTM denotes new team members. *Field Citation Ratio* represents relative citation performance compared to similarly aged articles in the same research group (sourced from the Dimensions database, 41 missing values). Robust standard errors in brackets are clustered at the grant level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Table A.5.3: The consequences of deviation (1st stage CFA)

1st stage DV: (ln) *Deviation*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Grant openness									
Curiosity	0.03***	0.03***	0.03***	0.02***	0.02***	0.03***	0.02***	0.02***	0.02***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Grant abstract wordcount (in 100s)	-0.03**	-0.03**	-0.03**	-0.03**	-0.03**	-0.03**	-0.03**	-0.03**	-0.03**
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Grant abstract wordcount sq. (in 100s)	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**	0.00**
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Authors count (ln)	0.05**	0.04**	0.05**	0.04**	0.04**	0.04**	0.04**	0.04*	0.04**
	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]
Share of new team members	0.04	0.04	0.04	0.05*	0.04*	0.05*	0.05*	0.04	0.04*
	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]
Affiliations of new team members (NTM) (baseline: Internal)									
Mixed (0/1)	0.01	-0.01	0.01	0.01	-0.01	0.01	0	-0.02	0
	[0.02]	[0.03]	[0.02]	[0.02]	[0.03]	[0.02]	[0.02]	[0.03]	[0.02]
External (0/1)	0.04*	0.06**	0.04*	0.04*	0.06**	0.04*	0.03	0.05*	0.03
	[0.02]	[0.03]	[0.02]	[0.02]	[0.03]	[0.02]	[0.02]	[0.03]	[0.02]
Distant new team members (0/1)	0.07**	0.07**	0.09***	0.07**	0.07**	0.09***	0.08***	0.08***	0.10***
	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.04]
First collaboration (0/1)	-0.03*	-0.03	-0.03	-0.03**	-0.03	-0.03	-0.03*	-0.02	-0.02
	[0.02]	[0.03]	[0.02]	[0.02]	[0.03]	[0.02]	[0.02]	[0.03]	[0.02]
First collab. x Affiliations of NTM (baseline: Internal)									
First collab. x Mixed		0.03			0.03			0.03	
		[0.04]			[0.04]			[0.04]	
First collab. x External		-0.03			-0.03			-0.04	
		[0.03]			[0.03]			[0.03]	
First collab. x Distant new team members			-0.03			-0.04			-0.03
			[0.03]			[0.03]			[0.03]
GH citation rate (ln; 3yrs pre-grant)	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05	-0.05
	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]	[0.06]
GH publication rate (ln; 3yrs pre-grant)	0.05**	0.05**	0.05**	0.06**	0.06**	0.06**	0.06**	0.06**	0.06**
	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]
GH multiple funding sources (0/1)	-0.04	-0.04	-0.03	-0.04	-0.04	-0.03	-0.03	-0.03	-0.03
	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]
GH absence (0/1)	0.04	0.05	0.03	0.03	0.04	0.03	0.04	0.05	0.04
	[0.05]	[0.05]	[0.05]	[0.05]	[0.05]	[0.05]	[0.05]	[0.05]	[0.05]
NTM publication rate (ln; 3yrs pre-grant)	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]
NTM citation rate (ln; 3yrs pre-grant)	0	0	0	-0.01	-0.01	-0.01	0	-0.01	0
	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.03]	[0.02]
NTM other grants (0/1)	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]	[0.02]
Time to publication (field-norm.)	-0.02	-0.02	-0.02	-0.02	-0.02	-0.01	-0.02	-0.02	-0.02
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Embeddedness in prior research	-0.1***	-0.1***	-0.1***	-0.1***	-0.1***	-0.1***	-0.1***	-0.1***	-0.1***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
Grant productivity (field-norm.)	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.04	-0.03	-0.04
	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]
Funding Source dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y
Constant	-1.3***	-1.3***	-1.4***	-1.3***	-1.3***	-1.4***	-1.3***	-1.3***	-1.4***
	[0.11]	[0.12]	[0.11]	[0.11]	[0.12]	[0.11]	[0.12]	[0.12]	[0.12]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
N	1296	1296	1296	1341	1341	1341	1277	1277	1277
adj. R-sq	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20

Notes: Estimates from the first-stage of a two-step control function approach; second-stage results are available in Table 4. *Deviation* is measured as the inverse cosine similarity between the publication and grant abstract. GH denotes grant holders. NTM denotes new team members. Models 1-3 correspond to the second stage using (ln) *Field Citation Ratio* as DV; Models 4-6 correspond to the second stage using *Top 90* as DV; Models 7-9 corresponds to the second stage using *Novelty* as DV. Robust standard errors in brackets are clustered at the grant level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A.5.2: The consequences of deviation on top citations and novelty (OLS)

DV:	*Top 90* (0/1) – Naïve OLS				*Novelty* (0/1) – Naïve OLS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deviation (ln)	0.019	0.017	0.016	0.018	0.001	0.002	0.004	0.002
	[0.030]	[0.030]	[0.030]	[0.030]	[0.031]	[0.032]	[0.031]	[0.032]
Authors count (ln)	-0.004	0.002	0.001	0.002	0.021	0.017	0.019	0.017
	[0.018]	[0.020]	[0.019]	[0.020]	[0.013]	[0.015]	[0.015]	[0.015]
Share of new team members	-0.026	-0.025	-0.024	-0.025	0.022	0.021	0.023	0.021
	[0.032]	[0.032]	[0.033]	[0.032]	[0.020]	[0.020]	[0.020]	[0.020]
Affiliations of new team members (NTM) (baseline: Internal)								
Mixed (0/1)	0.031	0.031	0.040	0.030	0.035**	0.035**	0.043*	0.035**
	[0.022]	[0.022]	[0.034]	[0.022]	[0.017]	[0.017]	[0.026]	[0.017]
External (0/1)	0.006	0.006	0.033	0.007	0.013	0.013	-0.003	0.013
	[0.024]	[0.024]	[0.031]	[0.024]	[0.021]	[0.021]	[0.026]	[0.021]
Distant new team members (0/1)	-0.027	-0.028	-0.028	-0.045	-0.015	-0.015	-0.015	-0.020
	[0.023]	[0.023]	[0.023]	[0.031]	[0.019]	[0.019]	[0.019]	[0.023]
First collaboration (0/1)		-0.015	0.014	-0.022		0.009	-0.002	0.007
		[0.020]	[0.031]	[0.023]		[0.016]	[0.028]	[0.018]
First collab. x Affiliations of NTM (baseline: Internal)								
First collab. x Mixed			-0.023				-0.011	
			[0.042]				[0.036]	
First collab. x External			-0.052				0.030	
			[0.036]				[0.034]	
First collab. x Distant new team members				0.029				0.008
				[0.035]				[0.030]
GH citation rate (ln; 3yrs pre-grant)	0.015	0.015	0.014	0.015	0.040	0.040	0.040	0.040
	[0.037]	[0.037]	[0.037]	[0.037]	[0.042]	[0.042]	[0.042]	[0.042]
GH publication rate (ln; 3yrs pre-grant)	0.005	0.005	0.004	0.005	-0.020	-0.020	-0.020	-0.020
	[0.021]	[0.021]	[0.021]	[0.021]	[0.017]	[0.017]	[0.017]	[0.017]
GH multiple funding (0/1)	0.018	0.018	0.018	0.016	-0.029	-0.029	-0.029	-0.030
	[0.022]	[0.022]	[0.022]	[0.022]	[0.020]	[0.020]	[0.020]	[0.020]
GH absence (0/1)	0.038	0.027	0.034	0.031	-0.052	-0.046	-0.053	-0.045
	[0.035]	[0.038]	[0.038]	[0.037]	[0.032]	[0.034]	[0.035]	[0.033]
NTM publication rate (ln; 3yrs pre-grant)	-0.026	-0.028	-0.033	-0.027	-0.004	-0.003	0.001	-0.003
	[0.037]	[0.037]	[0.038]	[0.037]	[0.040]	[0.039]	[0.040]	[0.039]
NTM citation rate (ln; 3yrs pre-grant)	0.014	0.011	0.011	0.012	-0.001	0.000	0.001	0.000
	[0.026]	[0.026]	[0.026]	[0.026]	[0.022]	[0.022]	[0.022]	[0.023]
NTM multiple funding (0/1)	-0.024	-0.023	-0.024	-0.023	0.049**	0.049**	0.048**	0.049**
	[0.021]	[0.021]	[0.021]	[0.021]	[0.020]	[0.020]	[0.020]	[0.020]
Time to publication (field-norm.)	-0.005	-0.005	-0.006	-0.006	0.003	0.003	0.004	0.003
	[0.009]	[0.009]	[0.009]	[0.009]	[0.011]	[0.011]	[0.011]	[0.011]
Embeddedness in prior research	0.083***	0.083***	0.084***	0.084***	-0.017*	-0.017*	-0.017*	-0.017*
	[0.017]	[0.017]	[0.017]	[0.017]	[0.010]	[0.010]	[0.010]	[0.010]
Grant productivity (field-norm.)	-0.011	-0.011	-0.010	-0.012	-0.016	-0.016	-0.018	-0.016
	[0.020]	[0.020]	[0.020]	[0.020]	[0.021]	[0.021]	[0.021]	[0.021]
Funding source (Baseline: FWO)								
Belspo (0/1)	0.043	0.043	0.044	0.045	0.064*	0.064*	0.065*	0.065*
	[0.035]	[0.035]	[0.035]	[0.035]	[0.033]	[0.033]	[0.033]	[0.033]
ERC (0/1)	0.072**	0.072**	0.073**	0.073**	0.034*	0.034*	0.034*	0.034*
	[0.029]	[0.029]	[0.029]	[0.029]	[0.019]	[0.019]	[0.019]	[0.019]
Other funder within the EU (0/1)	0.046*	0.046*	0.049*	0.046*	0.001	0.002	-0.003	0.002
	[0.027]	[0.027]	[0.027]	[0.027]	[0.023]	[0.023]	[0.024]	[0.023]
Other funder outside the EU (0/1)	0.040	0.039	0.042	0.039	0.017	0.018	0.015	0.018
	[0.030]	[0.030]	[0.029]	[0.030]	[0.027]	[0.027]	[0.027]	[0.027]
Constant	-0.061	-0.064	-0.088	-0.056	0.113	0.115	0.120	0.117
	[0.073]	[0.073]	[0.073]	[0.076]	[0.153]	[0.154]	[0.150]	[0.153]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y
N	1341	1341	1341	1341	1277	1277	1277	1277
adj. R-sq	0.109	0.109	0.109	0.109	0.033	0.032	0.032	0.032

Notes: Deviation is measured as the inverse cosine similarity between the publication and grant abstract. GH denotes grant holders. NTM denotes new team members. *Top 10* equals 1 if a publication is in the top 10% of the citation distribution by Field of Research and year of publication in our data. *Novelty* equals 1 if a publication introduces at least one new word to the corpus of science, and the word is used at least twice (adapted from Arts et al., 2023; 64 missing values). Robust standard errors, clustered at the grant level, are shown in brackets. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A.5.4: The consequences of deviation (DV: Citations)

DV: *Citations*	Naïve Poisson						Poisson CFA					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Deviation (ln)	-0.143	-0.123	-0.110	-0.109	-0.100	-0.101	2.305*	2.879*	2.595*	2.532*	2.526*	2.534*
	[0.156]	[0.156]	[0.156]	[0.156]	[0.157]	[0.156]	[1.399]	[1.622]	[1.473]	[1.482]	[1.474]	[1.458]
Authors count (ln)	0.125	0.152*	0.136	0.135	0.147*	0.148*	0.018	0.067	0.059	0.024	0.041	0.020
	[0.083]	[0.081]	[0.088]	[0.090]	[0.088]	[0.089]	[0.096]	[0.094]	[0.092]	[0.105]	[0.101]	[0.102]
Share of new team members		-0.168*	-0.145	-0.145	-0.137	-0.146		-0.384**	-0.320**	-0.316**	-0.296**	-0.353**
		[0.092]	[0.093]	[0.093]	[0.096]	[0.118]		[0.161]	[0.147]	[0.145]	[0.144]	[0.152]
Affiliations of new team members (NTM) (baseline: Internal NGH)												
Mixed (0/1)			-0.073	-0.072	-0.138	-0.141			-0.095	-0.094	-0.202	-0.218
			[0.096]	[0.096]	[0.153]	[0.150]			[0.098]	[0.097]	[0.156]	[0.156]
External (0/1)			-0.166*	-0.166*	-0.269*	-0.271*			-0.253**	-0.249**	-0.381**	-0.369**
			[0.094]	[0.094]	[0.145]	[0.146]			[0.113]	[0.112]	[0.186]	[0.183]
First collaboration (0/1)				0.004	-0.104	-0.105				0.105	-0.053	-0.044
				[0.077]	[0.129]	[0.129]				[0.095]	[0.151]	[0.152]
Affiliation of NTM (baseline: Internal x First collab.)												
Mixed x First collab.					0.105	0.108					0.183	0.187
					[0.164]	[0.164]					[0.184]	[0.186]
External x First collab.					0.168	0.173					0.221	0.230
					[0.159]	[0.159]					[0.208]	[0.209]
NTM citation rate (ln; 3yrs pre-grant)						0.064						-0.022
						[0.160]						[0.191]
NTM wt. publication rate (ln; 3yrs pre-grant)						0.011						0.141
						[0.122]						[0.124]
NTM multiple funding sources (0/1)						-0.019						0.032
						[0.085]						[0.095]
GH wt. citation rate (ln; 3yrs pre-grant)	0.124	0.129	0.139	0.139	0.144	0.132	0.277	0.336*	0.347*	0.339*	0.343*	0.354*
	[0.154]	[0.151]	[0.147]	[0.151]	[0.150]	[0.153]	[0.180]	[0.202]	[0.194]	[0.191]	[0.190]	[0.191]
GH publication rate (ln; 3yrs pre-grant)	-0.005	0.002	0.004	0.004	0.001	-0.002	-0.063	-0.082	-0.045	-0.040	-0.039	-0.044
	[0.099]	[0.097]	[0.097]	[0.098]	[0.098]	[0.098]	[0.126]	[0.139]	[0.127]	[0.126]	[0.125]	[0.123]
GH multiple funding sources (0/1)	0.036	0.041	0.048	0.048	0.046	0.055	0.076	0.106	0.103	0.101	0.095	0.077
	[0.088]	[0.087]	[0.086]	[0.086]	[0.085]	[0.090]	[0.094]	[0.102]	[0.098]	[0.097]	[0.096]	[0.103]
GH absence (0/1)	-0.041	-0.016	-0.025	-0.023	-0.061	-0.054	-0.026	0.015	0.039	0.111	0.071	0.051
	[0.150]	[0.147]	[0.147]	[0.152]	[0.160]	[0.162]	[0.194]	[0.208]	[0.194]	[0.193]	[0.199]	[0.211]
Time to publication (field-norm.)	-0.047	-0.044	-0.046	-0.045	-0.044	-0.043	-0.016	0.003	0.000	0.000	0.000	0.009
	[0.041]	[0.040]	[0.041]	[0.041]	[0.040]	[0.040]	[0.052]	[0.057]	[0.055]	[0.055]	[0.055]	[0.055]
Embeddedness in prior research	0.450***	0.453***	0.456***	0.457***	0.461***	0.460***	0.650***	0.709***	0.686***	0.678***	0.682***	0.686***
	[0.064]	[0.064]	[0.064]	[0.064]	[0.063]	[0.063]	[0.117]	[0.136]	[0.125]	[0.124]	[0.124]	[0.123]
Grant productivity (field-norm.)	-0.096	-0.100	-0.093	-0.093	-0.099	-0.100	-0.026	-0.022	-0.016	-0.016	-0.025	-0.023
	[0.097]	[0.097]	[0.096]	[0.096]	[0.095]	[0.095]	[0.107]	[0.118]	[0.111]	[0.110]	[0.108]	[0.110]
Funding source (Baseline: FWO)												
Belspo (0/1)	-0.108	-0.105	-0.107	-0.107	-0.104	-0.098	-0.753**	-0.880**	-0.814**	-0.801**	-0.791**	-0.790**
	[0.168]	[0.166]	[0.167]	[0.167]	[0.166]	[0.168]	[0.376]	[0.427]	[0.386]	[0.389]	[0.386]	[0.388]
ERC (0/1)	0.349***	0.346***	0.358***	0.358***	0.361***	0.365***	0.314**	0.298*	0.320**	0.321**	0.325**	0.339**
	[0.129]	[0.127]	[0.125]	[0.126]	[0.124]	[0.124]	[0.151]	[0.161]	[0.152]	[0.151]	[0.151]	[0.153]
Other funder within the EU (0/1)	-0.029	-0.015	0.044	0.044	0.029	0.032	-0.301	-0.332	-0.198	-0.185	-0.204	-0.211
	[0.114]	[0.113]	[0.122]	[0.123]	[0.123]	[0.124]	[0.221]	[0.240]	[0.207]	[0.205]	[0.213]	[0.214]
Other funder outside the EU (0/1)	-0.115	-0.100	-0.048	-0.048	-0.058	-0.055	-0.163	-0.167	-0.063	-0.058	-0.071	-0.063
	[0.126]	[0.124]	[0.126]	[0.127]	[0.125]	[0.125]	[0.164]	[0.183]	[0.164]	[0.161]	[0.160]	[0.159]
Constant	4.052***	4.110***	4.177***	4.178***	4.234***	4.226***	7.514***	8.351***	7.958***	7.882***	7.908***	7.948***
	[0.423]	[0.423]	[0.475]	[0.475]	[0.486]	[0.488]	[1.998]	[2.321]	[2.106]	[2.123]	[2.138]	[2.127]
N	1341	1341	1341	1341	1341	1341	1341	1341	1341	1341	1341	1341
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Log Like.	-24264.9	-24195.9	-24092.5	-24092.4	-24064.3	-24059.8						
1st stage residual							-2.439*	-3.003*	-2.707*	-2.641*	-2.631*	-2.638*
							[1.407]	[1.627]	[1.479]	[1.487]	[1.479]	[1.463]

Notes: Citations represents total citation count as of December 2022 (sourced from the Dimensions database). *Deviation* is measured as the inverse cosine similarity between the publication and grant abstract. GH denotes grant holders. NTM denotes new team members. Robust standard errors in brackets are clustered at the grant level. Significance levels: * p<0.10, ** p<0.05, *** p<0.01.

Table A.5.5: The consequences of deviation (sample includes most productive grants)

	1st stage DV: (ln) *Deviation*				2nd stage DV: (ln) *Field Citation Ratio*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deviation (ln)					2.124**	2.055**	2.073**	2.045**
					[0.951]	[0.960]	[0.979]	[0.960]
Grant openness								
Curiosity	0.024***	0.024***	0.023***	0.024***				
	[0.009]	[0.009]	[0.009]	[0.009]				
Grant abstract wordcount (in 100s)	-0.030**	-0.031**	-0.031**	-0.031**				
	[0.014]	[0.014]	[0.014]	[0.014]				
Grant abstract wordcount sq. (in 100s)	0.002**	0.002**	0.002**	0.002**				
	[0.001]	[0.001]	[0.001]	[0.001]				
Authors count (ln)	0.032*	0.041**	0.039**	0.041**	-0.010	-0.050	-0.046	-0.050
	[0.017]	[0.018]	[0.018]	[0.018]	[0.058]	[0.068]	[0.066]	[0.068]
Share of new team members	0.050*	0.051**	0.048*	0.051**	-0.239**	-0.241**	-0.228**	-0.241**
	[0.026]	[0.026]	[0.026]	[0.026]	[0.103]	[0.104]	[0.102]	[0.104]
Affiliations of new team members (NTM) (baseline: Internal)								
Mixed (0/1)	0.006	0.006	-0.013	0.006	0.016	0.016	0.096	0.016
	[0.018]	[0.018]	[0.028]	[0.018]	[0.067]	[0.067]	[0.107]	[0.067]
External (0/1)	0.031	0.031	0.036	0.031	-0.097	-0.094	-0.034	-0.092
	[0.020]	[0.020]	[0.025]	[0.020]	[0.078]	[0.078]	[0.107]	[0.078]
Distant new team members (0/1)	0.093***	0.092***	0.091***	0.090***	-0.133	-0.124	-0.124	-0.153
	[0.028]	[0.028]	[0.028]	[0.033]	[0.110]	[0.111]	[0.111]	[0.123]
First collaboration (0/1)		-0.026	-0.031	-0.027		0.120*	0.189**	0.110*
		[0.017]	[0.024]	[0.018]		[0.062]	[0.091]	[0.066]
First collab x Affiliation of NTM (baseline: Internal x First collab.)								
First collab x Mixed			0.033				-0.144	
			[0.034]				[0.119]	
First collab x External			-0.009				-0.068	
			[0.032]				[0.098]	
First collab. x Distant new team members				0.004				0.054
				[0.033]				[0.118]
GH citation rate (ln; 3yrs pre-grant)	-0.066	-0.065	-0.065	-0.065	0.277**	0.270**	0.270**	0.270**
	[0.057]	[0.056]	[0.056]	[0.056]	[0.130]	[0.131]	[0.132]	[0.131]
GH publication rate (ln; 3yrs pre-grant)	0.053**	0.052**	0.052*	0.052**	-0.017	-0.010	-0.011	-0.009
	[0.027]	[0.027]	[0.027]	[0.027]	[0.090]	[0.090]	[0.090]	[0.090]
GH multiple funding sources (0/1)	-0.030	-0.030	-0.029	-0.030	0.051	0.049	0.048	0.046
	[0.023]	[0.023]	[0.023]	[0.023]	[0.070]	[0.070]	[0.070]	[0.071]
GH absence (0/1)	0.039	0.021	0.027	0.021	-0.101	-0.013	-0.024	-0.002
	[0.040]	[0.042]	[0.043]	[0.043]	[0.125]	[0.130]	[0.135]	[0.132]
NTM publication rate (ln; 3yrs pre-grant)	0.055	0.053	0.050	0.053	-0.120	-0.106	-0.108	-0.103
	[0.041]	[0.042]	[0.042]	[0.041]	[0.135]	[0.135]	[0.135]	[0.136]
NTM citation rate (ln; 3yrs pre-grant)	0.011	0.008	0.007	0.008	0.029	0.043	0.047	0.043
	[0.023]	[0.023]	[0.023]	[0.023]	[0.078]	[0.078]	[0.079]	[0.078]
NTM other grants (0/1)	-0.018	-0.017	-0.017	-0.017	0.106	0.101	0.099	0.104
	[0.020]	[0.020]	[0.020]	[0.020]	[0.066]	[0.066]	[0.066]	[0.066]
Time to publication (field-norm.)	-0.014	-0.014	-0.014	-0.014	0.001	0.002	0.001	0.001
	[0.009]	[0.009]	[0.009]	[0.009]	[0.027]	[0.027]	[0.027]	[0.027]
Embeddedness in prior research	-0.066***	-0.065***	-0.065***	-0.065***	0.543***	0.534***	0.538***	0.533***
	[0.011]	[0.011]	[0.011]	[0.011]	[0.074]	[0.073]	[0.075]	[0.073]
Grant productivity (field-norm.)	0.001	0.001	0.001	0.001	-0.033	-0.033	-0.034	-0.033
	[0.016]	[0.016]	[0.016]	[0.016]	[0.042]	[0.042]	[0.041]	[0.042]
Funding source dummies	Y	Y	Y	Y	Y	Y	Y	Y
Constant	-1.430***	-1.428***	-1.419***	-1.427***	4.694***	4.593***	4.552***	4.591***
	[0.098]	[0.097]	[0.102]	[0.097]	[1.450]	[1.465]	[1.481]	[1.462]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y
N	1472	1472	1472	1472	1472	1472	1472	1472
adj. R-sq	0.236	0.237	0.237	0.236	0.178	0.178	0.178	0.178
1st stage residual					-2.137**	-2.063**	-2.081**	-2.053**
					[0.956]	[0.965]	[0.983]	[0.965]

Notes: Deviation is measured as the inverse cosine similarity between the publication and grant abstract. *Field Citation Ratio* represents relative citation performance compared to similarly aged articles in the same research group (sourced from the Dimensions database, 41 missing values). GH denotes grant holders. NTM denotes new team members. The sample with most productive grants include grants that produce up to 141 publications. Robust standard errors in brackets are clustered at the grant level. Significance levels:* p<0.10, ** p<0.05, *** p<0.01.

Table A.5.6: The consequences of deviation (sample includes fast publications)

	1st stage DV: (ln) *Deviation*				2nd stage DV: (ln) *Field Citation Ratio*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deviations (ln)					2.057*	1.932*	1.902	1.882
					[1.181]	[1.160]	[1.189]	[1.146]
Grant openness								
Curiosity	0.018**	0.018**	0.018*	0.018**				
	[0.009]	[0.009]	[0.009]	[0.009]				
Grant abstract wordcount (in 100s)	-0.026*	-0.027**	-0.027**	-0.028**				
	[0.014]	[0.014]	[0.014]	[0.014]				
Grant abstract wordcount sq. (in 100s)	0.002*	0.002**	0.002**	0.002**				
	[0.001]	[0.001]	[0.001]	[0.001]				
Authors count (ln)	0.026	0.040**	0.038**	0.040**	0.003	-0.035	-0.033	-0.033
	[0.017]	[0.019]	[0.019]	[0.019]	[0.060]	[0.074]	[0.073]	[0.074]
Share of new team members	0.057**	0.058**	0.055**	0.059**	-0.251**	-0.247**	-0.235**	-0.245**
	[0.025]	[0.025]	[0.025]	[0.025]	[0.114]	[0.114]	[0.112]	[0.114]
Affiliations of new team members (NTM) (baseline: Internal)								
Mixed (0/1)	0.006	0.006	-0.009	0.006	0.003	0.004	0.065	0.004
	[0.019]	[0.019]	[0.030]	[0.019]	[0.063]	[0.063]	[0.100]	[0.064]
External (0/1)	0.026	0.026	0.038	0.026	-0.102	-0.100	-0.027	-0.098
	[0.023]	[0.023]	[0.028]	[0.023]	[0.075]	[0.075]	[0.101]	[0.075]
Distant new team members (0/1)	0.062**	0.061**	0.060**	0.084**	-0.172*	-0.162*	-0.163*	-0.206
	[0.028]	[0.028]	[0.028]	[0.033]	[0.097]	[0.096]	[0.096]	[0.131]
First collaboration (0/1)		-0.035**	-0.031	-0.028		0.105	0.200**	0.088
		[0.017]	[0.026]	[0.019]		[0.069]	[0.099]	[0.071]
First collab x Affiliation of NTM (baseline: Internal x First collab.)								
First collab x Mixed			0.023				-0.120	
			[0.037]				[0.119]	
First collab x External			-0.021				-0.143	
			[0.033]				[0.103]	
First collab. x Distant new team members				-0.037				0.078
				[0.032]				[0.105]
GH citation rate (ln; 3yrs pre-grant)	-0.067	-0.064	-0.065	-0.065	0.327**	0.314**	0.311**	0.311**
	[0.058]	[0.057]	[0.057]	[0.057]	[0.143]	[0.142]	[0.143]	[0.141]
GH publication rate (ln; 3yrs pre-grant)	0.049**	0.049**	0.048*	0.049**	-0.001	0.008	0.008	0.010
	[0.025]	[0.025]	[0.025]	[0.025]	[0.091]	[0.090]	[0.091]	[0.089]
GH multiple funding sources (0/1)	-0.033	-0.033	-0.033	-0.031	0.103	0.099	0.098	0.094
	[0.023]	[0.023]	[0.023]	[0.023]	[0.069]	[0.069]	[0.069]	[0.069]
GH absence (0/1)	0.056	0.032	0.040	0.026	-0.012	0.066	0.075	0.080
	[0.045]	[0.047]	[0.048]	[0.047]	[0.142]	[0.138]	[0.145]	[0.138]
NTM publication rate (ln; 3yrs pre-grant)	0.018	0.015	0.010	0.014	-0.138	-0.125	-0.137	-0.123
	[0.046]	[0.046]	[0.046]	[0.046]	[0.116]	[0.116]	[0.114]	[0.116]
NTM citation rate (ln; 3yrs pre-grant)	-0.019	-0.023	-0.025	-0.024	0.125	0.137*	0.135	0.137*
	[0.023]	[0.023]	[0.023]	[0.023]	[0.081]	[0.082]	[0.085]	[0.082]
NTM other grants (0/1)	-0.011	-0.010	-0.009	-0.011	0.058	0.054	0.053	0.055
	[0.020]	[0.020]	[0.020]	[0.020]	[0.064]	[0.064]	[0.064]	[0.064]
Time to publication (field-norm.)	-0.020*	-0.020*	-0.021*	-0.020*	0.020	0.019	0.016	0.017
	[0.011]	[0.011]	[0.011]	[0.011]	[0.034]	[0.034]	[0.035]	[0.033]
Embeddedness in prior research	-0.075***	-0.073***	-0.074***	-0.074***	0.529***	0.516***	0.516***	0.513***
	[0.012]	[0.012]	[0.012]	[0.012]	[0.097]	[0.095]	[0.097]	[0.094]
Grant productivity (field-norm.)	-0.006	-0.007	-0.005	-0.006	-0.037	-0.036	-0.034	-0.038
	[0.032]	[0.032]	[0.032]	[0.032]	[0.080]	[0.080]	[0.080]	[0.079]
Funding source dummies	Y	Y	Y	Y	Y	Y	Y	Y
Constant	-1.344***	-1.344***	-1.340***	-1.352***	4.464***	4.302**	4.178**	4.250**
	[0.113]	[0.110]	[0.115]	[0.111]	[1.718]	[1.695]	[1.726]	[1.684]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Fields of Research FE	Y	Y	Y	Y	Y	Y	Y	Y
N	1422	1422	1422	1422	1422	1422	1422	1422
adj. R-sq	0.188	0.190	0.190	0.190	0.175	0.175	0.175	0.174
1st stage residual					-2.125*	-1.996*	-1.969	-1.945*
					[1.189]	[1.167]	[1.196]	[1.154]

Notes: Field Citation Ratio represents relative citation performance compared to similarly aged articles in the same research group (sourced from the Dimensions database, 41 missing values). *Deviation* is measured as the inverse cosine similarity between the publication and grant abstract. GH denotes grant holders. NTM denotes new team members. The sample with fast publications include articles with a publication date up to 70 days after the grant start date. Robust standard errors in brackets are clustered at the grant level. Significance levels:* p<0.10, ** p<0.05, *** p<0.01.

B Appendix Chapter 3

B.1 Supplementary Descriptives

Table B.1.1: Pairwise Correlations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)
(1) d_{ij}	1																						
(2) Pre-doc. Pubs	0.07*	1																					
(3) Co-Advisor	0.05*	0.03*	1																				
(4) Nr. Advisee	-0.06*	-0.01*	-0.09*	1																			
(5) Advisor Productivity	0.11*	0.02*	0.03*	0.06*	1																		
(6) Advisor Coauthors	0.20*	0.09*	0.09*	0.19*	0.29*	1																	
(7) Full Prof.	0.03*	0.01	0.00	0.08*	0.05*	0.11*	1																
(8) Adv. Research Breadth	0.15*	0.02*	0.02*	0.20*	0.47*	0.29*	0.06*	1															
(9) PhD Cohort	0.10*	0.08*	-0.08*	0.08*	0.08*	0.11*	-0.17*	0.26*	1														
(10) Year	0.07*	0.01	-0.05*	0.07*	0.07*	0.08*	-0.11*	0.21*	0.79*	1													
(11) Other Fields	-0.16*	-0.05*	-0.11*	0.01	-0.10*	-0.18*	-0.03*	-0.06*	0.0181*	0.0452*	1												
(12) Biology	0.03*	-0.04*	-0.01	-0.00	0.05*	0.09*	0.04*	0.00	-0.05*	-0.03*	-0.07*	1											
(13) Biomedical	0.29*	0.15*	0.19*	-0.17*	0.10*	0.25*	0.09*	0.02*	-0.12*	-0.08*	-0.33*	0.05*	1										
(14) Chemical	-0.09*	-0.00	-0.06*	0.13*	0.04*	0.09*	0.02*	0.04*	0.05*	0.06*	-0.03*	0.02*	-0.17*	1									
(15) Health	-0.16*	-0.02*	-0.06*	0.19*	-0.06*	-0.04*	-0.03*	0.04*	0.08*	0.07*	0.05*	-0.17*	-0.33*	0.20*	1								
(16) Engineering	0.03*	0.08*	0.11*	-0.08*	0.00	-0.00	0.00	0.01	-0.06*	-0.01	-0.10*	-0.19*	0.23*	-0.11*	-0.12*	1							
(17) Computing	-0.05*	-0.00	-0.05*	0.05*	-0.08*	-0.11*	-0.02*	0.01	0.03*	0.04*	0.00	-0.12*	-0.16*	-0.08*	0.13*	-0.02*	1						
(18) Psychology	0.15*	-0.04*	0.05*	-0.08*	0.06*	-0.06*	-0.00	0.05*	-0.07*	-0.03*	-0.07*	-0.11*	0.06*	-0.11*	-0.19*	0.10*	-0.04*	1					
(19) Publications	0.05*	0.16*	0.07*	-0.03*	-0.00	0.09*	0.06*	-0.04*	-0.16*	-0.02*	0.10*	0.04*	0.26*	0.05*	0.08*	0.19*	0.05*	0.02*	1				
(20) Coauthor-wt. Pubs.	-0.03*	0.10*	-0.01	-0.02*	-0.06*	-0.08*	0.04*	-0.09*	-0.18*	-0.06*	0.19*	-0.03*	0.08*	0.00	0.11*	0.15*	0.14*	0.05*	0.79*	1			
(21) Citations Avg.	0.04*	-0.02*	-0.02*	0.02*	0.05*	0.05*	0.00	0.03*	0.06*	0.07*	-0.04*	0.09*	-0.01	0.03*	-0.00	-0.03*	-0.01	-0.00	-0.02*	-0.05*	1		
(22) New Words	0.04*	0.04*	0.02*	-0.03*	-0.00	0.05*	0.02*	-0.01	-0.14*	-0.11*	-0.01	0.07*	0.15*	0.01	-0.03*	0.03*	-0.00	-0.03*	0.27*	0.17*	0.02		
(23) New Phrases	0.02	0.02*	0.02*	-0.01	-0.00	0.05*	0.02*	-0.03*	-0.20*	-0.18*	0.00	0.10*	0.15*	0.04*	0.04*	0.02*	0.00	-0.06*	0.33*	0.22*	0.04*	1	

Notes: Significance levels: * $p < 0.01$

B.2 Supplementary Analyses

Advisors' ID Engagement (continuous)

We reestimate our main models of PhD students' productivity and novelty, replacing the binary treatment d_{ij} with a continuous measure defined as the number of publications by the advisor that mention the center in the five years before the student's PhD start. In the tables below, we name this variable *Advisor ID Engagement*. Descriptively, advisors published on average 0.62 papers per year mentioning the center within the five years prior to the PhD start (SD = 1.61, max = 14.4).

Table B.2.1: PhD Students' Productivity (*Advisor ID Engagement* instead of d_{ij})

	Publications						Avg. Citations per Pub.			
	Simple count		Coauthor weighted		with Advisor		All pubs		with Advisor	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Advisor ID Engagement	0.005	-0.041***	-0.004	-0.025**	-0.003	-0.044*	-0.004	-0.016	-0.051*	-0.059*
	[0.020]	[0.011]	[0.012]	[0.011]	[0.028]	[0.024]	[0.023]	[0.020]	[0.030]	[0.035]
Pre-doctoral Pubs (0/1)	0.262***	0.181***	0.222***	0.147***	0.003	-0.056	0.011	-0.010	-0.011	-0.050
	[0.030]	[0.033]	[0.028]	[0.031]	[0.060]	[0.053]	[0.043]	[0.048]	[0.047]	[0.055]
Co-Advisor (0/1)	0.065**	-0.000	0.009	-0.027	0.058	0.068	-0.005	0.027	-0.030	-0.030
	[0.026]	[0.026]	[0.022]	[0.025]	[0.055]	[0.057]	[0.047]	[0.055]	[0.054]	[0.058]
Nr. Advisees (log)	-0.003		-0.015		-0.013		0.011		0.052	
	[0.027]		[0.022]		[0.055]		[0.040]		[0.044]	
Advisor Productivity (log)	-0.085***		-0.060***		-0.175***		0.148***		0.156***	
	[0.024]		[0.020]		[0.044]		[0.039]		[0.049]	
Advisor Coauthors (log)	0.060***		0.015		0.096***		-0.007		-0.050*	
	[0.013]		[0.011]		[0.029]		[0.020]		[0.027]	
Full prof. (0/1)	0.047		0.042		-0.061		0.045		0.050	
	[0.031]		[0.026]		[0.055]		[0.053]		[0.057]	
Advisor Research Breadth	-0.576***		-0.470***		-0.489**		-0.119		0.011	
	[0.118]		[0.108]		[0.199]		[0.177]		[0.210]	
Constant	0.521***	0.233***	-0.643***	-0.960***	0.178	-0.464***	0.600***	1.101***	0.640***	1.200***
	[0.105]	[0.030]	[0.099]	[0.027]	[0.232]	[0.066]	[0.184]	[0.069]	[0.209]	[0.082]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y		Y		Y
N	16799	16792	16799	16792	16799	16792	16799	16792	12910	12875
adj. R^2	0.396	0.506	0.346	0.457	0.195	0.340	0.170	0.226	0.152	0.218

Notes: Advisor ID Engagement: Advisor publications mentioning the ID center in the five years before the student's PhD start
Fixed-effect regressions using entropy weights. Robust standard errors clustered at the advisor level in brackets. Dependent vars. in $\log((x + 0.2))$.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.2.2: Novelty during PhD training (*Advisor ID Engagement* instead of d_{ij})

	Binary				Average			
	New Words		New Phrases		New Words		New Phrases	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Advisor ID Engagement	0.037**	0.019	-0.005	-0.028*	0.044***	0.034**	0.006	-0.021
	[0.015]	[0.014]	[0.017]	[0.016]	[0.014]	[0.015]	[0.025]	[0.024]
Pre-doctoral Pubs (0/1)	0.037**	0.026	0.066***	0.026	-0.012	-0.006	0.029	-0.002
	[0.015]	[0.016]	[0.022]	[0.028]	[0.018]	[0.020]	[0.033]	[0.046]
Co-Advisor (0/1)	-0.019	-0.030*	0.015	-0.030	-0.035	-0.051*	0.034	-0.032
	[0.017]	[0.018]	[0.024]	[0.026]	[0.024]	[0.027]	[0.037]	[0.043]
Nr. Advisees (log)	-0.026		-0.014		-0.031		0.000	
	[0.016]		[0.019]		[0.021]		[0.032]	
Advisor Productivity (log)	0.002		-0.035**		0.010		-0.037	
	[0.013]		[0.015]		[0.017]		[0.023]	
Advisor Coauthors (log)	0.005		0.003		0.000		-0.008	
	[0.008]		[0.009]		[0.009]		[0.014]	
Full prof. (0/1)	0.026		0.028		0.018		0.023	
	[0.018]		[0.021]		[0.022]		[0.032]	
Advisor Research Breadth	-0.087		-0.103		-0.001		0.053	
	[0.067]		[0.088]		[0.079]		[0.133]	
Constant	0.084	0.093***	0.425***	0.313***	-1.480***	-1.437***	-0.944***	-1.023***
	[0.061]	[0.024]	[0.079]	[0.033]	[0.079]	[0.033]	[0.120]	[0.059]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y		Y
N	10213	10148	10213	10148	10213	10148	10213	10148
adj. R^2	0.100	0.124	0.124	0.165	0.069	0.078	0.120	0.144

Notes: Advisor ID Engagement: Advisor publications mentioning the ID center in the five years before the student's PhD start
Fixed effect regressions using entropy weights. Robust standard errors clustered at the advisor level.
Cols.(5)-(8) dep.vars.in logs$((x + 0.2))$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Alternative definition of d_{ij}

We re-estimate our main models of PhD students' productivity and novelty, redefining the advisor-engagement window. We consider two variants. First, instead of five years before and five years after PhD start, we remove any recency window prior to the student's graduation: any publication that lists the center before the student graduates qualifies the advisor as center-engaged, even if the most recent mention predates entry by more than five years ($d_{ij}(ever)$). Second, we use five years before and two years after (d_{ij} (up to 2 yrs. into the PhD)). This captures engagement leading into training while minimizing overlap with publications likely influenced by the student. In both variations, we find consistent evidence of our main results.

Table B.2.3: PhD Students' Productivity (alternative $d_{ij}(ever)$)

	Publications						Avg. Citations per Pub.			
	Simple count		Coauthor weighted		with Advisor		All pubs		with Advisor	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$d_{ij}(ever)$	0.107*	-0.274***	0.060	-0.166***	0.119	-0.027	0.031	0.041	-0.046	0.525**
	[0.057]	[0.048]	[0.043]	[0.044]	[0.101]	[0.119]	[0.063]	[0.189]	[0.074]	[0.210]
Pre-doctoral Pubs (0/1)	0.220***	0.100***	0.184***	0.078***	-0.023	-0.086	0.004	-0.032	-0.020	-0.098
	[0.031]	[0.026]	[0.027]	[0.023]	[0.065]	[0.058]	[0.054]	[0.062]	[0.061]	[0.069]
Co-Advisor (0/1)	0.067**	-0.013	-0.003	-0.058**	0.134**	0.131**	-0.038	0.014	-0.071	-0.031
	[0.033]	[0.030]	[0.027]	[0.028]	[0.063]	[0.066]	[0.056]	[0.062]	[0.064]	[0.067]
Nr. Advisees (lg)	0.003		-0.011		-0.035		0.017		0.057	
	[0.027]		[0.022]		[0.069]		[0.045]		[0.050]	
Advisor Productivity (log)	-0.092***		-0.066***		-0.151***		0.168***		0.220***	
	[0.025]		[0.020]		[0.051]		[0.041]		[0.058]	
Advisor Coauthors (log)	0.052***		0.008		0.106***		0.003		-0.026	
	[0.016]		[0.012]		[0.039]		[0.024]		[0.032]	
Full. Prof. (0/1)	-0.017		-0.031		-0.074		0.033		0.036	
	[0.039]		[0.030]		[0.072]		[0.068]		[0.073]	
Advisor Research Breadth	-0.562***		-0.433***		-0.430*		-0.092		-0.051	
	[0.132]		[0.122]		[0.246]		[0.220]		[0.265]	
Constant	0.525***	0.398***	-0.638***	-0.860***	-0.112	-0.521***	0.591***	1.160***	0.532**	1.045***
	[0.118]	[0.037]	[0.095]	[0.037]	[0.255]	[0.100]	[0.202]	[0.134]	[0.250]	[0.133]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y		Y		Y
N	15683	15674	15683	15674	15683	15674	15683	15674	12042	12007
adj. R^2	0.403	0.502	0.359	0.450	0.206	0.347	0.186	0.240	0.173	0.251

Notes: $d_{ij}(ever) = 1$ if the student's advisor became a member of an interdisciplinary center before graduation; 0 otherwise.
Fixed-effect regressions using entropy weights. Robust standard errors clustered at the advisor level in brackets. Dependent vars. in logs$((x + 0.2))$.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.2.4: Novelty during PhD training (alternative $d_{ij}(ever)$)

	Binary				Average			
	New Words		New Phrases		New Words		New Phrases	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$d_{ij}(ever)$	0.032	0.029	0.114***	0.075	-0.010	0.073	0.120***	0.386**
	[0.027]	[0.103]	[0.033]	[0.082]	[0.029]	[0.117]	[0.045]	[0.180]
Pre-doctoral Pubs (0/1)	0.029	-0.002	0.044*	0.018	-0.008	-0.024	0.006	-0.021
	[0.018]	[0.017]	[0.026]	[0.036]	[0.021]	[0.023]	[0.038]	[0.055]
Co-Advisor (0/1)	-0.004	-0.017	0.035	0.006	-0.008	-0.024	0.058	0.018
	[0.021]	[0.024]	[0.030]	[0.032]	[0.029]	[0.037]	[0.045]	[0.054]
Nr. Advisees (lg)	-0.028		-0.016		-0.040		-0.013	
	[0.019]		[0.026]		[0.025]		[0.042]	
Advisor Productivity	0.003		-0.053**		0.009		-0.065**	
	[0.018]		[0.021]		[0.023]		[0.032]	
Advisor Coauthors (log)	-0.003		-0.018		-0.001		-0.046**	
	[0.010]		[0.013]		[0.012]		[0.019]	
Full. Prof. (0/1)	0.020		0.017		0.016		0.023	
	[0.025]		[0.030]		[0.029]		[0.045]	
Advisor Research Breadth	-0.082		-0.037		-0.051		0.049	
	[0.096]		[0.114]		[0.108]		[0.176]	
Constant	0.092	0.080	0.435***	0.200***	-1.470***	-1.494***	-0.834***	-1.325***
	[0.082]	[0.057]	[0.103]	[0.055]	[0.106]	[0.070]	[0.160]	[0.108]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y		Y
N	9264	9195	9264	9195	9264	9195	9264	9195
adj. R^2	0.115	0.149	0.127	0.168	0.084	0.095	0.104	0.132

Notes: Fixed effect regressions using entropy weights. Robust standard errors clustered at the advisor level.
$d_{ij}(ever) = 1$ if the student's advisor became a member of an interdisciplinary center before graduation; 0 otherwise.
Cols.(5)-(8) dep.vars.in logs$((x + 0.2))$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.2.5: PhD Students' Productivity (alternative d_{ij}, up to 2 yrs. into the PhD)

	Publications						Avg. Citations per Pub.			
	Simple count		Coauthor weighted		with Advisor		All pubs		with Advisor	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
d_{ij} up to 2 yrs. into the PhD	0.090	-0.297***	0.013	-0.200***	0.114	-0.059	0.044	0.209*	-0.105	0.814***
	[0.066]	[0.066]	[0.048]	[0.057]	[0.111]	[0.138]	[0.069]	[0.121]	[0.096]	[0.196]
Pre-doctoral Pubs (0/1)	0.213***	0.099***	0.185***	0.090***	0.042	-0.043	0.024	-0.029	0.010	-0.111
	[0.031]	[0.029]	[0.028]	[0.026]	[0.067]	[0.062]	[0.064]	[0.070]	[0.078]	[0.089]
Co-Advisor (0/1)	0.060*	-0.040	-0.001	-0.071**	0.149**	0.115	-0.004	0.039	-0.062	-0.103
	[0.036]	[0.038]	[0.029]	[0.032]	[0.067]	[0.072]	[0.060]	[0.076]	[0.073]	[0.080]
Nr. Advisees (lg)	0.006		-0.000		-0.016		0.020		0.082	
	[0.031]		[0.024]		[0.082]		[0.053]		[0.064]	
Advisor Productivity	-0.183***		-0.135***		-0.254***		0.284***		0.356***	
	[0.037]		[0.034]		[0.085]		[0.063]		[0.088]	
Advisor Coauthors (log)	0.069***		0.022*		0.137***		-0.014		-0.049	
	[0.016]		[0.012]		[0.041]		[0.028]		[0.036]	
Full. Prof. (0/1)	-0.019		-0.023		-0.117		0.070		0.060	
	[0.046]		[0.035]		[0.082]		[0.077]		[0.089]	
Advisor Research Breadth	-0.736***		-0.551***		-0.654**		0.089		0.119	
	[0.155]		[0.150]		[0.306]		[0.262]		[0.357]	
Constant	0.933***	0.442***	-0.335**	-0.828***	0.286	-0.494***	0.119	1.122***	0.038	1.012***
	[0.174]	[0.044]	[0.164]	[0.043]	[0.392]	[0.095]	[0.311]	[0.127]	[0.380]	[0.134]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y		Y		Y
N	15128	15121	15128	15121	15128	15121	15128	15121	11631	11599
adj. R^2	0.424	0.506	0.361	0.438	0.237	0.362	0.198	0.239	0.186	0.257

Fixed-effect regressions using entropy weights. Robust standard errors clustered at the advisor level in brackets. Dependent vars. in logs$((x + 0.2))$.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.2.6: Novelty during PhD training (alternative d_{ij}, up to 2 yrs. into the PhD)

	Binary				Average			
	New Words		New Phrases		New Words		New Phrases	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
d_{ij}(up to 2 yrs. into the PhD)	0.038	0.121**	0.121***	0.143*	-0.006	0.165**	0.143**	0.547***
	[0.037]	[0.060]	[0.046]	[0.080]	[0.038]	[0.074]	[0.062]	[0.127]
Pre-doc. Pubs (0/1)	0.018	-0.009	0.079***	0.046	-0.030	-0.035	0.056	0.023
	[0.021]	[0.020]	[0.029]	[0.038]	[0.025]	[0.028]	[0.043]	[0.058]
Co-Advisor (0/1)	-0.010	-0.024	0.013	-0.009	-0.014	-0.021	0.030	-0.011
	[0.026]	[0.030]	[0.036]	[0.037]	[0.036]	[0.048]	[0.054]	[0.062]
Nr. Advisees (0/1)	-0.033		-0.030		-0.069**		-0.038	
	[0.023]		[0.032]		[0.029]		[0.051]	
Advisor Productivity	0.006		-0.057		0.003		-0.061	
	[0.031]		[0.035]		[0.041]		[0.049]	
Advisor Coauthors	0.006		-0.008		0.009		-0.036*	
	[0.010]		[0.014]		[0.013]		[0.019]	
Full prof (0/1)	0.027		0.039		0.028		0.057	
	[0.030]		[0.032]		[0.034]		[0.048]	
Advisor Research Breadth	-0.146		-0.109		-0.117		0.002	
	[0.113]		[0.130]		[0.125]		[0.203]	
Constant	0.089	0.048	0.481***	0.195***	-1.414***	-1.521***	-0.819***	-1.359***
	[0.138]	[0.040]	[0.165]	[0.052]	[0.177]	[0.060]	[0.241]	[0.085]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y		Y
N	15128	15121	15128	15121	15128	15121	15128	15121
adj. R^2	0.424	0.506	0.361	0.438	0.237	0.362	0.198	0.239

Notes: Fixed effect regressions using entropy weights. Robust standard errors clustered at the advisor level.
Cols.(5)-(8) dep.vars.in logs$((x + 0.2))$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Substantive ID Engagement

An additional concern with the definition we adopt for the interdisciplinary engagement of advisors is that it relies on their institutional affiliation with the center. This affiliation may reflect an administrative arrangement that serves more as a signaling device than as evidence of a genuine commitment to integrating distinct disciplines in research. To the extent that some researchers in our sample may be substantially interdisciplinary in their work without necessarily engaging with the centers, examining the outcomes of their students could provide further reassurance about the actual role of the centers (if any).

While imperfect, we test for this possibility as follows. We define a binary variable, *Substantive ID Engagement*, which takes the value of one for advisors whose research variety (i.e., the average number of distinct disciplines they reference) falls within the top 75th percentile of our sample. To fix ideas, out of the 334 categories that constitute our universe of disciplines, advisors reference an average of 5.4 disciplines per paper, whereas those in the top quartile reference around 7. We re-estimate our main specifications for students' research novelty and productivity, replacing the d_{ij} indicator with the *Substantive ID Engagement* indicator.

On one hand, Table B.2.7 reports the estimates for PhD students' research novelty, where we find little association between *Substantive ID Engagement* and any of the novelty outcomes. On the other hand, Table B.2.8 presents the results for PhD students' productivity, showing that the estimated coefficient for *Substantive ID Engagement* is negative and statistically significant, confirming the productivity penalty documented in our baseline results. Taken together, we interpret these findings as providing further support for our center-based definition of interdisciplinary engagement, which reflects a tangible commitment by researchers to connect and collaborate with faculty from other academic units to stimulate their research. Moreover, the absence of an observable "upside" in terms of either novelty or citation impact may indicate that the interdisciplinarity at the advisor level is not easily transmitted to students and that distinct organization settings,

such as the centers we study, enable such transmission.

Table B.2.7: Novelty during PhD training (*Substantive ID Engagement*)

	Binary				Average			
	New Words		New Phrases		New Words		New Phrases	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1.pr_top75_variety	-0.001	-0.014	-0.031*	-0.006	-0.001	-0.019	-0.037	-0.002
	[0.010]	[0.018]	[0.017]	[0.026]	[0.014]	[0.028]	[0.030]	[0.044]
Pre-doc. Pubs (0/1)	0.035***	0.025*	0.035	0.003	0.021	0.012	0.009	-0.037
	[0.012]	[0.014]	[0.022]	[0.023]	[0.017]	[0.020]	[0.036]	[0.039]
Co-Advisor (0/1)	-0.003	-0.000	-0.006	0.013	-0.003	-0.001	-0.005	0.021
	[0.010]	[0.012]	[0.017]	[0.020]	[0.015]	[0.018]	[0.029]	[0.032]
Nr. Advisees (0/1)	0.004		-0.012		0.008		-0.021	
	[0.009]		[0.017]		[0.013]		[0.028]	
Advisor Productivity	-0.016		-0.022		-0.017		-0.044	
	[0.010]		[0.016]		[0.016]		[0.027]	
Advisor Coauthors	-0.002		0.007		-0.006		0.010	
	[0.005]		[0.008]		[0.006]		[0.013]	
Full Prof. (0/1)	-0.001		0.014		-0.010		0.021	
	[0.012]		[0.020]		[0.016]		[0.032]	
Advisor Breadth	-0.014		-0.023		-0.016		-0.029	
	[0.011]		[0.018]		[0.015]		[0.030]	
Constant	0.039*	0.041**	0.213***	0.194***	-1.496***	-1.497***	-1.148***	-1.166***
	[0.023]	[0.019]	[0.033]	[0.027]	[0.032]	[0.026]	[0.057]	[0.044]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y		Y
N	10159	10093	10159	10093	10159	10093	10159	10093
adj. R^2	0.087	0.098	0.126	0.157	0.060	0.072	0.113	0.135

Notes: Fixed effect regressions using entropy weights. Robust standard errors clustered at the advisor level.
$d_{ij}(ever) = 1$ if the student's advisor became a member of an interdisciplinary center before graduation; 0 otherwise.
Cols.(5)-(8) dep.vars.in logs$((x + 0.2))$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.2.8: PhD Students' Productivity (*Substantive ID Engagement*)

	Publications						Avg. Citations per Pub.			
	Simple count		Coauthor weighted		with Advisor		All pubs		with Advisor	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1.pr_top75_variety	-0.041**	-0.051**	-0.038**	-0.039*	-0.006	-0.043	0.072*	-0.035	0.101**	-0.068
	[0.019]	[0.025]	[0.018]	[0.023]	[0.040]	[0.058]	[0.041]	[0.055]	[0.047]	[0.060]
Pre-doc. Pubs (0/1)	0.181***	0.131***	0.132***	0.096***	-0.210***	-0.213***	-0.030	-0.026	-0.021	-0.018
	[0.028]	[0.024]	[0.022]	[0.020]	[0.049]	[0.050]	[0.047]	[0.045]	[0.053]	[0.048]
Co-Advisor (0/1)	-0.035**	-0.018	-0.085***	-0.053***	0.014	0.084**	-0.041	-0.017	-0.064	-0.047
	[0.015]	[0.016]	[0.014]	[0.014]	[0.031]	[0.033]	[0.034]	[0.035]	[0.040]	[0.040]
Nr. Advisees (0/1)	0.015		0.017		-0.091***		0.014		0.043	
	[0.015]		[0.012]		[0.035]		[0.037]		[0.042]	
Advisor Productivity	0.019		0.022		0.038		0.226***		0.248***	
	[0.018]		[0.015]		[0.037]		[0.043]		[0.047]	
Advisor Coauthors	0.019**		-0.022***		0.050***		0.035**		0.029	>
	[0.008]		[0.007]		[0.018]		[0.016]		[0.018]	
Full Prof. (0/1)	0.001		-0.003		-0.050		0.054		0.031	
	[0.020]		[0.017]		[0.043]		[0.051]		[0.058]	
Advisor Breadth	-0.044**		-0.021		0.073**		-0.051		-0.059	
	[0.017]		[0.016]		[0.035]		[0.044]		[0.051]	
Constant	0.084***	0.133***	-0.881***	-0.958***	-0.424***	-0.361***	0.960***	1.167***	1.003***	1.242***
	[0.028]	[0.021]	[0.027]	[0.020]	[0.058]	[0.047]	[0.072]	[0.044]	[0.086]	[0.051]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y		Y		Y
N	16730	16723	16730	16723	16730	16723	16730	16723	13300	13270
adj. R^2	0.350	0.419	0.283	0.365	0.184	0.265	0.140	0.224	0.121	0.223

Fixed-effect regressions using entropy weights. Robust standard errors clustered at the advisor level in brackets. Dependent vars. in $\log((x+0.2))$.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Alternative estimation method (Poisson)

Table B.2.9: PhD Students' Research Productivity: Poisson

	Publications						Avg. Citations per Pub.	
	Simple count		Coauthor weighted		with Advisor		All pubs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
d_{ij}	0.157**	-0.249***	0.160**	-0.289***	0.122	-0.054	-0.019	-0.311
	[0.077]	[0.083]	[0.068]	[0.098]	[0.099]	[0.118]	[0.182]	[0.195]
Pre-doctoral Pubs (0/1)	0.246***	0.119***	0.295***	0.135***	0.064	-0.011	-0.114	-0.138*
	[0.040]	[0.034]	[0.045]	[0.042]	[0.059]	[0.047]	[0.084]	[0.080]
Co-Advisor (0/1)	0.037	-0.049	-0.055	-0.123**	0.078	0.093	-0.030	-0.183**
	[0.048]	[0.038]	[0.048]	[0.052]	[0.067]	[0.073]	[0.086]	[0.093]
Nr. Advisees (lg)	0.013		-0.017		0.017		-0.142*	
	[0.043]		[0.044]		[0.060]		[0.080]	
Advisor Productivity	-0.161***		-0.139***		-0.173***		0.488***	
	[0.034]		[0.035]		[0.044]		[0.141]	
Advisor Coauthors (log)	0.066***		0.018		0.112***		-0.112*	
	[0.023]		[0.018]		[0.031]		[0.061]	
Full. Prof. (0/1)	-0.008		-0.018		-0.093		-0.166	
	[0.049]		[0.055]		[0.064]		[0.130]	
Advisor Research Breadth	-0.523***		-0.550***		-0.304		0.999	
	[0.166]		[0.190]		[0.205]		[0.674]	
Constant	0.839***	0.641***	-0.756***	-1.046***	0.342*	0.063	0.355	2.812***
	[0.155]	[0.058]	[0.155]	[0.072]	[0.199]	[0.093]	[0.684]	[0.174]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y		Y
N	11539	11510	11539	11510	11527	11318	9042	8995
pseudo R^2	0.165	0.227	0.077	0.109	0.129	0.214	0.230	0.386

Notes: d_{ij}=1 if advisor demonstrates engagement in interdisciplinary research within a 5-year window around PhD entry.
Poisson FE regressions using entropy weights. Robust standard errors clustered at the advisor level in brackets.
*:p<0.10, **:p<0.05, ***:p<0.01

Subsample: Extended window of analysis

Table B.2.10: PhD Students' Research Novelty (Graduated by 2020, window: up to 5 years post graduation)

	Binary				Average			
	New Words		New Phrases		New Words		New Phrases	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Advisor ID Engagement	0.646***	0.514***	0.274	-0.421	0.531***	0.274	0.540*	-0.842
	[0.172]	[0.170]	[0.246]	[0.616]	[0.120]	[0.206]	[0.326]	[0.827]
Pre-doctoral Pubs (0/1)	0.008	0.009	0.037	0.031	-0.020	0.002	-0.016	0.024
	[0.024]	[0.022]	[0.034]	[0.049]	[0.027]	[0.028]	[0.054]	[0.079]
Co-Advisor (0/1)	-0.018	-0.001	0.022	0.051	-0.025	-0.006	0.049	0.096
	[0.028]	[0.026]	[0.040]	[0.035]	[0.034]	[0.036]	[0.055]	[0.064]
Nr. Advisees (log)	-0.064**		-0.077***		-0.067**		-0.118***	
	[0.026]		[0.027]		[0.028]		[0.045]	
Advisor Productivity Top25 (0/1)	0.027		0.003		0.033		-0.019	
	[0.025]		[0.032]		[0.033]		[0.049]	
Advisor Coauthors (log)	-0.008		-0.019		-0.007		-0.041*	
	[0.012]		[0.015]		[0.011]		[0.023]	
Full. Prof. (0/1)	0.056**		0.048**		0.042*		0.090***	
	[0.022]		[0.023]		[0.024]		[0.035]	
Advisor Research Breadth Top25 (0/1)	0.019		-0.016		0.030		-0.031	
	[0.023]		[0.035]		[0.027]		[0.051]	
Constant	0.114***	0.064	0.308***	0.199***	-1.438***	-1.510***	-1.064***	-1.274***
	[0.044]	[0.040]	[0.062]	[0.060]	[0.052]	[0.048]	[0.102]	[0.102]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y		Y
N	8606	8578	8606	8578	8606	8578	8606	8578
adj. R^2	0.115	0.153	0.129	0.191	0.089	0.111	0.114	0.154

Notes: Fixed effect regressions using entropy weights. Robust standard errors clustered at the advisor level.

Advisor ID Engagement: Advisor publications mentioning the ID center in the five years before the student's PhD start

Cols.(5)-(8) dep.vars.in logs$((x + 0.2))$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.2.11: PhD Students' Productivity (Graduated by 2020, window: up to 5 years post graduation)

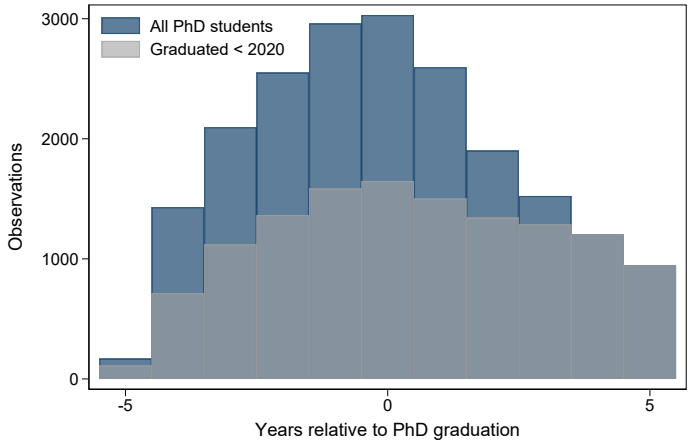
	Publications						Avg. Citations per Pub.			
	Simple count		Coauthor weighted		with Advisor		All pubs		with Advisor	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Advisor ID Engagement	0.751***	1.848***	0.367**	1.223***	0.284	0.835**	-0.129	-0.755	0.260	0.929
	[0.246]	[0.400]	[0.187]	[0.404]	[0.570]	[0.390]	[0.255]	[0.664]	[0.255]	[0.715]
Pre-doctoral Pubs (0/1)	0.212***	0.086**	0.179***	0.054	-0.031	-0.184**	0.023	0.118*	-0.030	0.085
	[0.049]	[0.038]	[0.044]	[0.044]	[0.098]	[0.089]	[0.068]	[0.066]	[0.073]	[0.086]
Co-Advisor (0/1)	0.045	-0.015	-0.007	-0.029	0.084	0.117	0.032	0.033	0.121*	0.156
	[0.054]	[0.054]	[0.041]	[0.051]	[0.107]	[0.087]	[0.067]	[0.078]	[0.063]	[0.098]
Nr. Advisees (log)	-0.124**		-0.055		-0.136		0.016		0.087	
	[0.049]		[0.042]		[0.089]		[0.070]		[0.066]	
Advisor Productivity (log)	-0.060		-0.039		0.007		0.098		0.157***	
	[0.049]		[0.034]		[0.090]		[0.069]		[0.061]	
Advisor Coauthors (log)	0.048*		0.017		-0.022		0.039		-0.012	
	[0.029]		[0.020]		[0.048]		[0.031]		[0.033]	
Full. Prof. (0/1)	0.080*		0.030		0.068		0.070		0.049	
	[0.041]		[0.032]		[0.077]		[0.063]		[0.066]	
Advisor Research Breadth	-0.475**		-0.390**		-0.586		-0.406		-0.169	
	[0.206]		[0.160]		[0.398]		[0.334]		[0.315]	
Constant	0.499**	0.266***	-0.688***	-0.938***	-0.306	-0.515***	0.701**	1.037***	0.414	0.967***
	[0.224]	[0.055]	[0.155]	[0.055]	[0.435]	[0.126]	[0.322]	[0.097]	[0.295]	[0.127]
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Science Field FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Academic Unit FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Advisor FE		Y		Y		Y		Y		Y
N	11592	11587	11592	11587	11592	11587	11592	11587	8961	8937
adj. R^2	0.422	0.515	0.410	0.477	0.250	0.352	0.182	0.242	0.188	0.226

Notes: Advisor ID Engagement: Advisor publications mentioning the ID center in the five years before the student's PhD start

Fixed-effect regressions using entropy weights. Robust standard errors clustered at the advisor level in brackets. Dependent vars. in $\log((x + 0.2))$.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure B.2.1: Histogram of PhD Students by Graduation Year



Subsample: PhD Students as Corresponding Author

Due to the limited time span of our data after students' graduation, a potential concern is that our results may primarily reflect the advisor's research trajectory rather than the students' own. At this stage, the majority of PhD students' publications are expected to be produced in collaboration with their advisors, consistent with the mentorship they receive during their doctoral studies. Descriptive statistics further indicate that, on average, up to 70% of students' yearly publications are coauthored with their advisor, and they appear as corresponding authors in only about 20% of publications. This share decreases after graduation, as shown in Figure B.2.2. Consequently, the citation premium reported in our baseline results (Table 3.8) may be driven by the advisor's prominence and networks rather than by the interdisciplinarity of the research itself.

To examine this issue more directly, we conduct a supplementary analysis restricting attention to publications where PhD students appear as corresponding authors—an indicator of intellectual leadership and greater independence in the research process. The dependent variables include: (i) a binary indicator for whether the student has any publication as corresponding author, (ii) the total number of such publications, and (iii) their average citation count. Results are reported in Table B.2.12.

The estimates suggest a nuanced relationship between advisor interdisciplinary engagement (d_{ij}) and students' independent research productivity. In specifications with advisor fixed effects, we uncover positive and statistically significant coefficients for both the likelihood of publishing as a corresponding author and the number of such publications. This suggests that, within the same advisor, greater interdisciplinary engagement is associated with more independence among their students. By contrast, the effects on citations remain negative, albeit weakly significant, indicating that when students lead these publications, they do not experience the citation premium observed in our main analysis (Table 3.8).

Taken together, these findings suggest that while advisors' interdisciplinary engagement may foster students' independent publication activity,

it does not translate into higher citation impact when students take the lead. This reinforces the interpretation that the citation premium in our main results likely reflects advisors' visibility and collaborative networks rather than students' autonomous research. It also highlights that early-stage researchers remain closely tied to their advisors' intellectual orbit, with true independence and potentially distinct citation advantages emerging only later in their academic trajectory. At the same time, the results suggest that students may gradually respond to their advisors' interdisciplinary engagement by pursuing more independent publications, either as a natural step in their development or as a strategic response to carve out their own research space—an insight that warrants further investigation.

Figure B.2.2: Students' Publications with their Advisor and as Corresponding Author

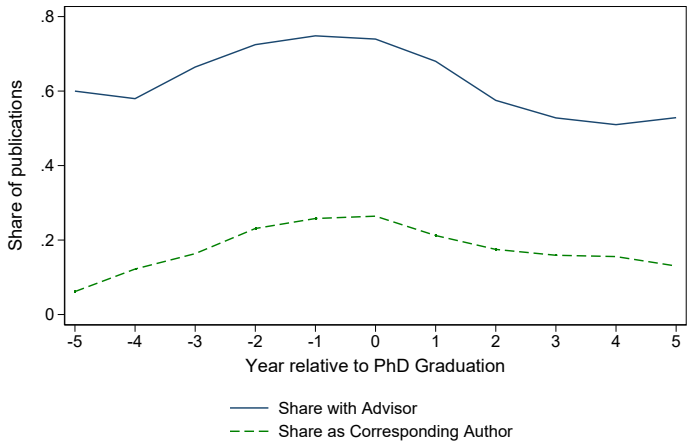


Table B.2.12: PhD Students' Productivity as Corresponding Author

	Publications as Corresponding Author					
	Any (0/1)		Count		Average Citations	
	(1)	(2)	(3)	(4)	(5)	(6)
d_{ij}	-0.061**	0.125**	-0.104*	0.227**	-0.142*	-0.749*
	[0.026]	[0.051]	[0.058]	[0.088]	[0.080]	[0.434]
Pre-doctoral Pubs (0/1)	0.059***	0.036*	0.120***	0.063	0.028	-0.045
	[0.021]	[0.021]	[0.044]	[0.044]	[0.062]	[0.100]
Co-Advisor (0/1)	-0.002	-0.005	0.009	0.001	0.053	-0.171
	[0.021]	[0.021]	[0.042]	[0.042]	[0.091]	[0.130]
Nr. Advisees (log)	-0.048**		-0.098**		0.046	
	[0.020]		[0.044]		[0.069]	
Advisor Productivty (log)	-0.038**		-0.086**		0.023	
	[0.017]		[0.038]		[0.099]	
Advisor Coauthors (log)	-0.004		-0.009		0.076**	
	[0.010]		[0.023]		[0.032]	
Full. Prof. (0/1)	-0.034		-0.058		0.087	
	[0.026]		[0.058]		[0.089]	
Advisor Research Breadth	-0.053		-0.180		-0.641**	
	[0.093]		[0.198]		[0.322]	
Constant	0.400***	0.124***	-0.778***	-1.372***	0.695	1.361***
	[0.078]	[0.035]	[0.179]	[0.066]	[0.428]	[0.220]
N	15748	15740	15748	15740	5061	4947
adj. R^2	0.184	0.287	0.218	0.338	0.259	0.365

Notes: d_{ij}=1 if advisor demonstrates engagement in interdisciplinary research within a 5-year window around PhD entry. Fixed-effect regressions using entropy weights. Robust standard errors clustered at the advisor level in brackets. Cols. (3)-(6): Dependent var. in logs$((x + 0.2))$. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.3 Interdisciplinary Indicators

Following Stirling et al.(2007) and Wang et al.(2015), interdisciplinarity can be characterized by three complementary components: *variety*, *balance*, and *disparity*. In terms of notation, a publication's references can be expressed in terms of n_c distinct ASJC (All Science Journal Classifications) disciplines. For each discipline i where $(i = 1, \ldots, n_c)$, x_i denotes the number of references to discipline i, and disciplines are sorted in *ascending* order of x_i. Moreover, the distance between any two distinct disciplines i and j can be denoted by d_{ij} $(i \neq j)$. With this notation in mind, we define the following indicators:

- Variety (V): Variety captures how many distinct disciplines are referenced:

$$V = n_c. \tag{1}$$

- Balance (B): Balance reflects how evenly references are distributed across disciplines. We compute the Gini coefficient G, and a balance index $B = 1 - G$, so that higher values indicate greater balance, indicating evenness in the distribution of disciplines. $B = 1$ denotes maximum evenness and $B = 0$ denotes extreme imbalance.

$$G = \frac{\sum_{i=1}^{n_c} (2i - n_c - 1)\, x_i}{n_c \sum_{k=1}^{n_c} x_k}, \qquad B = 1 - G \in [0, 1], \tag{2}$$

- Disparity (D): Disparity measures the average pairwise distance among the disciplines cited:

$$D = \frac{\sum_{i \neq j} d_{ij}}{n_c(n_c - 1)}. \tag{3}$$

  Following Zhang et al.(2016), distance is defined as one minus a cosine similarity derived from cross-citations. Using all Scopus-indexed papers published in 2020, we construct an ASJC cross-citation matrix and compute:

$$d_{ij} = 1 - \frac{R_{ij} + R_{ji}}{\sqrt{\big(TC_i + TR_i\big)\big(TC_j + TR_j\big)}}, \tag{4}$$

  where:

  R_{ij} is the number of references from set i (publications in discipline i) to set j. TC_k is the total citations received by set k *from other disciplines*. TR_k is the total references made by set k *to other disciplines*.

B.4 Academic Units by Research Orientation

Table B.4.1: Distribution by Research Orientation

Research Type	PhD Students	Fraction of the Sample	Mean d_{ij} (0/1)
Fundamental/Basic	1,764	48.1	0.03
Translational/Clinical	1,131	30.9	0.25
Hybrid	771	21.0	0.08

Table B.4.2: Classification of Academic Units by Research Orientation

Orientation	Academic Units (*names abbreviated*)
Translational / Clinical	Abdominal Surgical Oncology; Abdominal Transplantation; Nursing and Midwifery; Allergy and Clinical Immunology; Anesthesiology and Algology; Biomedical MRI; Cardiology and Cardiovascular Imaging; Clinical Psychology; Clinical and Experimental Endocrinology; Contextual and Clinical Psychiatry; Imaging & Pathology; Developmental Psychiatry; ICU and Transplantation; Locomotor and Neurological Disorders; Movement Control & Neuroplasticity; Neuropsychiatry; Nuclear Medicine & Molecular Imaging; Ophthalmology; Pediatric Oncology; Radiology and Radiopharmaceutical; Rehabilitation and Neurorehabilitation Sciences; Surgical, Thoracic and Vascular Surgery; Translational Cell & Tissue; Translational Breast Cancer ; Translational MRI; Translational in GastroIntestinal Disorders; Tumor Immunology and Immunotherapy; Virology and Gene Therapy; Woman and Child
Fundamental / Basic	Adaptive Immunology; Animal Physiology and Neurobiology; Astronomy and Astrophysics; Biochemistry, Molecular and Structural Biology; Bioengineering and Catalysis; Chemical and Process Engineering; Cognitive Neurology; Computational Biology; Computer Science and Informatics; Dynamics in Biological Systems; Ecology, Evolution and Biodiversity Conservation; Electrical, Electronic and Micro/Nano Systems; Food and Microbial Technology; Geography and Environmental Sciences; Human Genetics; Mathematics, Physics and Quantum Solid State Physics; Mechanics, Mechatronics, and Robotics; Molecular and Cellular Signaling; Molecular Biotechnology of Plants and Micro-organisms; Neurobiology and Neurophysiology; Lab for Protein Phosphorylation and Proteomics; Statistics, Numerical and Applied Mathematics; Surface Chemistry and Sustainable Catalysis; Theoretical Physics and Chemistry
Hybrid	Digestive and Molecular Digestive Oncology; Gynaecological Oncology; Hepatology; Development and Regeneration; Human Movement Biomechanics; Lab Angiogenesis and Vascular Metabolism; Lab Biosignaling & Therapeutics; Lab Cellular Metabolism and Regulation; Lab for Disease Mechanisms in Cancer; Soft Matter, Rheology and Technology; Skeletal Biology and Engineering; Other Surgical subunits;

C Appendix Chapter 4

C.1 Supplementary Data Description

Table C.1.1: Funding Success by Industry

	Funded (0/1)		Applicants
Industry	mean	(sd)	N
Agriculture, Horticulture & Livestock	0.06	(0.23)	546
Banking, Insurance & Financial Services	0.04	(0.19)	394
Biotechnology and Life Sciences	0.10	(0.29)	10979
Business Services	0.06	(0.24)	19662
Chemicals, Petroleum, Rubber & Plastic	0.08	(0.27)	2579
Communications	0.07	(0.25)	824
Computer Hardware	0.08	(0.27)	386
Computer Software	0.05	(0.22)	16411
Construction	0.07	(0.25)	1177
Food & Tobacco Manufacturing	0.07	(0.26)	626
Industrial, Electric & Electronic Machines	0.09	(0.29)	8073
Information Services	0.00	(0.00)	31
Leather, Stone, Clay & Glass Products	0.07	(0.25)	366
Media & Broadcasting	0.04	(0.19)	1343
Metals & Metal Products	0.08	(0.27)	1145
Mining & Extraction	0.06	(0.23)	87
Miscellaneous Manufacturing	0.08	(0.27)	471
Printing & Publishing	0.03	(0.18)	275
Property Services	0.03	(0.18)	357
Public Administration, Education, Health	0.04	(0.20)	1514
Retail	0.06	(0.23)	889
Textiles & Clothing Manufacturing	0.07	(0.25)	328
Transport Manufacturing	0.10	(0.30)	934
Transport, Freight & Storage	0.05	(0.21)	349
Travel, Personal & Leisure	0.05	(0.21)	989
Utilities	0.07	(0.26)	601
Waste Management & Treatment	0.06	(0.24)	473
Wholesale	0.06	(0.24)	3949
Wood, Furniture & Paper Manufacturing	0.08	(0.27)	420
Other	0.05	(0.22)	2234

Table C.1.2: Pairwise Correlations

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(13)	(14)	(15)	(16)
(1)	Rejected	1														
(2)	Firm age	-0.03*	1													
(3)	Employees	-0.05*	0.40*	1												
(4)	Revenue	-0.00	0.20*	0.19*	1											
(5)	Patent	-0.20*	0.14*	0.12*	0.04*	1										
(6)	Proposal WC	-0.12*	0.03*	0.02*	0.01*	0.07*	1									
(7)	Requested Grant	-0.19*	0.02*	0.09*	-0.02*	0.19*	0.09*	1								
(8)	Exceptional Grant	-0.04*	0.00	0.01*	-0.01*	0.07*	0.01*	0.43*	1							
(9)	Multiple Submissions	-0.20*	0.02*	0.02*	0.05*	0.15*	0.12*	0.12*	0.01*	1						
(10)	Applicants per Call	-0.01*	-0.10*	-0.04*	-0.02*	-0.02*	-0.01*	0.07*	-0.04*	0.03*	1					
(11)	Congestion Index	-0.02*	-0.00	0.01*	-0.01*	0.00	0.00	0.19*	0.02*	-0.02*	0.17*	1				
(13)	Alternative Framing	0.07*	-0.01*	-0.01*	-0.01*	-0.06*	-0.20*	-0.05*	-0.02*	-0.07*	0.01*	0.01*	1			
(14)	Innovativeness Framing	0.04	-0.02*	-0.01*	-0.05	0.01	-0.22*	-0.00	0.00	0.00	0.00	0.01	-0.27*	1		
(15)	Feasibility Framing	-0.02*	-0.00	0.03	0.01*	0.01*	-0.07*	0.00	0.02	0.01*	0.00	0.00	-0.31*	0.04*	1	
(16)	Contextually Novel	-0.06*	0.00	0.00	-0.00	0.03*	0.15*	0.03*	0.02*	0.05*	0.01	-0.01*	-0.03*	-0.03*	-0.02*	1

Notes: *:p<0.01.

C.2 Supplementary Analyses

Table C.2.1: Analysis of Rejection Probability (Placebo Framing)

DV = Rejected (0/1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(1+Firm age)	0.003	0.003	0.003	0.004	0.004	0.004	0.004	0.004
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
ln(1+Employees)	-0.017***	-0.017***	-0.016***	-0.016***	-0.016***	-0.016***	-0.016***	-0.016***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Revenue (0/1)	0.004	0.004	0.003	0.003	0.003	0.003	0.003	0.003
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Patent (0/1)	-0.107***	-0.107***	-0.107***	-0.110***	-0.110***	-0.110***	-0.110***	-0.110***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Multiple Submissions (0/1)	-0.157***	-0.156***	-0.156***	-0.155***	-0.155***	-0.155***	-0.155***	-0.155***
	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]
Proposal Word Count	-0.057***	-0.054***	-0.050***	-0.052***	-0.050***	-0.050***	-0.050***	-0.050***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
ln(Requested Grant)	-0.016	-0.017	-0.016	-0.017	-0.017	-0.017	-0.017	-0.005
	[0.011]	[0.011]	[0.011]	[0.012]	[0.012]	[0.012]	[0.012]	[0.014]
Exceptional Grant Request (0/1)	-0.041***	-0.041***	-0.040***	-0.039***	-0.039***	-0.039***	-0.039***	-0.039***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Congestion Index	0.077***	0.077***	0.077***	0.083***	0.083***	0.083***	0.083***	0.083***
	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]
ln(Applicants per Call)	-0.010	-0.009	-0.009	-0.010	-0.010	-0.010	-0.010	-0.010
	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]
Contextually Novel (0/1)		-0.028***	-0.027***	-0.025***	-0.025***	-0.027***	-0.025***	-0.024***
		[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
Alternative Framing (0/1)			-0.006***	-0.005***	-0.008**	-0.008***	-0.008**	-0.008**
			[0.001]	[0.002]	[0.003]	[0.003]	[0.003]	[0.003]
Importance Framing			0.001		0.005	0.004	0.005	0.005
			[0.003]		[0.004]	[0.004]	[0.004]	[0.004]
Team Capability Framing				-0.005**	-0.003	-0.003	-0.002	-0.003
				[0.002]	[0.003]	[0.003]	[0.003]	[0.003]
Contextually Novel X Importance Framing						0.006		
						[0.005]		
Contextually Novel X Team Capability Framing							-0.004	
							[0.004]	
Contextually Novel X Congestion Index								-0.054***
								[0.020]
Constant	1.360***	1.350***	1.345***	1.327***	1.325***	1.325***	1.326***	1.323***
	[0.047]	[0.046]	[0.047]	[0.048]	[0.048]	[0.048]	[0.048]	[0.048]
Call FE	y	y	y	y	y	y	y	y
Country FE	y	y	y	y	y	y	y	y
Industry FE	y	y	y	y	y	y	y	y
N	77916	77916	77916	77916	77916	77916	77916	77916
adj. R-sq	0.137	0.137	0.138	0.139	0.139	0.139	0.139	0.139

Notes: Logistic regression estimates. Robust standard errors clustered at the application cutoff level in brackets. *p<0.10, **p<0.05, ***p<0.01.

Table C.2.2: Analysis of Rejection Probability, Split by Evaluator Congestion (Placebo Framing)

	Low Congestion			High Congestion		
DV=Rejected(0/1)	(1)	(2)	(3)	(4)	(5)	(6)
ln(1+Firm age)	0.001	0.001	0.001	0.006	0.006	0.006
	[0.004]	[0.004]	[0.004]	[0.006]	[0.006]	[0.006]
ln(1+Employees)	-0.014***	-0.014***	-0.014***	-0.018***	-0.018***	-0.018***
	[0.003]	[0.003]	[0.003]	[0.004]	[0.004]	[0.004]
Revenue (0/1)	-0.036***	-0.036***	-0.036***	-0.019**	-0.019**	-0.019**
	[0.008]	[0.008]	[0.008]	[0.009]	[0.009]	[0.009]
Patent (0/1)	-0.095***	-0.095***	-0.095***	-0.085***	-0.085***	-0.085***
	[0.007]	[0.007]	[0.007]	[0.008]	[0.008]	[0.008]
Multiple Submissions (0/1)	-0.147***	-0.147***	-0.147***	-0.160***	-0.160***	-0.160***
	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
Proposal Word Count	-0.048***	-0.048***	-0.048***	-0.055***	-0.055***	-0.056***
	[0.008]	[0.008]	[0.008]	[0.007]	[0.007]	[0.007]
ln(Requested Grant)	-0.020	-0.019	-0.020	0.019	0.019	0.019
	[0.044]	[0.044]	[0.044]	[0.025]	[0.025]	[0.025]
Exceptional Grant Request (0/1)	-0.039***	-0.038***	-0.039***	-0.030***	-0.030***	-0.030***
	[0.007]	[0.007]	[0.007]	[0.004]	[0.004]	[0.004]
Congestion Index	0.056	0.055	0.055	0.096***	0.096***	0.097***
	[0.038]	[0.038]	[0.039]	[0.022]	[0.022]	[0.022]
ln(Applicants per Call)	-0.005	-0.004	-0.005	-0.001	-0.001	0.000
	[0.024]	[0.024]	[0.024]	[0.014]	[0.014]	[0.014]
Contextually Novel (0/1)	-0.002	-0.003	-0.002	-0.009**	-0.009**	-0.008**
	[0.005]	[0.006]	[0.005]	[0.004]	[0.004]	[0.004]
Alternative Framing (0/1)	-0.010	-0.018**	-0.010	-0.030***	-0.034***	-0.029***
	[0.009]	[0.008]	[0.009]	[0.010]	[0.011]	[0.010]
Importance Framing	0.007	0.002	0.007	0.006	0.004	0.006
	[0.007]	[0.007]	[0.007]	[0.005]	[0.005]	[0.005]
Team Capability Framing	-0.011**	-0.011*	-0.010*	0.001	0.001	0.002
	[0.006]	[0.006]	[0.005]	[0.003]	[0.003]	[0.004]
Contextually Novel × Importance Framing		0.021*			0.010	
		[0.011]			[0.008]	
Contextually Novel × Team Capability Framing			-0.006			-0.006
			[0.009]			[0.007]
Constant	1.251***	1.256***	1.254***	1.282***	1.284***	1.284***
	[0.087]	[0.086]	[0.088]	[0.079]	[0.077]	[0.079]
Call FE	y	y	y	y	y	y
Country FE	y	y	y	y	y	y
Industry FE	y	y	y	y	y	y
N	17506	17506	17506	17443	17443	17443
adj. R^2	0.124	0.124	0.124	0.168	0.168	0.168

Notes: Logistic regression estimates. Robust standard errors clustered at the application cutoff level in brackets.
Low (High) Congestion is defined as *Congestion Index* values below (above) the 25 (75) percentile, equivalent to -0.08 (0.14).
H4 refers to Hypothesis 4. *p<0.10, **p<0.05, ***p<0.01.

C.3 Further Robustness Checks

- First, in Table C.3.1 and Table C.3.2 we repeat our estimations using logistic regression instead of a linear probability model and find consistent support for our arguments in both the full sample regression, as well as for the split sample by Congestion Index.
- Second, in Table C.3.3 and Table C.3.4 we repeat our analysis using an alternative operationalization of our framing variables. We use the "raw" share of words in each framing category relative to total proposal length (*Innovativeness Share* and *Feasibility Share*, respectively), rather than the normalized by application deadline and industry used in the main analysis. The findings remain substantively similar to our main results.
- Third, in Table C.3.5 and Table C.3.6 we repeat our analyses focusing on conventional ideas instead (*Contextually Typical*), showing that conventional ideas are more likely to be rejected, and that these ideas are penalized when adopting feasibility framing, which is the exactly opposite pattern of results reported in our main analysis.
- Finally, following Hainmueller (2011), we use entropy balancing to rule out the concern that our framing effects merely reflect systematic differences between funded and rejected ventures (differences not fully captured by the linear controls in our main regressions), while avoiding additional functional form assumptions. All control variables entered the balancing algorithm except our focal predictors: *Innovativeness* and *Feasibility Framing*, *Contextually Novel*, and *Congestion Index*. Table C.3.9 shows that the procedure achieves balance even considering the first three moments (mean, variance, skewness) across the two groups. Using this weighted sample, we find consistent results concerning our framing variables in the balanced full sample. By contrast, our results are less consistent for the split-sample analysis: under high congestion, the interaction between *Feasibility Framing* and *Contextually Novel* remains negative but not statistically significant.

Table C.3.1: Analysis of Rejection Probability (Logit)

DV = Rejected (0/1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(1+Firm age)	0.011	0.011	0.010	0.010	0.010	0.009	0.010	0.010
	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]
ln(1+Employees)	-0.085***	-0.085***	-0.085***	-0.085***	-0.084***	-0.084***	-0.084***	-0.084***
	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]
Revenue (0/1)	0.007	0.005	0.005	0.005	0.006	0.006	0.006	0.006
	[0.023]	[0.023]	[0.023]	[0.023]	[0.023]	[0.023]	[0.023]	[0.023]
Patent (0/1)	-0.537***	-0.538***	-0.533***	-0.534***	-0.534***	-0.534***	-0.534***	-0.534***
	[0.022]	[0.022]	[0.022]	[0.022]	[0.022]	[0.022]	[0.022]	[0.022]
Multiple Submissions (0/1)	-1.291***	-1.286***	-1.278***	-1.279***	-1.278***	-1.278***	-1.278***	-1.277***
	[0.088]	[0.088]	[0.087]	[0.087]	[0.087]	[0.087]	[0.087]	[0.086]
Proposal Word Count	-0.399***	-0.382***	-0.374***	-0.369***	-0.385***	-0.384***	-0.385***	-0.385***
	[0.023]	[0.022]	[0.023]	[0.022]	[0.023]	[0.023]	[0.023]	[0.023]
ln(Requested Grant)	-0.234***	-0.235***	-0.234***	-0.234***	-0.234***	-0.234***	-0.234***	-0.233***
	[0.024]	[0.024]	[0.024]	[0.024]	[0.024]	[0.024]	[0.024]	[0.024]
Exceptional Grant Request (0/1)	0.298***	0.303***	0.307***	0.305***	0.306***	0.306***	0.306***	0.309***
	[0.072]	[0.073]	[0.072]	[0.072]	[0.072]	[0.072]	[0.072]	[0.073]
Congestion Index	0.047	0.045	0.043	0.044	0.046	0.046	0.046	0.113
	[0.070]	[0.071]	[0.070]	[0.071]	[0.071]	[0.071]	[0.071]	[0.085]
ln(Applicants per Call)	-0.083	-0.078	-0.077	-0.078	-0.077	-0.077	-0.077	-0.077
	[0.087]	[0.087]	[0.087]	[0.087]	[0.087]	[0.087]	[0.087]	[0.087]
Contextually Novel (0/1)		-0.145***	-0.144***	-0.145***	-0.145***	-0.139***	-0.136***	-0.137***
		[0.028]	[0.028]	[0.028]	[0.028]	[0.029]	[0.029]	[0.026]
Alternative Framing (0/1)			0.160***	0.154***	0.133***	0.133***	0.132***	0.133***
			[0.024]	[0.026]	[0.023]	[0.023]	[0.023]	[0.023]
Innovativeness Framing			-0.035**		-0.040***	-0.036**	-0.040***	-0.039***
			[0.014]		[0.014]	[0.015]	[0.014]	[0.014]
Feasibility Framing (H1)				-0.043***	-0.046***	-0.046***	-0.041***	-0.046***
				[0.011]	[0.011]	[0.011]	[0.011]	[0.011]
Contextually Novel × Innovativeness Framing							-0.017	
							[0.030]	
Contextually Novel × Feasibility Framing (H2)								-0.023
								[0.029]
Contextually Novel × Congestion Index (H3)								-0.262**
								[0.130]
Constant	5.734***	5.730***	5.665***	5.654***	5.719***	5.714***	5.714***	5.724***
	[1.034]	[1.030]	[1.010]	[1.010]	[1.013]	[1.011]	[1.013]	[1.006]
Call FE	y	y	y	y	y	y	y	y
Country FE	y	y	y	y	y	y	y	y
Industry FE	y	y	y	y	y	y	y	y
N	77717	77717	77717	77717	77717	77717	77717	77717
pseudo R^2	0.124	0.125	0.125	0.125	0.125	0.125	0.125	0.126

Notes: Logistic regression estimates. Robust standard errors clustered at the application cutoff level in brackets. *p<0.10, **p<0.05, ***p<0.01.
H1, *H2*, and *H3* refer to Hypothesis 1, Hypothesis 2, and Hypothesis 3, respectively.

Table C.3.2: Analysis of Rejection Probability, Split by Evaluator Congestion (Logit)

	Low Congestion			High Congestion		
DV=Rejected(0/1)	(1)	(2)	(3)	(4)	(5)	(6)
ln(1+Firm age)	-0.003	-0.003	-0.003	0.038	0.038	0.037
	[0.025]	[0.025]	[0.025]	[0.033]	[0.033]	[0.033]
ln(1+Employees)	-0.071***	-0.071***	-0.071***	-0.104***	-0.105***	-0.105***
	[0.016]	[0.016]	[0.016]	[0.020]	[0.020]	[0.020]
Revenue (0/1)	-0.032	-0.032	-0.033	0.011	0.010	0.011
	[0.042]	[0.042]	[0.042]	[0.044]	[0.044]	[0.044]
Patent (0/1)	-0.537***	-0.537***	-0.537***	-0.447***	-0.447***	-0.448***
	[0.045]	[0.045]	[0.045]	[0.051]	[0.051]	[0.051]
Multiple Submissions (0/1)	-1.170***	-1.170***	-1.170***	-1.517***	-1.516***	-1.517***
	[0.100]	[0.100]	[0.100]	[0.155]	[0.155]	[0.155]
Proposal Word Count	-0.391***	-0.390***	-0.391***	-0.480***	-0.484***	-0.478***
	[0.042]	[0.043]	[0.042]	[0.039]	[0.039]	[0.039]
ln(Requested Grant)	-0.307***	-0.307***	-0.307***	-0.179***	-0.178***	-0.179***
	[0.028]	[0.028]	[0.028]	[0.042]	[0.042]	[0.042]
Exceptional Grant Request (0/1)	0.318*	0.320*	0.318*	0.297**	0.297**	0.297**
	[0.167]	[0.167]	[0.168]	[0.138]	[0.138]	[0.138]
Congestion Index	-0.197	-0.198	-0.197	0.280	0.278	0.281
	[0.199]	[0.199]	[0.199]	[0.185]	[0.184]	[0.186]
ln(Applicants per Call)	0.042	0.042	0.042	-0.168	-0.169	-0.168
	[0.137]	[0.137]	[0.137]	[0.111]	[0.111]	[0.111]
Contextually Novel (0/1)	-0.071	-0.061	-0.072	-0.172***	-0.196***	-0.136**
	[0.051]	[0.047]	[0.054]	[0.056]	[0.061]	[0.057]
Alternative Framing (0/1)	0.155***	0.155***	0.155***	0.080	0.080	0.079
	[0.050]	[0.050]	[0.050]	[0.052]	[0.052]	[0.052]
Innovativeness Framing	-0.089***	-0.082***	-0.089***	-0.070***	-0.087***	-0.071***
	[0.029]	[0.031]	[0.029]	[0.020]	[0.025]	[0.020]
Feasibility Framing	-0.040**	-0.040**	-0.041	-0.057***	-0.057***	-0.037*
	[0.020]	[0.020]	[0.025]	[0.018]	[0.018]	[0.020]
Contextually Novel × Innovativeness Framing		-0.030			0.073	
		[0.066]			[0.056]	
Contextually Novel × Feasibility Framing *(H4)*			0.004			-0.091**
			[0.058]			[0.044]
Constant	5.038***	5.032***	5.039***	4.830***	4.848***	4.812***
	[0.907]	[0.907]	[0.906]	[0.491]	[0.491]	[0.486]
Call FE	y	y	y	y	y	y
Country FE	y	y	y	y	y	y
Industry FE	y	y	y	y	y	y
N	19384	19384	19384	19436	19436	19436
pseudo R^2	0.114	0.114	0.114	0.153	0.153	0.153

Notes: Logistic regression estimates. Robust standard errors clustered at the application cutoff level in brackets.
Low (High) Congestion is defined as *Congestion Index* values below (above) the 25 (75) percentile, equivalent to -0.08 (0.14).
H4 refers to Hypothesis 4. *p<0.10, **p<0.05, ***p<0.01.

Table C.3.3: Analysis of Rejection Probability (Alt. Framing Operationalization)

DV = Rejected (0/1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(1+Firm age)	0	0.001	0	0	0	0	0	0
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
ln(1+Employees)	-0.015***	-0.014***	-0.014***	-0.014***	-0.014***	-0.014***	-0.014***	-0.014***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Revenue (0/1)	-0.020***	-0.020***	-0.021***	-0.020***	-0.020***	-0.020***	-0.020***	-0.020***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Patent (0/1)	-0.097***	-0.097***	-0.096***	-0.096***	-0.096***	-0.096***	-0.096***	-0.096***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Multiple Submissions (0/1)	-0.153***	-0.152***	-0.151***	-0.151***	-0.150***	-0.150***	-0.150***	-0.150***
	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Proposal Word Count	-0.055***	-0.052***	-0.051***	-0.049***	-0.053***	-0.053***	-0.053***	-0.053***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
ln(Requested Grant)	-0.047***	-0.047***	-0.047***	-0.047***	-0.047***	-0.047***	-0.047***	-0.047***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Exceptional Grant Request (0/1)	0.061***	0.061***	0.062***	0.061***	0.061***	0.061***	0.061***	0.062***
	[0.016]	[0.016]	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]
Congestion Index	-0.005	-0.006	-0.007	-0.007	-0.007	-0.006	-0.007	0.004
	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.014]
ln(Applicants per Call)	-0.011	-0.011	-0.011	-0.01	-0.011	-0.011	-0.011	-0.011
	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]
Contextually Novel (0/1)		-0.026***	-0.026***	-0.026***	-0.026***	-0.019***	-0.024***	-0.025***
		[0.005]	[0.005]	[0.005]	[0.005]	[0.006]	[0.005]	[0.005]
Alternative Framing (0/1)			0.021***	0.023***	0.011**	0.011**	0.011**	0.011***
			[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Innovativeness Share			-0.016***		-0.021***	-0.018***	-0.021***	-0.021***
			[0.004]		[0.005]	[0.005]	[0.005]	[0.005]
Feasibility Share				-0.019***	-0.025***	-0.025***	-0.023***	-0.025***
				[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Contextually Novel X Innovativeness Share (H1)						-0.020*		
						[0.010]		
Contextually Novel X Feasibility Share (H1)							-0.011	
							[0.013]	
Contextually Novel X Congestion Index (H2)								-0.048**
								[0.021]
Constant	1.521***	1.520***	1.517***	1.510***	1.531***	1.527***	1.530***	1.533***
	[0.122]	[0.120]	[0.118]	[0.119]	[0.120]	[0.119]	[0.120]	[0.119]
Call FE	y	y	y	y	y	y	y	y
Country FE	y	y	y	y	y	y	y	y
Industry FE	y	y	y	y	y	y	y	y
N	77717	77717	77717	77717	77717	77717	77717	77717
adj. R-sq	0.143	0.144	0.144	0.144	0.144	0.144	0.144	0.145

Notes: Linear regression estimates. Robust standard errors clustered at the application cutoff level in brackets. *p<0.10, **p<0.05, ***p<0.01.
H1, *H2*, and *H3* refer to Hypothesis 1, Hypothesis 2, and Hypothesis 3, respectively.

Table C.3.4: Analysis of Rejection Probability, Split by Evaluator Congestion (Alt. Operationalization of Framing)

	Low Congestion			High Congestion		
DV=Rejected(0/1)	(1)	(2)	(3)	(4)	(5)	(6)
ln(1+Firm age)	-0.003	-0.003	-0.003	0.004	0.004	0.004
	[0.004]	[0.004]	[0.004]	[0.005]	[0.005]	[0.005]
ln(1+Employees)	-0.011***	-0.011***	-0.011***	-0.017***	-0.017***	-0.017***
	[0.003]	[0.003]	[0.003]	[0.004]	[0.004]	[0.004]
Revenue (0/1)	-0.031***	-0.031***	-0.032***	-0.020**	-0.020**	-0.019**
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Patent (0/1)	-0.093***	-0.093***	-0.093***	-0.081***	-0.081***	-0.082***
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Multiple Submissions (0/1)	-0.143***	-0.143***	-0.143***	-0.163***	-0.163***	-0.163***
	[0.013]	[0.013]	[0.013]	[0.011]	[0.011]	[0.011]
Proposal Word Count	-0.054***	-0.054***	-0.054***	-0.062***	-0.062***	-0.062***
	[0.006]	[0.006]	[0.006]	[0.007]	[0.007]	[0.007]
ln(Requested Grant)	-0.056***	-0.056***	-0.056***	-0.032***	-0.032***	-0.032***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Exceptional Grant Request (0/1)	0.060*	0.060*	0.060*	0.062**	0.062**	0.062**
	[0.033]	[0.033]	[0.033]	[0.025]	[0.025]	[0.025]
Congestion Index	-0.03	-0.03	-0.029	0.028	0.027	0.028
	[0.035]	[0.035]	[0.035]	[0.027]	[0.027]	[0.027]
ln(Applicants per Call)	0.012	0.012	0.011	-0.021	-0.021	-0.021
	[0.022]	[0.022]	[0.022]	[0.019]	[0.019]	[0.019]
Contextually Novel (0/1)	-0.014	-0.001	-0.016*	-0.033***	-0.040***	-0.025**
	[0.008]	[0.008]	[0.008]	[0.011]	[0.014]	[0.011]
Alternative Framing (0/1)	0.009	0.009	0.009	-0.003	-0.003	-0.003
	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]
Innovativeness Share	-0.033***	-0.026**	-0.033***	-0.031***	-0.035***	-0.031***
	[0.009]	[0.010]	[0.009]	[0.005]	[0.007]	[0.005]
Feasibility Share	-0.028***	-0.028***	-0.030***	-0.028***	-0.028***	-0.020**
	[0.009]	[0.009]	[0.010]	[0.008]	[0.008]	[0.010]
Contextually Novel X Innovativeness Share		-0.035*			0.021	
		[0.020]			[0.021]	
Contextually Novel X Feasibility Share (H4)			0.012			-0.043*
			[0.025]			[0.024]
Constant	1.570***	1.563***	1.571***	1.327***	1.332***	1.322***
	[0.095]	[0.095]	[0.095]	[0.069]	[0.069]	[0.068]
Call FE	y	y	y	y	y	y
Country FE	y	y	y	y	y	y
Industry FE	y	y	y	y	y	y
N	19384	19384	19384	19529	19529	19529
adj. R-sq	0.123	0.123	0.123	0.165	0.165	0.165

Notes: Linear regression estimates. Robust standard errors clustered at the application cutoff level in brackets.
Low (High) Congestion is defined as *Congestion Index* values below (above) the 25 (75) percentile, equivalent to -0.08 (0.14).
H4 refers to Hypothesis 4. *p<0.10, **p<0.05, ***p<0.01.

Table C.3.5: Analysis of Rejection Probability (Contextually Typical)

DV = Rejected (0/1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(1+Firm age)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
ln(1+Employees)	-0.015***	-0.014***	-0.014***	-0.014***	-0.014***	-0.014***	-0.014***	-0.014***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Revenue (0/1)	-0.020***	-0.020***	-0.020***	-0.020***	-0.020***	-0.020***	-0.020***	-0.020***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Patent (0/1)	-0.097***	-0.096***	-0.095***	-0.096***	-0.095***	-0.095***	-0.095***	-0.095***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Multiple Submissions (0/1)	-0.153***	-0.150***	-0.149***	-0.149***	-0.149***	-0.149***	-0.149***	-0.149***
	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Proposal Word Count	-0.055***	-0.045***	-0.043***	-0.042***	-0.045***	-0.045***	-0.045***	-0.045***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
ln(Requested Grant)	-0.047***	-0.048***	-0.047***	-0.047***	-0.047***	-0.047***	-0.047***	-0.047***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Exceptional Grant Request (0/1)	0.061***	0.061***	0.061***	0.061***	0.061***	0.061***	0.061***	0.061***
	[0.016]	[0.016]	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]	[0.015]
Congestion Index	-0.005	-0.006	-0.007	-0.007	-0.007	-0.007	-0.007	-0.009
	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]
ln(Applicants per Call)	-0.011	-0.010	-0.010	-0.010	-0.010	-0.010	-0.010	-0.010
	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]
Contextually Typical (0/1)		0.042***	0.041***	0.041***	0.041***	0.036***	0.039***	0.041***
		[0.004]	[0.004]	[0.004]	[0.004]	[0.005]	[0.004]	[0.004]
Alternative Framing (0/1)			0.020***	0.023***	0.010**	0.010**	0.010**	0.010**
			[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Innovativeness Framing			-0.016***		-0.021***	-0.028***	-0.021***	-0.021***
			[0.004]		[0.005]	[0.007]	[0.005]	[0.005]
Feasibility Framing				-0.019***	-0.025***	-0.025***	-0.028***	-0.025***
				[0.005]	[0.005]	[0.005]	[0.007]	[0.005]
Contextually Typical × Innovativeness Framing							0.016*	
							[0.009]	
Contextually Typical × Feasibility Framing								0.010
								[0.011]
Contextually Typical × Congestion Index								0.007
								[0.014]
Constant	1.521***	1.473***	1.470***	1.463***	1.484***	1.485***	1.485***	1.484***
	[0.122]	[0.125]	[0.123]	[0.124]	[0.124]	[0.125]	[0.124]	[0.124]
Call FE	y	y	y	y	y	y	y	y
Country FE	y	y	y	y	y	y	y	y
Industry FE	y	y	y	y	y	y	y	y
N	77717	77717	77717	77717	77717	77717	77717	77717
adj. R^2	0.138	0.140	0.140	0.140	0.141	0.141	0.141	0.141

Notes: Linear regression estimates. Robust standard errors clustered at the application cutoff level in brackets. *p<0.10, **p<0.05, ***p<0.01.

Table C.3.6: Analysis of Rejection Probability, Split by Evaluator Congestion (Contextually Typical)

	Low Congestion			High Congestion		
DV=Rejected(0/1)	(1)	(2)	(3)	(4)	(5)	(6)
ln(1+Firm age)	-0.003	-0.003	-0.003	0.004	0.004	0.004
	[0.004]	[0.004]	[0.004]	[0.005]	[0.005]	[0.005]
ln(1+Employees)	-0.011***	-0.011***	-0.011***	-0.017***	-0.017***	-0.017***
	[0.003]	[0.003]	[0.003]	[0.004]	[0.004]	[0.004]
Revenue (0/1)	-0.031***	-0.031***	-0.031***	-0.020**	-0.020**	-0.019**
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Patent (0/1)	-0.093***	-0.093***	-0.093***	-0.081***	-0.081***	-0.081***
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Multiple Submissions (0/1)	-0.141***	-0.142***	-0.141***	-0.162***	-0.162***	-0.162***
	[0.013]	[0.013]	[0.013]	[0.012]	[0.012]	[0.012]
Proposal Word Count	-0.045***	-0.045***	-0.045***	-0.055***	-0.055***	-0.055***
	[0.006]	[0.006]	[0.006]	[0.007]	[0.007]	[0.007]
ln(Requested Grant)	-0.057***	-0.057***	-0.057***	-0.032***	-0.032***	-0.032***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Exceptional Grant Request (0/1)	0.060*	0.061*	0.060*	0.060**	0.060**	0.061**
	[0.033]	[0.033]	[0.033]	[0.024]	[0.024]	[0.025]
Congestion Index	-0.029	-0.029	-0.029	0.031	0.031	0.031
	[0.034]	[0.034]	[0.034]	[0.028]	[0.028]	[0.028]
ln(Applicants per Call)	0.013	0.013	0.013	-0.020	-0.020	-0.020
	[0.023]	[0.023]	[0.023]	[0.019]	[0.019]	[0.019]
Contextually Typical (0/1)	0.039***	0.036***	0.040***	0.041***	0.041***	0.031***
	[0.007]	[0.007]	[0.008]	[0.007]	[0.009]	[0.009]
Alternative Framing (0/1)	0.021**	0.021**	0.021**	0.007	0.007	0.007
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Innovativeness Framing	-0.013***	-0.017***	-0.013***	-0.011***	-0.011*	-0.011***
	[0.005]	[0.006]	[0.005]	[0.003]	[0.006]	[0.003]
Feasibility Framing	-0.005	-0.005	-0.005	-0.009***	-0.009***	-0.017***
	[0.004]	[0.004]	[0.005]	[0.003]	[0.003]	[0.004]
Contextually Typical × Innovativeness Framing		0.010			-0.001	
		[0.008]			[0.010]	
Contextually Typical × Feasibility Framing			0.000			0.023***
			[0.007]			[0.007]
Constant	1.509***	1.501***	1.509***	1.246***	1.246***	1.251***
	[0.098]	[0.099]	[0.096]	[0.074]	[0.075]	[0.072]
Call FE	y	y	y	y	y	y
Country FE	y	y	y	y	y	y
Industry FE	y	y	y	y	y	y
N	20046	20046	20046	19529	19529	19529
adj. R^2	0.124	0.124	0.124	0.165	0.165	0.165

Notes: Linear regression estimates. Robust standard errors clustered at the application cutoff level in brackets.
Low (High) Congestion is defined as *Congestion Index* values below (above) the 25 (75) percentile, equivalent to -0.08 (0.14).
Contextually Typical: Percentile 25 or below of atypicality. *p<0.10, **p<0.05, ***p<0.01.

Table C.3.7: Analysis of Rejection Probability, (Rejection Rates)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(1+Firm age)	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
ln(1+Employees)	-0.015***	-0.015***	-0.015***	-0.015***	-0.015***	-0.015***	-0.015***	-0.015***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Revenue (0/1)	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***	-0.019***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Patent (0/1)	-0.097***	-0.097***	-0.096***	-0.096***	-0.096***	-0.096***	-0.096***	-0.096***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Multiple Submissions (0/1)	-0.152***	-0.151***	-0.150***	-0.150***	-0.150***	-0.150***	-0.150***	-0.150***
	[0.011]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Proposal Word Count	-0.056***	-0.053***	-0.050***	-0.050***	-0.052***	-0.052***	-0.052***	-0.052***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
ln(Requested Grant)	-0.020**	-0.020**	-0.019**	-0.019**	-0.019**	-0.019**	-0.019**	-0.019**
	[0.008]	[0.008]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.008]
Exceptional Grant Request (0/1)	0.040**	0.041**	0.041**	0.041**	0.041**	0.041**	0.041**	0.041**
	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]
Call Rejection Rate	0.454***	0.460***	0.461***	0.461***	0.462***	0.462***	0.462***	0.447***
	[0.092]	[0.092]	[0.092]	[0.092]	[0.092]	[0.092]	[0.092]	[0.093]
Contextually Novel (0/1)		-0.028***	-0.028***	-0.028***	-0.028***	-0.026***	-0.026***	-0.070**
		[0.005]	[0.005]	[0.005]	[0.005]	[0.006]	[0.005]	[0.027]
Alternative Framing (0/1)			0.026***	0.025***	0.021***	0.021***	0.021***	0.021***
			[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Innovativeness Framing			-0.005**		-0.006**	-0.005*	-0.006**	-0.006**
			[0.002]		[0.002]	[0.002]	[0.002]	[0.002]
Feasibility Framing				-0.006***	-0.007***	-0.007***	-0.006***	-0.007***
				[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Contextually Novel × Innovativeness Framing							-0.005	
							[0.006]	
Contextually Novel × Feasibility Framing								-0.005
								[0.006]
Contextually Novel × Call Rejection Rate								0.058
								[0.038]
Constant	0.938***	0.931***	0.920***	0.920***	0.929***	0.927***	0.928***	0.939***
	[0.147]	[0.146]	[0.145]	[0.144]	[0.145]	[0.145]	[0.145]	[0.145]
Call FE	y	y	y	y	y	y	y	y
Country FE	y	y	y	y	y	y	y	y
Industry FE	y	y	y	y	y	y	y	y
N	77717	77717	77717	77717	77717	77717	77717	77717
adj. R^2	0.141	0.142	0.143	0.143	0.143	0.143	0.143	0.143

Notes: Linear probability model estimates. Robust standard errors clustered at the application cutoff level in brackets.
* p<0.10, ** p<0.05, *** p<0.01.

Table C.3.8: Analysis of Rejection Probability, Split by Rejection Rate

	Low Rejection Rate			High Rejection Rate		
	(1)	(2)	(3)	(5)	(6)	(7)
ln(1+Firm age)	0.005	0.005	0.005	-0.003	-0.003	-0.003
	[0.008]	[0.008]	[0.008]	[0.003]	[0.003]	[0.003]
ln(1+Employees)	-0.014***	-0.014***	-0.014***	-0.005*	-0.005*	-0.005*
	[0.004]	[0.005]	[0.004]	[0.002]	[0.002]	[0.002]
Revenue (0/1)	-0.012	-0.012	-0.012	-0.023**	-0.023**	-0.023**
	[0.011]	[0.011]	[0.011]	[0.009]	[0.009]	[0.009]
Patent (0/1)	-0.092***	-0.092***	-0.092***	-0.096***	-0.096***	-0.096***
	[0.008]	[0.007]	[0.008]	[0.007]	[0.007]	[0.007]
Multiple Submissions (0/1)	-0.281***	-0.281***	-0.281***	-0.109***	-0.109***	-0.109***
	[0.015]	[0.016]	[0.016]	[0.009]	[0.009]	[0.009]
Proposal Word Count	-0.084***	-0.083***	-0.084***	-0.030***	-0.030***	-0.030***
	[0.007]	[0.007]	[0.007]	[0.005]	[0.005]	[0.005]
ln(Requested Grant)	-0.037***	-0.037***	-0.037***	-0.059***	-0.059***	-0.059***
	[0.007]	[0.007]	[0.007]	[0.002]	[0.002]	[0.002]
Exceptional Grant Request (0/1)	0.035*	0.035*	0.035*	0.028	0.028	0.028
	[0.017]	[0.017]	[0.017]	[0.029]	[0.029]	[0.029]
Call Rejection Rate	0.859***	0.862***	0.859***	0.701**	0.694**	0.699**
	[0.062]	[0.062]	[0.062]	[0.243]	[0.242]	[0.243]
Contextually Novel (0/1)	-0.015*	-0.010	-0.014	-0.006	-0.002	-0.010
	[0.008]	[0.009]	[0.010]	[0.008]	[0.008]	[0.007]
Alternative Framing (0/1)	0.015*	0.015*	0.015*	0.017*	0.017*	0.017*
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
Innovativeness Framing	0.004	0.007	0.004	-0.010**	-0.007*	-0.010**
	[0.006]	[0.007]	[0.006]	[0.004]	[0.003]	[0.004]
Feasibility Framing	-0.012**	-0.012**	-0.011**	-0.006*	-0.006*	-0.008**
	[0.004]	[0.004]	[0.005]	[0.003]	[0.003]	[0.003]
Contextually Novel × Innovativeness Framing		-0.013			-0.013	
		[0.014]			[0.011]	
Contextually Novel × Feasibility Framing			-0.001			0.010
			[0.014]			[0.006]
Constant	1.056***	1.051***	1.056***	1.086***	1.088***	1.089***
	[0.077]	[0.077]	[0.077]	[0.214]	[0.212]	[0.214]
Call FE	y	y	y	y	y	y
Country FE	y	y	y	y	y	y
Industry FE	y	y	y	y	y	y
N	20496	20496	20496	20630	20630	20630
adj. R^2	0.099	0.099	0.099	0.130	0.130	0.130

Notes: Linear regression estimates. Robust standard errors clustered at the application cutoff level in brackets.
Low (High) Rejection Rate defined as values below (above) the 25th (75th) percentile, equal to 0.64 (0.82), respectively.
* p<0.10, ** p<0.05, *** p<0.01.

Table C.3.9: Entropy Balancing Output

	Funded			Rejected (before balance)			Rejected (after balance)		
	Mean	Variance	Skewness	Mean	Variance	Skewness	Mean	Variance	Skewness
Firm Age	1.79	0.77	0.12	1.75	0.82	0.01	1.79	0.77	0.12
Proposal Word Count	2.73	0.22	-2.03	2.61	0.35	-1.66	2.73	0.22	-2.02
Employees	1.53	1.80	0.50	1.38	1.76	0.68	1.53	1.80	0.50
Patent (0/1)	0.59	0.24	-0.37	0.40	0.24	0.42	0.59	0.24	-0.37
Revenue (0/1)	0.60	0.24	-0.43	0.61	0.24	-0.45	0.60	0.24	-0.43
Multiple Submissions (0/1)	0.89	0.10	-2.50	0.82	0.15	-1.70	0.89	0.10	-2.50
ln(Applicants per Cutoff)	7.02	1.13	-1.49	7.15	0.86	-1.58	7.02	1.13	-1.49
ln(Requested Grant)	4.74	2.35	1.14	5.39	2.95	0.25	4.74	2.35	1.14
Exceptional Grant Request (0/1)	0.01	0.01	8.99	0.02	0.02	6.53	0.01	0.01	8.99

Table C.3.10: Analysis of Rejection Probability with Entropy Weights

DV: Rejected (0/1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(1+Firm age)	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
ln(1+Employees)	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Proposal Word Count	-0.002	0.001	-0.003	-0.001	-0.006	-0.006	-0.006	-0.006
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
Revenue (0/1)	-0.058***	-0.058***	-0.058***	-0.058***	-0.058***	-0.058***	-0.058***	-0.058***
	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]	[0.009]
Patent (0/1)	0.005	0.005	0.004	0.005	0.004	0.005	0.004	0.004
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
ln(Requested Grant)	-0.036***	-0.037***	-0.037***	-0.037***	-0.037***	-0.037***	-0.037***	-0.036***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
Exceptional Grant Request (0/1)	0.068***	0.069***	0.069***	0.068***	0.069***	0.068***	0.069***	0.071***
	[0.024]	[0.024]	[0.024]	[0.024]	[0.024]	[0.024]	[0.024]	[0.024]
Multiple Submissions (0/1)	-0.086***	-0.085***	-0.085***	-0.085***	-0.085***	-0.085***	-0.085***	-0.085***
	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]	[0.010]
Alternative Framing (0/1)	0.049***	0.048***	0.042***	0.042***	0.035***	0.035***	0.035***	0.035***
	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]	[0.008]
ln(Applicants per Cutoff)	-0.016	-0.015	-0.015	-0.015	-0.015	-0.015	-0.015	-0.015
	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]
Congestion Index	0.033**	0.032**	0.032**	0.033**	0.033**	0.033**	0.033**	0.055***
	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.016]	[0.019]
Contextually Novel (0/1)		-0.030***	-0.030***	-0.030***	-0.030***	-0.033***	-0.028***	-0.030***
		[0.006]	[0.006]	[0.006]	[0.006]	[0.007]	[0.008]	[0.006]
Innovativeness Framing			-0.012***		-0.013***	-0.015***	-0.013***	-0.013***
			[0.005]		[0.005]	[0.005]	[0.005]	[0.005]
Feasibility Framing				-0.011**	-0.012***	-0.012***	-0.011**	-0.012***
				[0.004]	[0.004]	[0.004]	[0.005]	[0.004]
Contextually Novel X Innovativeness Framing							0.009	
							[0.010]	
Contextually Novel X Feasibility Framing								-0.005
								[0.009]
Contextually Novel X Congestion Index								-0.084***
								[0.032]
Constant	1.655***	1.646***	1.662***	1.665***	1.684***	1.687***	1.681***	1.676***
	[0.091]	[0.090]	[0.090]	[0.091]	[0.091]	[0.091]	[0.091]	[0.090]
Call FE	y	y	y	y	y	y	y	y
Country FE	y	y	y	y	y	y	y	y
Industry FE	y	y	y	y	y	y	y	y
N	77717	77717	77717	77717	77717	77717	77717	77717
adj. R-sq	0.085	0.086	0.086	0.086	0.086	0.086	0.086	0.086

Notes: Linear probability with entropy weights. Robust clustered standard errors in brackets.
Sample is weighted following Hainmueller (2011), to achieve balance on all 1st, 2nd, and 3rd moments of control variables.
* p<0.10, ** p<0.05, *** p<0.01.

Table C.3.11: Analysis of Rejection Probability with Entropy Weights, Split by Congestion

DV: Rejected (0/1)	Low Congestion Index			High Congestion Index		
	(1)	(2)	(3)	(5)	(6)	(7)
ln(1+Firm age)	-0.016**	-0.016**	-0.016**	0.013	0.013	0.013
	[0.007]	[0.007]	[0.007]	[0.009]	[0.009]	[0.009]
ln(1+Employees)	0.005	0.005	0.005	-0.014**	-0.014**	-0.014**
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Proposal Word Count	-0.007	-0.006	-0.006	-0.036**	-0.037**	-0.035**
	[0.013]	[0.013]	[0.013]	[0.015]	[0.016]	[0.015]
Revenue (0/1)	-0.095***	-0.095***	-0.095***	-0.050***	-0.050***	-0.049***
	[0.016]	[0.017]	[0.017]	[0.018]	[0.018]	[0.018]
Patent (0/1)	0.022*	0.022*	0.022*	-0.024	-0.024	-0.024
	[0.012]	[0.012]	[0.012]	[0.015]	[0.015]	[0.015]
ln(Requested Grant)	-0.057***	-0.057***	-0.058***	-0.029***	-0.029***	-0.029***
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
Exceptional Grant Request (0/1)	0.063	0.063	0.063	0.064	0.063	0.064
	[0.051]	[0.051]	[0.051]	[0.044]	[0.044]	[0.044]
Multiple Submissions (0/1)	-0.068***	-0.068***	-0.069***	-0.128***	-0.128***	-0.127***
	[0.019]	[0.019]	[0.019]	[0.023]	[0.023]	[0.023]
Alternative Framing (0/1)	0.036**	0.036**	0.036**	0.021	0.020	0.021
	[0.016]	[0.016]	[0.016]	[0.017]	[0.017]	[0.017]
ln(Applicants per Cutoff)	0.053*	0.053*	0.053**	-0.010	-0.010	-0.010
	[0.027]	[0.027]	[0.027]	[0.027]	[0.027]	[0.027]
Congestion Index	0.051	0.051	0.051	0.013	0.012	0.014
	[0.054]	[0.054]	[0.054]	[0.056]	[0.056]	[0.056]
Contextually Novel (0/1)	-0.010	-0.008	-0.004	-0.035***	-0.039***	-0.024
	[0.012]	[0.013]	[0.014]	[0.013]	[0.015]	[0.016]
Innovativeness Framing	-0.021***	-0.020**	-0.021***	-0.020**	-0.023**	-0.020**
	[0.008]	[0.009]	[0.008]	[0.010]	[0.011]	[0.010]
Feasibility Framing	-0.010	-0.010	-0.007	-0.018**	-0.018**	-0.011
	[0.008]	[0.008]	[0.010]	[0.009]	[0.009]	[0.010]
Contextually Novel X Innovativeness Framing		-0.005			0.013	
		[0.017]			[0.020]	
Contextually Novel X Feasibility Framing			-0.015			-0.028
			[0.018]			[0.018]
Constant	1.753***	1.750***	1.745***	1.514***	1.521***	1.499***
	[0.258]	[0.259]	[0.261]	[0.164]	[0.165]	[0.163]
Call FE	y	y	y	y	y	y
Country FE	y	y	y	y	y	y
Industry FE	y	y	y	y	y	y
N	20046	20046	20046	19529	19529	19529
adj. R-sq	0.100	0.100	0.100	0.133	0.133	0.133

Notes: Linear probability with entropy weights. Robust clustered standard errors in brackets.

Sample is weighted following Hainmueller (2011), to achieve balance on all 1st, 2nd, and 3rd moments of control variables.

* p<0.10, ** p<0.05, *** p<0.01

C.4 Details on Framing Variables

Table C.4.1: Framing Vocabulary

Category	Vocabulary keywords
Innovativeness	creative; emerging; first; innovative; groundbreaking; latest; novel; revolutionary; unique; unparalleled; unprecedented; breakthrough.
Feasibility	accurate; advanced; compatible; careful; detailed; cohesive; nuanced; powerful; quality; reproducible; rigorous; robust.

Example of framing emphasis on *Innovativeness*:

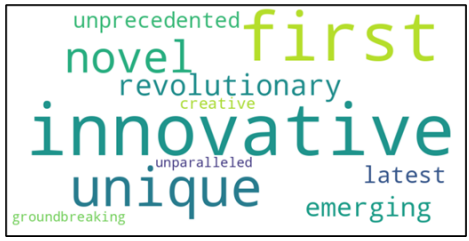
- The [VENTURE NAME] project is a ***breakthrough*** indoor air-cleaning technology targeting not just the elimination of indoor air pollution but also the ***creation*** of healthy new air thanks to a powerful combination of real plants with technology and design [TECHNOLOGY]. The ***novelty*** of the [VENTURE NAME] project is based on the introduction of the forced-air system in the natural air purification arena [. . .].

Example of framing emphasis on *Feasibility*:

- To fill the customer and societal need for better industrial energy optimization, we have created a ***robust*** and secure hardware-neutral [TECHNOLOGY] to optimize energy, manage assets, integrate and analyze data. [TECHNOLOGY] is ***compatible*** with pre-installed hardware and control systems. [VENTURE NAME] will be based on this ***validated*** [TECHNOLOGY] but will include significant upgrades [. . .].

Figure C.4.1: Word Clouds Based on Absolute Frequency

(a) Innovativeness



(b) Feasibility

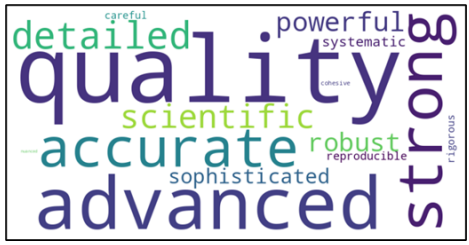
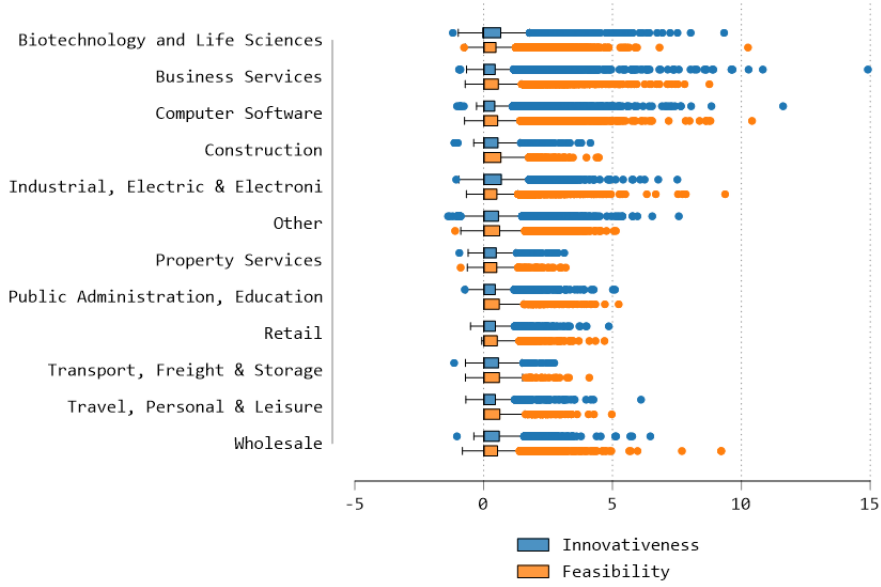


Figure C.4.2: Standardized Framing Emphasis by Industry



C.5 Details on Congestion Index

We begin by noting the variation in submission volume across specific application deadlines and industries. To illustrate this point, Figures A1 and A2 report the raw number of application submissions for selected industries. Moreover, this variation can be further uncovered by removing the purely "scale" factor. We define *rel_variation* as the percentage deviation in the number of funding applications submitted within a specific industry and application deadline period relative to that industry's historical average. Positive values indicate above-average application volume (e.g., $rel_variation = 10$ means ten percent more applications than the historical average, etc.), and we plot this variation for the previously selected industries in Figures A3 and A4 below. These figures suggest a relatively high magnitude of variation in application submissions across time and industries.

To capture idiosyncratic fluctuations in the volume of applications, we define *Congestion Index* following the residual-based approach put forward by Ferguson & Carnabucci (2017). We first build an unbalanced panel of 1,206 observations (30 industries $\times$ 61 month-year application deadlines from the EIC Accelerator), and calculate the industry-specific historical average application volume (*avg_applications_sector*). I then generate the demeaned log-application count by deadline to isolate deviations from secular trends by industry. Next, we regress this log-transformed demeaned count of applications by sector on a full set of industry fixed effects and application deadline fixed effects, using robust standard errors to account for heteroskedasticity.

The residuals from this regression can be interpreted as the idiosyncratic congestion, i.e., unexplained surges or shortfalls in application volume relative to both industry, specific baselines and time/deadline shocks. Positive residuals indicate higher than expected congestion (which relates to evaluator overload). Crucially, this approach improves upon raw counts or industry-normalized application counts by simultaneously controlling for cross-sector and temporal shocks, ensuring congestion variation is plausibly driven by idiosyncratic factors net of systematic variation.

We provide further validation of the properties of our *Congestion Index* variable. In Figure B.5.4 we report the kernel density; Figure B.5.3 report the essentially flat relationship between the raw number of applications and our *Congestion Index* variable, with an estimated correlation coefficient of 0.05. We further report that the distribution of the *Congestion Index* is centered around zero across industries. Finally, in Table B.5.1 we report a balance test of the *Congestion Index*, which demonstrates no systematic association with other variables used in the analysis of idea rejection.

Figure B.5.1: Number of Applications over Time, Selected Industries

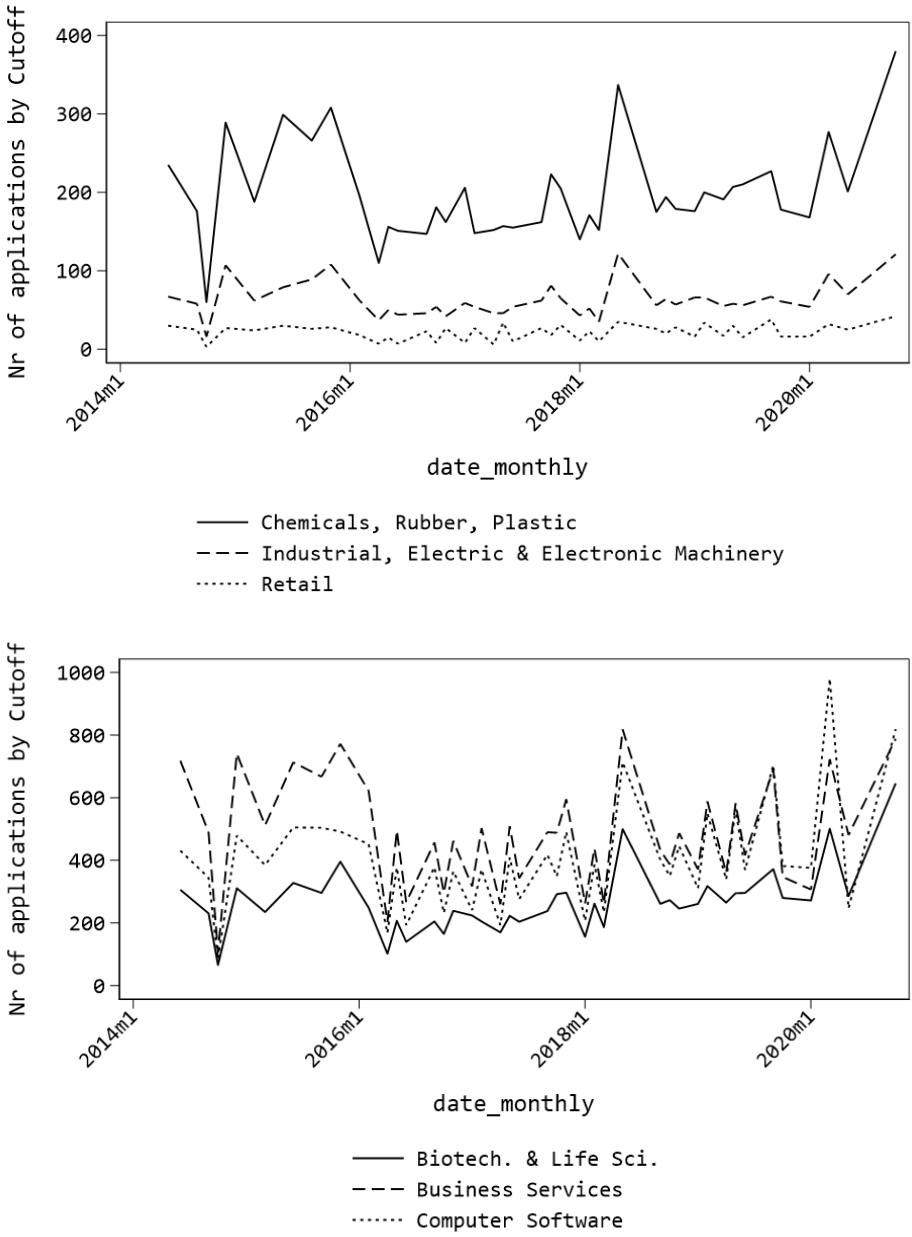


Figure B.5.2: Congestion Index Density

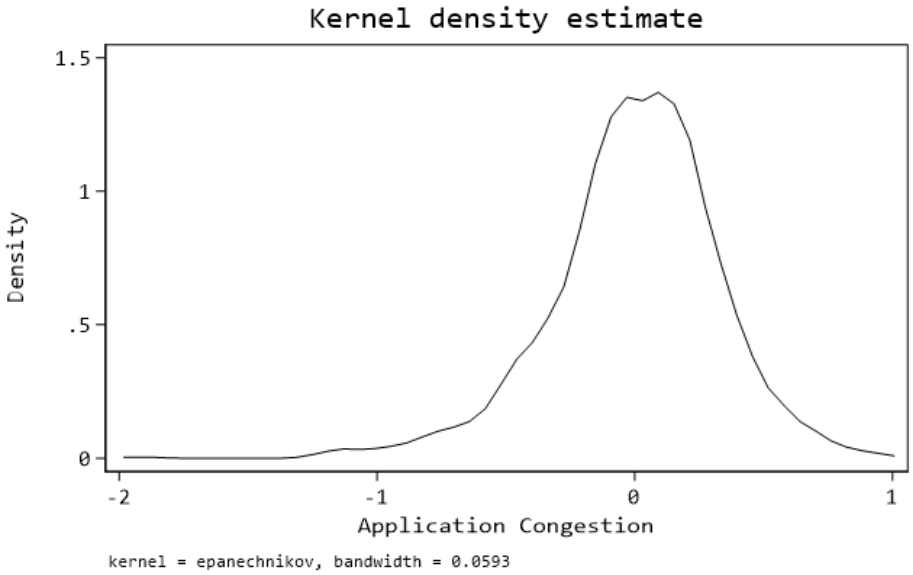


Figure B.5.3: Scatterplot Congestion Index and Submissions

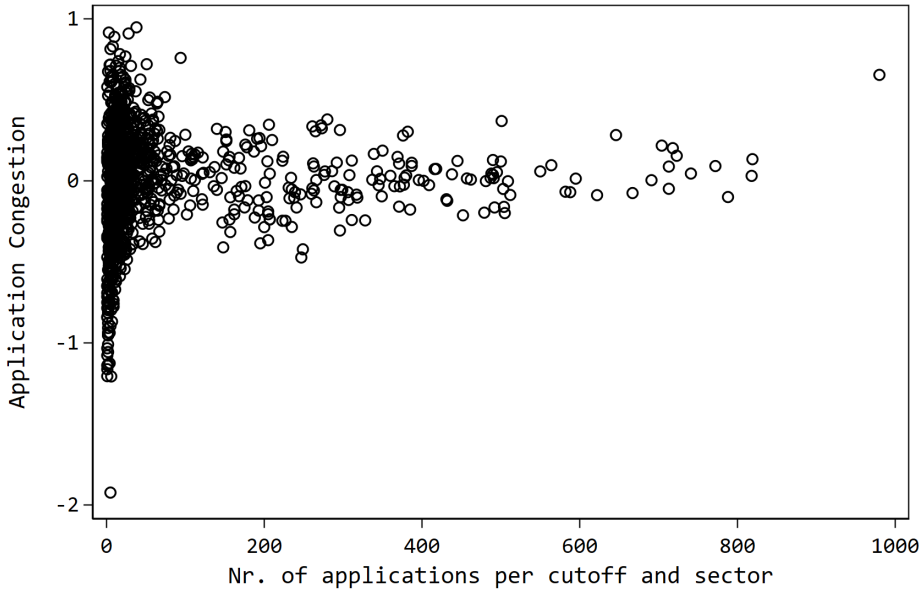


Figure B.5.4: Boxplot Congestion Index by Industry

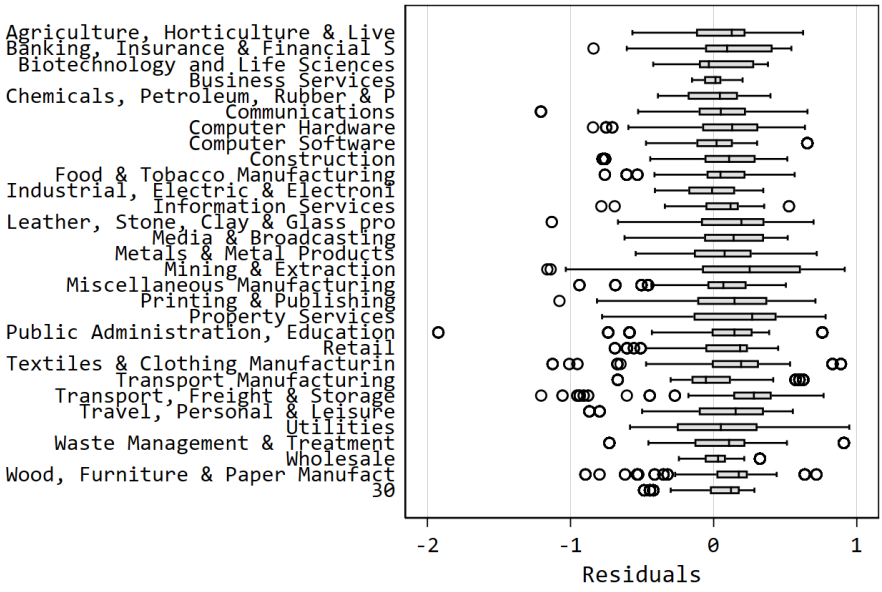


Table B.5.1: Balance of Congestion Index

DV: Congestion Index	(1)
ln(1+Firm Age)	0.000
	[0.000]
ln(1+Employees)	0.006*
	[0.003]
Proposal Word Count	0.000
	[0.005]
Revenue (0/1)	-0.004
	[0.005]
Patent (0/1)	0.000
	[0.008]
Multiple Submissions (0/1)	-0.011
	[0.012]
N	78412

Notes: OLS estimates with robust standard errors. *:$p < 0.1$

Find all doctoral defences at KU Leuven under:
https://research.kuleuven.be/en/doctoraldefences/index.

Faculty of Economics and Business
Naamsetraat 69
3000 LEUVEN, België
tel. +32 16 32 66 12
feb.leuven@kuleuven.be
www.kuleuven.be